Systematics, Biogeography, and Phylogeography of *Thylamys* Mouse Opossums, a Recent Radiation of Neotropical Marsupials

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

Thomas Christopher Giarla

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

Sharon A. Jansa

August 2013

ACKNOWLEDGEMENTS

This dissertation would have been impossible without the kindness, support, and advice of my advisor, Sharon Jansa, who pushed me to follow my independent interests and provided me with the resources and connections that have led to its successful completion. Robert Voss was a crucial collaborator, especially as I was first developing my project. His eye for detail improved this project immeasurably, and he contributed substantially to Chapters 1 and 2. Discussions with my committee members, Keith Barker, Ken Kozak, and Andrew Simons, enriched my understanding of important methodological issues and helped narrow my research questions. Discussions and comments on earlier drafts of each chapter were generously provided by Andrew Simons, Keith Barker, Robert Voss, Juan Diaz, Silvia Pavan, and Robert Anderson.

This work was supported by two grants from the National Science Foundation: DEB-1110365 to Sharon Jansa and me and DEB-0743062 to Sharon Jansa and Robert Voss. Additional funds were provided by the University of Minnesota, the Bell Museum of Natural History, the American Society of Mammalogists, and the Society of Systematic Biologists. Molecular work for this project was completed with help from Tracy Smelter, Tim Sosa, Courtney Comar, and Lorissa Fujishin. Museum curators, curatorial assistants, and collectors from all over the world generously provided tissues and morphological specimens. During fieldwork in Argentina, Gabriel Martin was my colleague, guide, and translator, and for that I am grateful. I would like to acknowledge the amazing, encouraging team of mentors, colleagues, and friends that I have met at the

University of Minnesota. Without them, this whole endeavor would have been much less fun and enriching. I also want to thank my dear friend Rafael Walker, who encouraged me to pursue a Ph.D. in the first place and supported me from start to finish. Finally, I thank my husband, Dan Covich, who moved half-way across the country and altered his career plans to be with me in Minnesota. He helped me escape the academic bubble every now and then, keeping me sane, healthy, and happy over the past six years.

DEDICATION

I dedicate this dissertation to my parents, who fostered my love for biology by filling my childhood with bug collections, exotic vacations, odd pets, and a backyard biome worth exploring.

ABSTRACT

This project broadly explores the systematics, biogeography, and phylogeography of *Thylamys* mouse opossums, a genus of Neotropical marsupials from central and southern South America. Chapter 1 is part of a collaborative work with Robert Voss and Sharon Jansa. In it, we resolve longstanding issues surrounding *Thylamys* taxonomy and nomenclature using mitochondrial DNA sequences and morphology, and provide the first phylogenetic hypothesis for all recognized species in this genus. We recognize nine species but also uncover numerous morphologically cryptic mitochondrial haplogroups within four species. In Chapter 2, I assess the evolutionary independence of a subset of these morphologically cryptic lineages within the montane species *T. pallidior*, *T. sponsorius*, and *T. venustus*. I find evidence to support the existence of two lineages within each of the three species, and also conduct tests to determine the number of nuclear loci needed to confidently test species limits. In Chapter 3, I examine the biogeographic history of *Thylamys* and its monotypic sister-genus *Lestodelphys*, considering the impact of habitat type and physical barriers on range evolution and cladogenesis. In Chapter 4, I test predictions regarding the impact of late Quaternary glacial cycles on the evolutionary history of six montane cryptic lineages. I estimate divergence times and demographic shifts for each lineage, and find limited support for the core predictions. Two supplementary files are provided online as part of this dissertation: a file containing 15 phylogenetic trees for each of the loci considered in Chapter 2 (Online Supplementary File 1) and a file containing GenBank accession

numbers and tissue voucher numbers for the sequences included in the supermatrix in Chapter 3 (Online Supplementary File 2).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

SPECIES LIMITS AND PHYLOGENETIC RELATIONSHIPS IN THE DIDELPHID MARSUPIAL GENUS *THYLAMYS* BASED ON MITOCHONDRIAL DNA SEQUENCES AND MORPHOLOGY

INTRODUCTION

Species of the didelphid marsupial genus *Thylamys*, commonly known as fat-tailed mouse opossums, are widely distributed in savanna woodlands, thornscrub, semidesert shrublands, shrubby steppe, montane thickets, and other more-or-less open, nonforest habitats from western Peru and eastern Brazil southward to central Chile and Patagonia. Their common English name refers to the seasonal deposition of caudal fat (**Figure 1**), a phylogenetically derived trait that *Thylamys* shares with its sister taxon *Lestodelphys* (see Morton 1980). Together, *Thylamys*, *Lestodelphys*, and four other didelphid genera (*Chacodelphys*, *Cryptonanus*, *Gracilinanus*, *Marmosops*) comprise the tribe Thylamyini, one of two well-supported suprageneric groups of small, mouse-like or "murine" opossums (Voss and Jansa 2009).

Although the family Didelphidae is a predominantly lowland rainforest clade, *Thylamys* is the most species-rich and widely distributed of several opossum lineages that invaded nonforest habitats in the Miocene (Jansa et al. *in review*). Because geographic patterns of endemism and cladogenesis among fat-tailed mouse opossums are potentially informative about the late Tertiary history of nonforest environments in South America, the systematics of this group are of considerable biogeographic interest. Unfortunately,

neither the number of valid species in the genus nor their evolutionary relationships has been convincingly resolved by previous studies.

Figure 1. *Thylamys pallidior* from Chubut, Argentina photographed by Darío Podestá.

TAXONOMIC HISTORY

Most of the nominal taxa now included in *Thylamys* were referred to the "Elegans Group" of *Marmosa* by Tate (1933), who recognized eight of them as valid species. Subsequent authors have recognized as few as one to as many as ten species in this group, which was first recognized as a valid genus in its currently accepted sense by Gardner and Creighton (1989). Although the current classification (Creighton and Gardner 2008) is based on an impressive number of recent revisionary studies (Palma and Yates 1998; Flores et al. 2000; Meynard et al. 2002; Palma et al. 2002; Solari 2003; Braun et al. 2005; Carmignotto and Monfort 2006), no study to date has included material from all of the species now regarded as valid.

Another notable aspect of most previous systematic analyses of *Thylamys* is an exclusive reliance either on molecular or morphological methods. Both approaches have unique advantages and limitations. Analyses based on DNA sequencing, for example, have the advantage of abundant character data, but sequence data can be difficult to interpret in the absence of phenotypic criteria linking gene trees to name-bearing types and other unsequenced material. Additionally, gene sequencing is relatively expensive, and the restricted availability of preserved tissue samples has hitherto limited molecular studies to treatments of no more than a few dozen individuals. By contrast, morphological analyses have the advantage of almost cost-free data collection and abundant material, but they are often hindered by a poverty of taxonomically informative characters.

Herein we combine both approaches by analyzing >100 mitochondrial gene sequences representing populations throughout the known distribution of *Thylamys*, including exemplars of all currently recognized species and most of the nominal taxa purported to be valid. We associated these sequence data with phenotypic character variation by examining morphological voucher material, and we assessed the taxonomic implications of our results by comparing voucher material with holotypes, neotypes, and other museum specimens.

Species Recognition Criteria

Despite a plethora of published species definitions (reviewed by de Queiroz 1998), most evolutionary biologists agree that contemporaneous species should be genetically independent lineages. For a number of theoretical and practical reasons, mtDNA sequence monophyly and morphological diagnosability are now routinely used

as joint criteria for species recognition by vertebrate systematists (Zamudio and Greene 1997; Frost et al. 1998; Patton et al. 2000; Wiens and Penkrot, 2002). Because substitution rates and haplotype coalescence are, in general, much more rapid for mitochondrial than for nuclear genes (Moore 1995; Palumbi and Cipriano 1998), mitochondrial sequence data are potentially more informative than nuclear DNA with regard to recent lineage formation. Morphological diagnosability, on the other hand, is important as a proxy for nuclear-gene divergence (Wiens and Penkrot 2002) because it allows haplotype groups to be associated with appropriate clade names, and because it allows geographic ranges to be mapped without sequencing every specimen.

We recognize that some valid species might not be recovered as monophyletic haplotype groups due to incomplete lineage sorting of ancestral polymorphisms (Funk and Omland 2003), and we are aware that some genetically isolated lineages might not have evolved diagnostic morphological characters (Baker and Bradley 2006). However, given that speciation is a process, practical problems of delimitation will be encountered no matter what operational criteria are employed (Wiens and Penkrot 2002; Sites and Marshall 2003). In effect, our criteria are probably conservative, and in this respect our operationally defined species constitute hypotheses to be tested by future research.

Although a high degree of sequence divergence is not a necessary or sufficient criterion for species recognition (Ferguson 2002; Baker and Bradley 2006), it is relevant to note that congeneric didelphid species recognized on the basis of haplotype monophyly and morphological diagnosability (e.g., by Mustrangi and Patton 1997; Patton and da Silva 1997; Patton et al. 2000; Costa et al. 2003; Solari 2007) differ by an

average K2P-corrected pairwise divergence of about 10% at the cytochrome-*b* (CYTB) locus. This is (coincidentally) close to the threshold value for species delimitation suggested by Bradley and Baker (2001), but some didelphids that maintain unambiguously diagnostic morphological differences in sympatry—and so would be recognized as valid species by any current definition—have much less divergent CYTB sequences (e.g., 4% between *Philander mcilhennyi* and *P. opossum*; Patton and da Silva 1997). Therefore, although we estimate pairwise sequence divergence for comparative purposes, we do not necessarily consider divergence values to be critical indicators of species status.

MATERIALS AND METHODS

Museum Collections

Morphological specimens and tissue samples used in the course of this study or cited in this report are preserved in the following collections: AMNH (American Museum of Natural History, New York), ANSP (Academy of Natural Sciences, Philadelphia), BMNH (Natural History Museum, London), CML (Colección de Mamíferos Lillo, Universidad Nacional de Tucumán, Tucumán), CNP (Centro Nacional Patagónico, Puerto Madryn), FMNH (Field Museum, Chicago), IADIZA (Instituto Argentino de Investigaciones en Zonas Áridas, Mendoza), MLP (Museo de La Plata, La Plata), MMNH (J.F. Bell Museum of Natural History, University of Minnesota, St. Paul), MNHN (Muséum National d'Histoire Naturelle, Paris), MSB (Museum of Southwestern Biology, University of New Mexico, Albuquerque), MVZ (Museum of Vertebrate Zoology, University of California, Berkeley), MZUSP (Museu de Zoologia da Universidad do São Paulo, São Paulo), NMW (Naturhistorisches Museum Wien,

Vienna), OMNH (Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, Norman), TTU (Museum of Texas Tech University, Lubbock), UMMZ (University of Michigan Museum of Zoology, Ann Arbor), USNM (National Museum of Natural History, Smithsonian Institution, Washington), UWBM (University of Washington Burke Museum, Seattle), ZMUC (Zoological Museum of the University of Copenhagen, Copenhagen), ZSM (Zoologische Staatssammlung München, Munich).

Taxon Sampling

We sequenced 131 specimens of *Thylamys* for this study (Appendix 1 and Appendix 2), including representative material of all species currently recognized as valid by authors (Gardner 2005; Creighton and Gardner 2008; Teta et al. 2009). Within most species, multiple individuals from throughout the known geographic range were selected for sequencing in order to sample intraspecific genetic variation as completely as possible. In addition, we attempted to include representative material of all described nominal taxa (synonyms and "subspecies") to resolve longstanding nomenclatural issues.

Outgroup taxa were selected based on the phylogenetic results of Voss and Jansa (2009). These include *Lestodelphys halli* (the sister taxon of *Thylamys*) and exemplar species of three other thylamyine genera (*Cryptonanus*, *Gracilinanus*, *Marmosops*). On the assumption of thylamyine monophyly, sequences downloaded from GenBank for the marmosine species *Monodelphis domestica* (complete mitochondrial genome accession number AJ508398) were used to root all of our trees.

DNA Extraction and Gene Sequencing

For the majority of specimens, DNA was extracted using a Qiagen DNA Minikit (Qiagen Inc.) from kidney, liver, or muscle tissue that had been preserved in ethanol or frozen in the field. However, we also extracted DNA from pieces of dried tissue snipped from skins or scraped from skeletal specimens. All extractions and subsequent sequence amplifications for dried tissues were conducted in a UV-sterilized hood to avoid contamination with foreign DNA. To further minimize the risk of contamination, all procedures conducted prior to amplification were conducted in a laboratory in which amplified mammalian DNA has never been present. Most of the museum specimens from which such material was removed are decades old and have been handled extensively, so it was necessary to thoroughly wash the samples in order to remove foreign DNA and potential PCR inhibitors. First, we washed whole pieces of the dried tissue in 800 µL of 100% ethanol by vortexing the sample for one minute and letting it rest overnight. The following day, we removed the liquid, resuspended the tissue in 100% ethanol, vortexed the sample for an additional minute, and then immediately repeated the same wash. We repeated a similar set of three washes (one overnight, then two for one minute) using 70% ethanol and then, finally, using water, for a total of nine washes per sample. We extracted DNA using a Qiagen DNA Minikit (Qiagen Inc.) after the washed tissue samples were digested for 48 hours in a solution of tissue lysis buffer and as much proteinase K as was necessary to completely digest the sample, typically 30–40 µL.

We amplified and sequenced three mitochondrial protein-coding genes for this study: cytochrome *b* (CYTB, ~1.1 kb), cytochrome *c* oxidase subunit II (COX2, ~0.7 kb), and NADH dehydrogenase 2 (ND2, ~1.0 kb). We sequenced CYTB for all individuals

(131 ingroup, five outgroup), whereas COX2 and ND2 were sequenced for all of the outgroup taxa and for a subset of 28 individuals representing all major CYTB haplogroups (see Results). For frozen and ethanol-preserved samples, amplification of CYTB and ND2 required two steps: one round to amplify the whole gene, followed by a second round of amplification using internal primers, which, depending on the gene and taxon, yielded two or three appropriately sized fragments for sequencing. The targeted COX2 sequence was short enough to amplify and sequence in one step. Museum specimens typically yielded highly degraded DNA, which necessitated the use of dozens of specific primer pairs that amplified 150–300 bp fragments. All primer sequences can be found in Appendix 3.

We cleaned PCR products of excess primers and unincorporated nucleotides prior to sequencing using Exonuclease I and Shrimp Alkaline Phosphatase (Hanke and Wink 1994). We sequenced these products in both directions on an ABI 3730 automated sequencer (Applied Biosystems Inc.) using amplification primers and dye-terminator chemistry (BigDye ver. 3.1 Cycle Sequencing Kit, Applied Biosystems Inc., Foster City, California). We used Sequencher 4.7 (GeneCodes Inc., Ann Arbor, Michigan) to compile and edit the sequences and MEGA 4.0 (Tamura et al. 2007) to determine sequence characteristics and genetic distances. All sequences have been deposited in GenBank (CYTB: HM583364–HM583499; COX2: HM583500–HM583532; ND2: HM583533–HM583566).

Phylogenetic Analyses

DNA sequences for each gene were aligned with reference to translated amino acid sequences; no insertions or deletions were necessary. We performed phylogenetic

analyses using Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI). Analyses were conducted with four datasets: one for each of the three genes alone and one that includes the three genes in combination. The CYTB dataset included multiple sequences for most of the species, ranging from one representative individual (*T. karimii*) to 48 (*T. venustus*). In contrast, the ND2 and COX2 datasets include no more than three sequences from any given haplogroup (see Results). Because of the disparity in size among the single-gene datasets, the combined-gene dataset includes CYTB sequences from only those individuals for which we had also sequenced ND2 and COX2.

Maximum parsimony and maximum likelihood analyses. For each of the four datasets, we implemented MP analyses in PAUP* 4.0 (Swofford 2002) using a heuristic search strategy with TBR branch swapping, ten random-addition replicates, and with all characters unordered and equally weighted. We assessed nodal support using 5000 bootstrap pseudoreplicates with TBR branch swapping and ten random-addition replicates. For the CYTB dataset, we first pruned out identical sequences in order to maximize computational efficiency.

We conducted ML analyses of each of the four datasets in GARLI ver. 0.96 (Zwickl 2006) by specifying the best-fit model for each dataset as determined in MrModelTest 2.3 (Nylander 2004) and allowing the model parameters to be estimated from the data. We did not alter the default parameters used in the program's genetic algorithm, and we set the analyses to automatically terminate after 20,000 generations with no significant change in the log-likelihood score. In order to ensure that the program had arrived at a stable log-likelihood estimate, we repeated each search five times and

report only the estimate with the highest log-likelihood. We measured nodal support using 500 bootstrap replicates and the same settings used in the individual ML analyses.

Data partitioning in a Bayesian framework. Combining data from multiple genes and treating each codon position identically assumes that each site within the dataset evolves under the same nucleotide substitution model. If this assumption is invalid, as is often the case, the dataset can be partitioned into independent units (e.g., Bull et al. 1993; Nylander et al. 2004) in order to avoid misleading results (Brown and Lemmon 2007). For each of the four datasets, we evaluated the relative fit of different data-partitioning strategies in a Bayesian framework. For the three single-gene datasets, we tested whether a three-partition (by codon position) model fit the data significantly better that a single (unpartitioned) model. For the combined-gene dataset, we evaluated the relative fit of four partitioning strategies: (1) a single partition for all genes in combination (combined: unpartitioned); (2) three partitions, one for each of the three genes (by gene); (3) three partitions, one for each of the codon positions across all genes (combined: by codon position); and (4) nine partitions, one for each codon position for each of the three genes (by gene by codon position).

Several statistical approaches have been suggested in the literature for evaluating the relative fit of different partitioning schemes, including Bayes factors (Kass and Raftery 1995), the Akaike Information Criterion (AIC; Akaike, 1974), and the Bayesian Information Criterion (BIC; Schwarz, 1978). Although Bayes factors have been shown to be reasonably effective tools for choosing among partitioning schemes (Brown and Lemmon 2007), the number of parameters contained in alternative models is not explicitly incorporated into the calculation of Bayes factors. The AIC and BIC both apply

penalties to the addition of parameters, mitigating the problems associated with over-parameterization (McGuire et al. 2007). Therefore, to account for potential over-parameterization, we only ranked partitioning strategies using the AIC and BIC.

To implement this approach, we obtained log-likelihood scores and model parameter values for each of the partitioning strategies described above for the four datasets using BI as implemented in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). We first determined the best-fitting nucleotide substitution model for each individual partition using the AIC as implemented in MrModeltest. We specified the appropriate model for each partition in MrBayes and, for strategies with more than one partition, we unlinked all substitution parameters across partitions. We estimated branch lengths proportionally to each partition and allowed the substitution rate to vary across each partition by setting the rate prior option to "variable." We ran two independent replicates of the Metropolis-Coupled Markov Chain Monte Carlo analysis for 5×10^6 generations, sampling every 100^{th} generation, with one cold chain and three incrementally heated chains.

Once these runs were completed, we examined a plot of the log-likelihood scores per generation using Tracer 1.4 (Rambaut and Drummond 2007) to determine the appropriate number of generations to discard as burn-in. For all strategies and datasets, we discarded the first 500,000 generations (5,000 saved samples)—well after the log-likelihood scores stabilized—and used the remaining 45,000 samples to estimate mean log-likelihood, nodal posterior probability (BPP) values, mean branch lengths, and other parameters. In order to compare the results of different partitioning strategies, we calculated the harmonic mean log-likelihood (HML) across all parameters for each

strategy in Tracer. We then used the HML values to calculate the AIC and BIC scores according to the methods of McGuire et al. (2007).

Topology tests. Previously published molecular and morphological research on *Thylamys* has produced several phylogenies that differ from ours. In order to evaluate these alternative hypotheses, we conducted topology tests using Shimodaira-Hasegawa (SH) likelihood-ratio tests (Shimodaira and Hasegawa 1999). First, an ML analysis was performed in GARLI following the same protocol described above, except that constraint trees were used to infer the best tree that conformed to the alternative hypothesis in question. We then implemented the SH test in PAUP* using 1000 Resampling Estimated Log Likelihoods (RELL) replicates to compare the constrained ML tree to the best ML tree from the above analysis, allowing the program to estimate all parameters.

Parsimony ancestral state reconstruction. Species and haplogroups that comprise the Elegans Group (see Results) occur on both the eastern and western sides of the Andes. We used the combined-gene ML topology to reconstruct the ancestral distributions of Elegans Group lineages relative to the Andes. Tips were coded as occurring east of the Andes (cis-Andean), west of the Andes (trans-Andean), or both (ambiguous). We optimized ancestral areas on the topology using the parsimony criterion for ancestral state optimization of unordered characters as implemented in MacClade 4.08 (Maddison and Maddison, 2000).

Morphology

We recorded measurement data and scored qualitative morphological characters from voucher specimens in order to assess the phenotypic distinctness of mitochondrial

haplotype clades and to provide a basis for comparisons with holotypes and other unsequenced material. Except as noted otherwise below, all analyzed character data were obtained from adult specimens as determined by dental criteria. Specimens were judged to be *juvenile* if dP3 was still in place, *subadult* if dP3 had been shed but P3 was still incompletely erupted, and *adult* if the permanent dentition was complete. We examined skins, skulls, and fluid preserved specimens for taxonomic variation in qualitative morphological characters. Surveyed characters included those previously described in the didelphid taxonomic literature (e.g., Tate, 1933; Flores et al. 2000; Solari 2003; Voss and Jansa, 2003; Carmignotto and Monfort 2006) and others newly reported herein. Nomenclature used to describe observed variation in integumental, osteological, and dental features follows standard anatomical usage as reviewed by Voss and Jansa (2009). Morphological and taxonomic results are discussed in detail in Giarla et al. (2010).

RESULTS

Sequence Characteristics

Our cytochrome-*b* dataset comprises aligned sequence data for 131 ingroup individuals. In 115 cases, we obtained sequence from the complete CYTB gene (1149 bp); the remaining 16 CYTB sequences range in length from 139 to 1038 bp. In addition, we sequenced 283–1044 bp of the ND2 gene and 178–684 bp of the COX2 gene from 28 individuals representing all but one of the major lineages recovered from analyses of the CYTB dataset. We combined the resulting sequences with CYTB sequences from the same individuals to obtain a concatenated, three-gene alignment that was 2877 bp long. We were unable to amplify ND2 or COX2 from any of the undescribed Peruvian specimens (*Thylamys* sp.; see Results) and therefore scored those genes as missing data

for the only example of the undescribed Peruvian form included in our combined-gene analysis.

Nucleotide substitution characteristics were estimated for each gene using the best-fitting model. For all three genes, a general-time-reversible (GTR) model best fits the data (**Table 1**). Each model also incorporates an among-site rate heterogeneity parameter (α) as approximated by the gamma distribution and a proportion of invariant sites parameter (p_{inv}). Base frequency parameter estimates (πA, πC, πG, πT) demonstrate that all three genes are AT-rich with a marked paucity of guanine residues, similar to other estimates of the base composition of mammalian mitochondrial DNA (Irwin et al. 1991; Patton et al. 1996). None of the three genes showed any departure from base-compositional stationarity among individuals (CYTB: $\chi^2 = 86.19$ [$df = 405$], $P = 1.00$; ND2: $\chi^2 = 51.45$ [$df = 69$], $P = 0.94$; COX2: $\chi^2 = 14.59$ [$df = 78$], $P = 1.00$).

Table 1. Results of Maximum-likelihood Model Fitting with Single-gene Datasets

	CYTB	**CYTB**	**ND2**	**COX2**
No. sequences	137	35	34	32
Model	GTR + I + Γ	GTR + I + Γ	GTR + I + Γ	GTR + I + Γ
πA	0.35	0.35	0.37	0.36
πC	0.28	0.28	0.30	0.22
πG	0.05	0.06	0.05	0.11
πT	0.32	0.31	0.29	0.31
rAC	0.82	1.16	0.34	1.04
rAG	39.2	41.11	11.4	18.94
rAT	1.81	2.35	0.63	1.93
rCG	1.47	0.77	0.27	0.38
rCT	18.9	23.43	4.58	21.07
α	1.79	1.93	1.28	1.15
p_{inv}	0.54	0.55	0.36	0.57
LnL best tree	-10723.24	-8867.26	-8867.43	-4341.90

Phylogenetic Analyses

For each of the four datasets, phylogenetic relationships were inferred using MP, ML, and BI analyses. Maximum parsimony analysis of each dataset resulted in well-resolved strict consensus trees (**Table 2**). Because we recovered the same species-level clades in all analyses of each dataset, we show only the ML phylogenies in this report and superimpose nodal support values from each analysis on that tree. For MP and ML analyses, support for a node is considered strong if that node receives bootstrap values 75% or greater, moderate for values between 50% and 75%, and weak for values of 50% or less. For BI, nodes are considered strongly supported for BPP values greater than or equal to 0.95 and weakly supported for values less than 0.95.

Table 2. Results of Parsimony Analyses of Molecular Datasets

	CYTB (Full taxon sampling)	**CYTB (Subsample)**	**ND2**	**COX2**	**Concat.**
Ingroup sequences	100	29	28	27	29
Outgroup sequences	6	6	6	5	6
Aligned sites	1149	1149	1044	684	2877
Variable sites	501	489	590	257	1336
% variable sites	43.6	42.6	56.5	37.6	46.5
Informative sites	456	431	493	210	1134
Minimum-length trees	324	1	1	6	2
Length of best tree(s)	2022	1767	1819	820	4431
Consistency Index	0.35	0.40	0.48	0.44	0.44
Retention Index	0.65	0.67	0.70	0.65	0.68
Resolved nodes (%)	72 (72.7)	28 (100)	27 (100)	24 (92.3)	28 (100)

Individual BI analyses were conducted for each data partitioning strategy and results were ranked by the AIC and BIC according to their fit. For each dataset, a GTR+I+Γ nucleotide substitution model was found to fit each partition best. For the CYTB and ND2 single-gene datasets, the "by codon" partitioning scheme fit the data

better than the unpartitioned scheme. The unpartitioned scheme fit the data best for the single-gene COX2 dataset. For the combined-gene dataset, the partitioning strategy that pooled codon positions across the three genes fit the data best.

Species limits and phylogeographic structure. Nine of the monophyletic groups of *Thylamys* that we recovered by analyzing cytochrome-*b* sequence data appear to represent valid species (**Figure 2**, **Table 3**). In order to simplify the presentation of these results, we anticipate our taxonomic conclusions by using Latin binomials for reciprocally monophyletic groups of sequences obtained from morphologically differentiated series of voucher specimens, and we use alphabetical designations (A, B, C) for reciprocally monophyletic groups of apparently conspecific sequences ("haplogroups").

Table 3. Uncorrected Cytochrome-b p-Distances (scaled as percent sequence divergence; below diagonal), K2P-Corrected Distances (above diagonal), and Intraspecific Distances (diagonal).[1]

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. *T. elegans*	**2.5**	17.1	15.3	10.3	14.8	10.6	16.5	10.7	17.7	16.0
2. *T. karimii*	14.9	**-**	15.7	16.3	16.6	15.1	18.7	15.5	15.6	18.4
3. *T. macrurus*	13.5	13.9	**0.8**	15.2	15.7	14.6	15.7	14.7	18.5	15.9
4. *T. pallidior*	9.4	14.3	13.4	**2.7**	13.7	10.9	18.2	10.6	16.3	16.3
5. *T. pusillus*	13.1	14.6	13.8	12.2	**5.1**	13.4	17.9	13.7	18.0	16.7
6. *T. sp.*	9.7	13.4	13.0	9.9	12.1	**2.9**	17.2	6.8	16.9	15.9
7. *T. sponsorius*	14.4	16.1	13.8	15.6	15.4	14.9	**1.8**	17.7	18.8	14.1
8. *T. tatei*	9.8	13.8	13.1	9.7	12.3	6.4	15.3	**0.1**	16.8	16.1
9. T. *velutinus*	15.4	13.7	16.0	14.3	15.6	14.8	16.1	14.7	**0.2**	15.9
10. *T. venustus*	14.1	15.9	14.0	14.3	14.6	14.1	12.5	14.2	14.0	**3.2**

[1] Off-diagonal table entries are mean pairwise distances among all sequences assigned to row and column taxa. Intraspecific distances are averaged over all pairwise comparisons of conspecific sequences; mean differences across the basal split within species (as reported in the text) are generally higher. K2P-corrected values were computed for comparison with previous studies of didelphid CYTB sequence variation (e.g., Patton et al. 1996, 2000), most of which used this model-based distance metric.

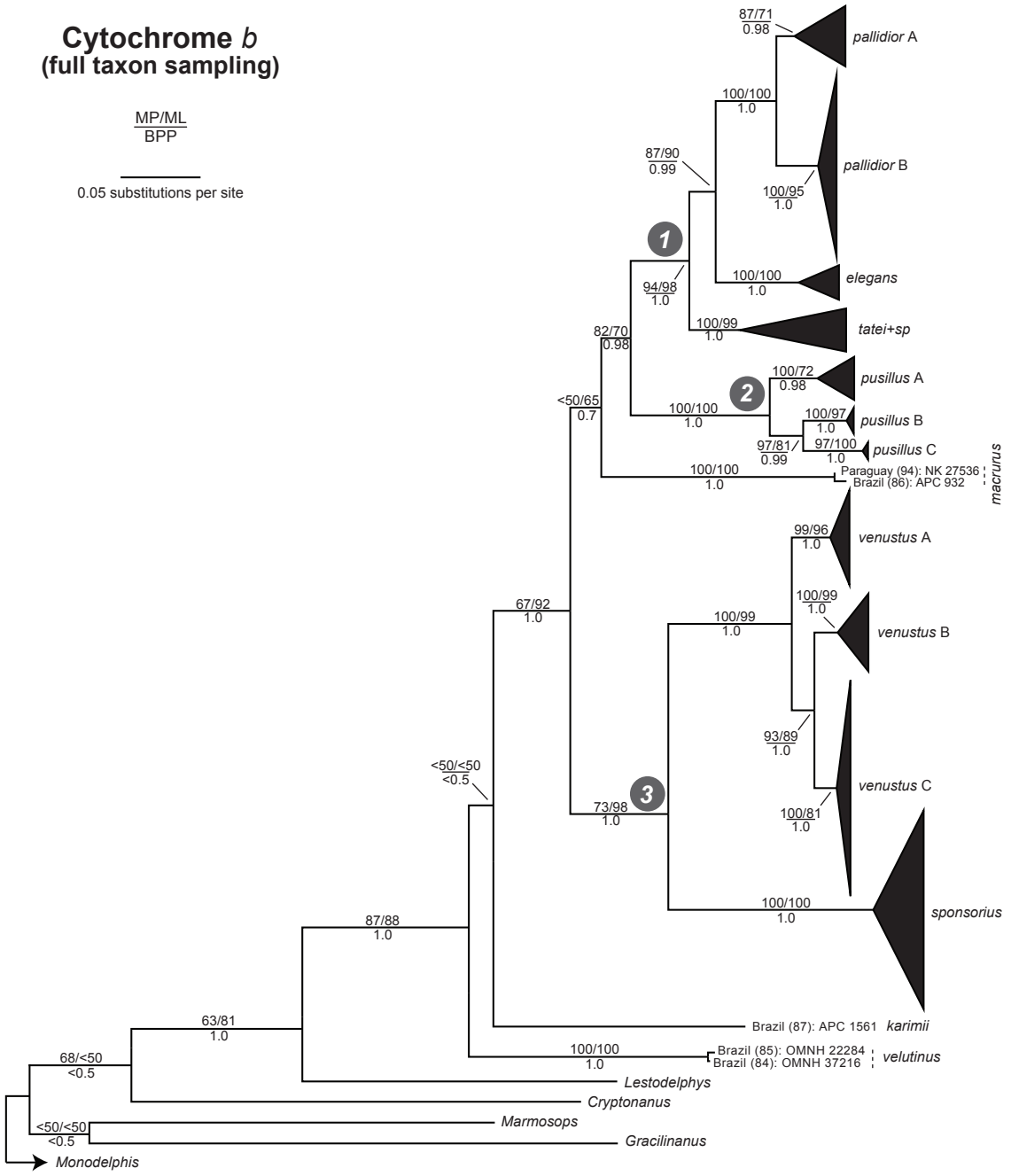

Figure 2. The tree based on ML analysis of cytochrome-b sequences for 131 ingroup terminals and six individuals representing five outgroup genera (GTR+I+Γ, ln-likelihood = -10723.24). Maximum parsimony (MP) and maximum likelihood (ML) bootstrap values and Bayesian Posterior Probabilities (BPP) are indicated at all nodes. Individual specimens of *Thylamys macrurus*, *T. karimii*, and *T. velutinus* are labeled with country of origin, locality number (in parentheses; see Appendix 2), and an alphanumeric specimen identifier (see Appendix 1). Other species (and haplogroups) are cartooned with black triangles, the breadth (vertical dimension) of which is proportional to the number of individuals sampled for each clade, and the length (horizontal extent) of which is proportional to genetic diversity within each clade. Clades numbered in gray circles are presented as subtrees in subsequent figures.

The first two lineages to branch within *Thylamys* are Brazilian endemic species—*Thylamys karimii* and *T. velutinus* (both sensu Carmignotto and Monfort 2006)—that are represented in these data by only one and two individual(s), respectively. Also sparsely sampled is *T. macrurus*, a species that inhabits eastern Paraguay and the adjoining Brazilian state of Mato Grosso do Sul (Carmignotto and Monfort 2006; Voss et al. 2009). Three clades recovered by all analyses of the CYTB data incorporate the remaining species: the Elegans Group (node 1), *Thylamys pusillus* (node 2), and the Venustus Group (node 3). The Elegans Group (**Figure 3**) consists of at least three species that collectively range from Patagonia to Peru (**Figure 4**). *Thylamys elegans,* a name that we apply in the restricted sense of Meynard et al. (2002) and Solari (2003), is represented in these data by five sequences from central Chile that form a robustly supported monophyletic group in all of our analyses. Despite the fact that *T. elegans* has a geographically restricted distribution, the uncorrected average sequence divergence among these haplotypes is moderately high, about 2.5% (**Table 3**). Sequence divergence across the basal split in this species is somewhat higher (3.8%), but it is far less than the 8.6% difference that Meynard et al. (2002) reported among their Chilean specimens. Although we are unable to account for this discrepancy (we sequenced only one of the specimens that Meynard et al. included in their study, NK 27583), we note that Braun et al. (2005) reanalyzed Meynard et al.'s *T. elegans* sequences and obtained an average divergence value of 1.94%.

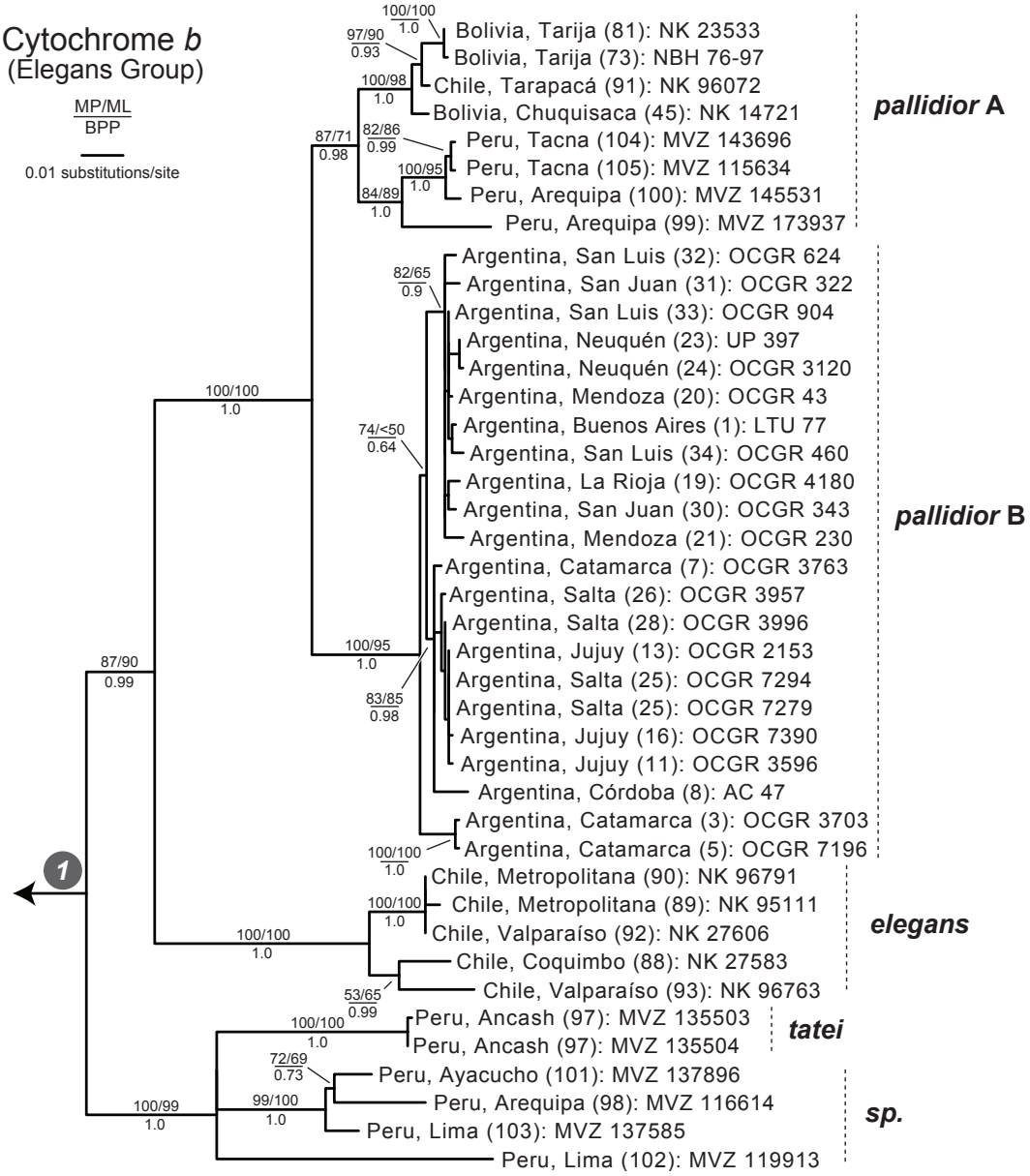

Figure 3. Relationships among 41 cytochrome-b sequences of the Elegans Group. This subtree shows the full details of relationships at node 1 (**Figure 2**) as resolved by ML analysis. Phylogenetic terminals in this tree are identified by country and state/department/province of origin, locality number (in parentheses; see Appendix 2), and an alphanumeric specimen identifier (see Appendix 1). Nodal support values are provided for all nodes recovered in common by MP, ML, and BI analyses.

A second species in the Elegans Group is *Thylamys pallidior,* a name that we apply to specimens from southern Peru, northern Chile, the Bolivian altiplano, and Argentina. The monophyly of *T. pallidior* is robustly supported in all of our analyses of

CYTB sequence data, but average uncorrected sequence divergence is moderately high (5.0%) across the basal split in this species, which separates two reciprocally monophyletic haplogroups. The northern haplogroup (*pallidior* A) occurs in Peru, Bolivia, and northern Chile and comprises two well supported subgroups differing by a net uncorrected p-distance of 2.5%. The southern haplogroup (*pallidior* B), although restricted to Argentina, is widely distributed in that country, where it occurs from high elevations (> 4000 m) in the Andes to near sea level in the Pampas; despite this impressive ecogeographic range, genetic diversity within *pallidior* B is low (about 1.0%) and there is little apparent phylogeographic structure among the 22 member sequences that we analyzed.

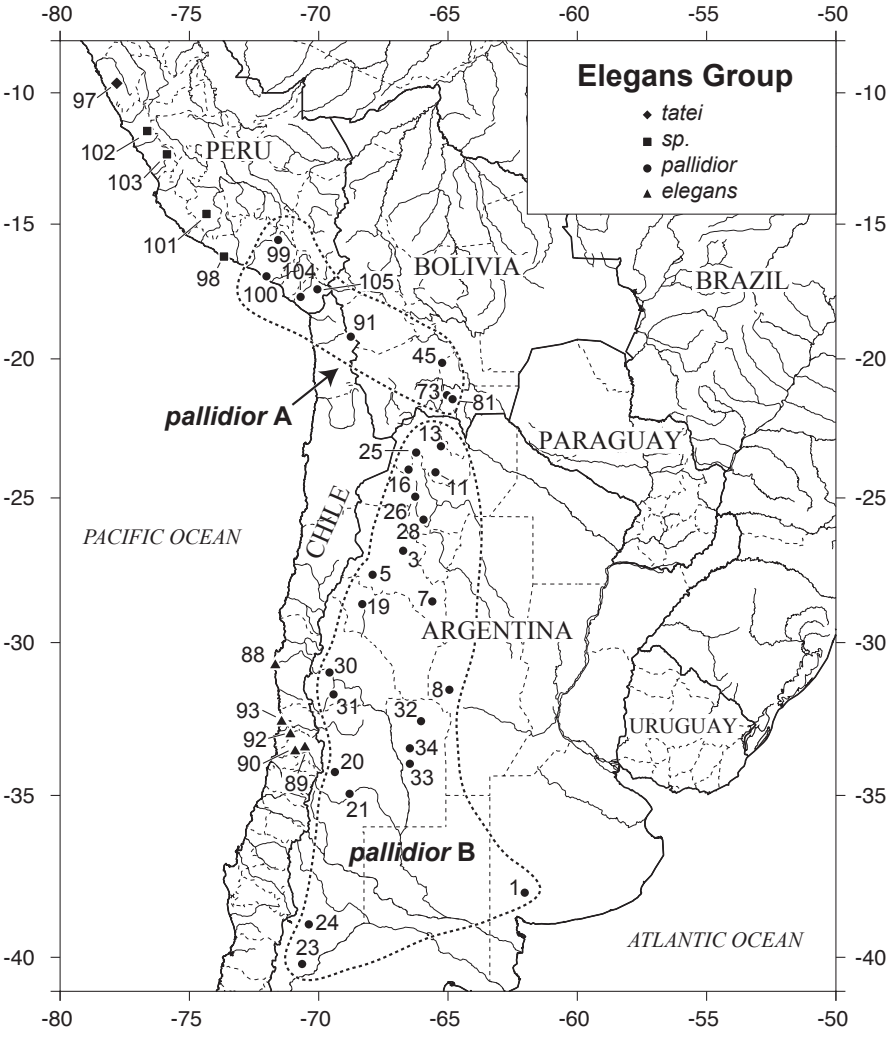

Figure 4. Map of collecting localities for specimens of the Elegans Group sequenced for this study. Numbers are keyed to entries in the gazetteer (Appendix 2).

The third species that we recognize in the Elegans Group is *Thylamys tatei*, represented by two individuals from west-central Peru (Ancash) that differ in sequence at only one nucleotide position. Interestingly, this genetically and morphologically distinctive taxon is part of an unresolved trichotomy with two other divergent lineages, also from western Peru, that may represent undescribed species (E. Palma, personal commun.). *Thylamys pusillus*, a name that we apply to three allopatric haplogroups (**Figure 5**), is a lowland species from the tropical and subtropical dry forests and savannas of southeastern Bolivia, western Paraguay, and northern Argentina (**Figure 6**). The three haplogroups recovered in our analysis correspond to those recently reported by Teta et al. (2009). *Thylamys pusillus* A is represented by six non-identical sequences from the Chaco Boreal (north of the Río Pilcomayo) that exhibit an average uncorrected intraclade distance of 1.9%. Genetic variation is substantially less within *pusillus* B from the Chaco Austral (south of the Pilcomayo, 0.6%) and *pusillus* C from the Mesopotamian region of northeastern Argentina (between the Paraná and Uruguay rivers, 0.4%). Although the reciprocal monophyly of all three haplogroups is robustly supported by our MP and Bayesian analyses, we note that our geographic samples are tightly clustered with large unsampled gaps between them, and that the monophyly of *pusillus* A is only moderately supported by ML bootstrapping.

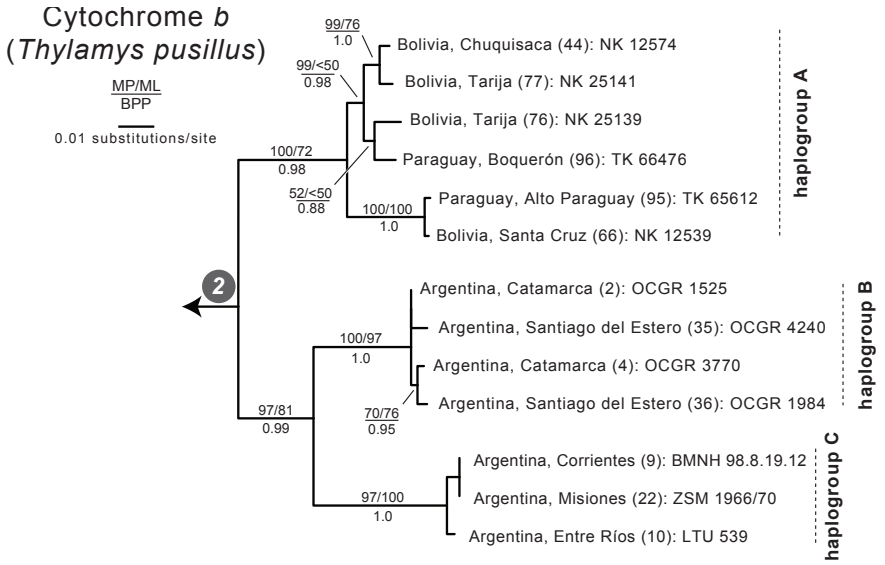

Figure 5. Relationships among 13 cytochrome-b sequences of *Thylamys pusillus*. This subtree shows the full details of relationships at node 2 (**Figure 2**) as resolved by ML analysis. Phylogenetic terminals in this tree are identified by country and state/department/province of origin, locality number (in parentheses; see Appendix 2), and an alphanumeric specimen identifier (see Appendix 1). Nodal support values are provided for all nodes recovered in common by MP, ML, and BI analyses.

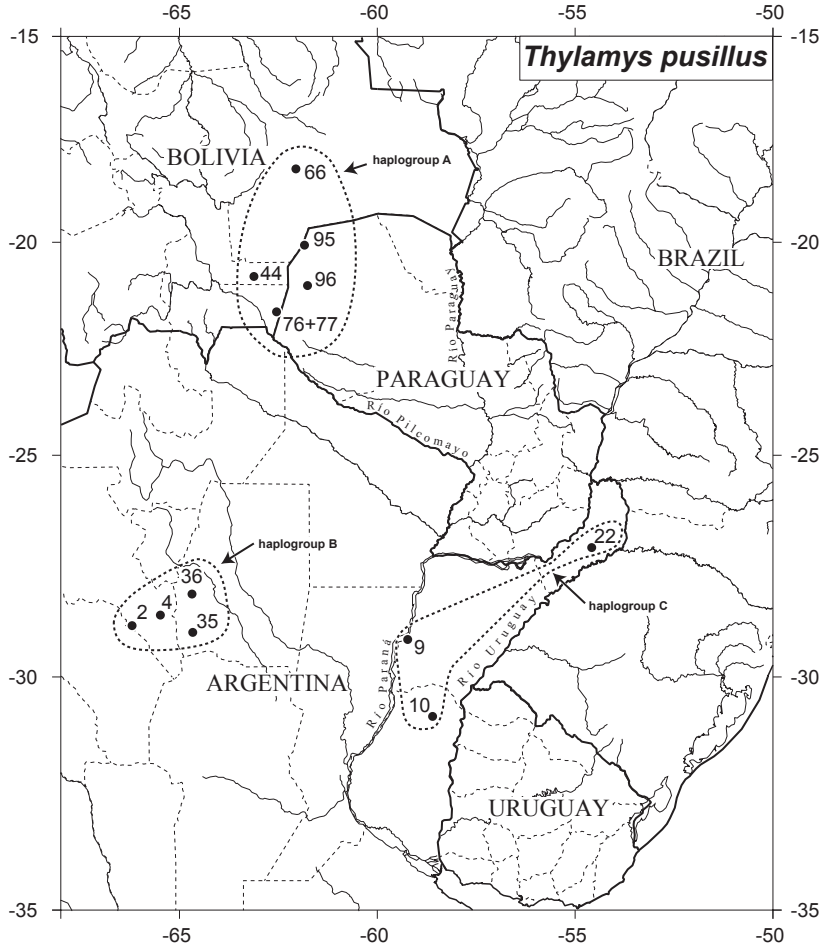

Figure 6. Map of collecting localities for specimens of Thylamys pusillus sequenced for this study. Numbers are keyed to entries in the gazetteer (Appendix 2).

The Venustus Group (**Figure 7**) includes two species, both of which are abundantly represented in our analyses by specimens collected on the eastern Andean slopes and foothills of Bolivia and northwestern Argentina (**Figure 8**) Although our analyses recovered the same basal mtDNA sequence dichotomy in this group previously reported by Braun et al. (2005), we use the name *sponsorius* for the species that they called *cinderella*. *Thylamys sponsorius* and *T. venustus* have partially overlapping distributions, but *T. sponsorius* ranges farther south into Argentina, and *T. venustus* ranges farther north into Bolivia. Where their latitudinal ranges overlap, *T. sponsorius* tends to occur at higher elevations than *T. venustus*, but allopatric populations of *T. venustus* in central Bolivia can occur at elevations as high as those at which *T. sponsorius* occurs further south. Mean intraspecific genetic diversity for *Thylamys sponsorius* is 1.8%, and specimens from Argentina form a strongly supported clade to the exclusion of conspecific Bolivian material, which appears as a shallowly branched paraphyletic group. By contrast, phylogeographic structure is much more apparent within *T. venustus*, which includes three robustly supported and spatially segregated haplogroups (**Figure 9**). *Thylamys venustus* A is represented by 12 sequences (with an average intraclade divergence value of 1.2%) from seven localities at high elevations (2450–4000 m) in central Bolivia, where voucher specimens were collected in the departments of La Paz, Chuquisaca, and Cochabamba. *Thylamys venustus* B is represented by 10 sequences (with an average intraclade divergence value of 1.1%) from seven localities at somewhat lower elevations (1695–2950 m) in eastern Cochabamba and western Santa Cruz. *Thylamys venustus* C is represented by 26 sequences (with an average intraclade divergence of only

0.6%) from 18 localities in foothill and middle elevations (349–2100 m) in western Santa Cruz southward to Argentina (Jujuy and Tucumán).

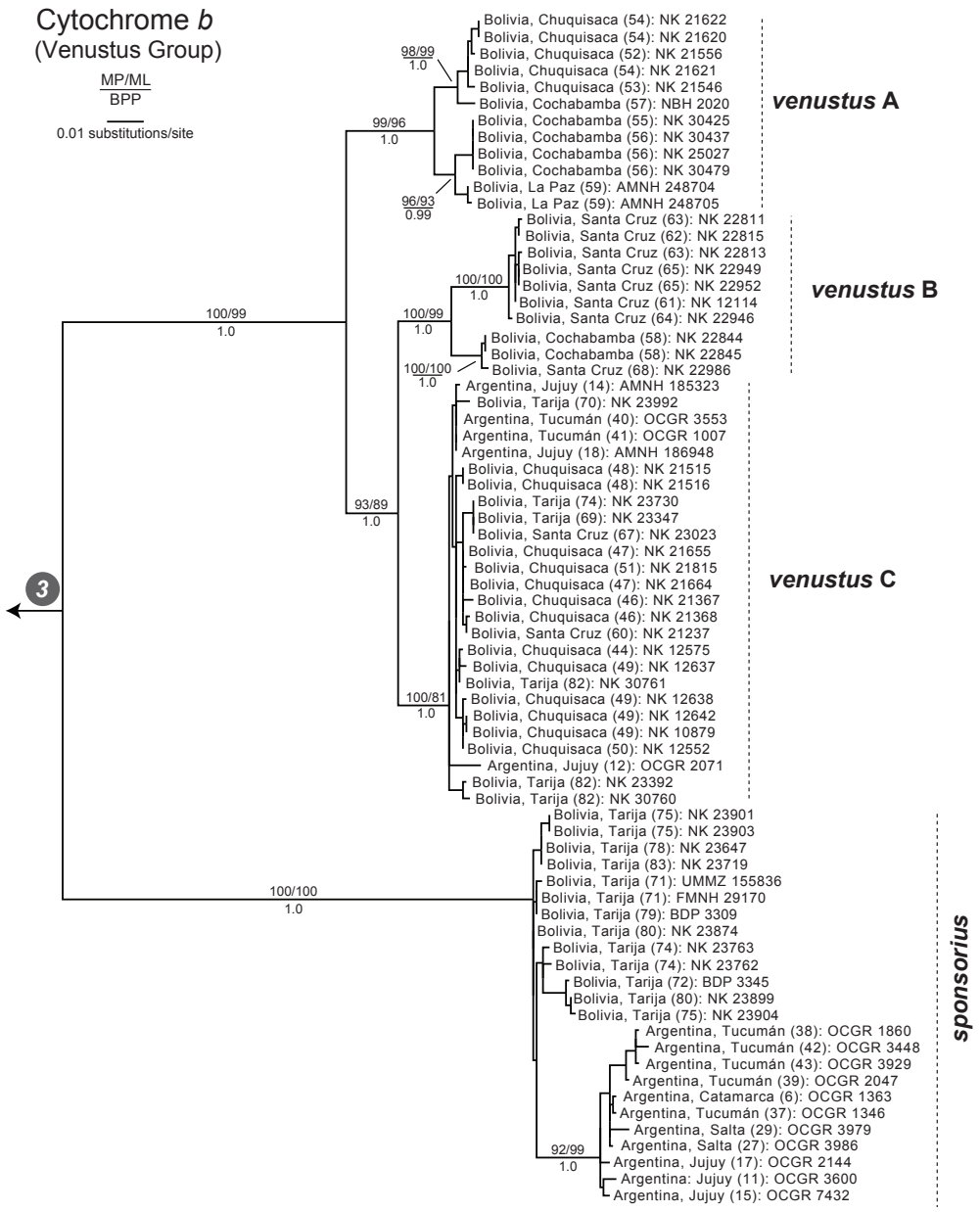

Figure 7. Relationships among 72 cytochrome-*b* sequences of the Venustus Group. This subtree shows the full details of relationships at node 3 (**Figure 2**) as resolved by ML analysis. Phylogenetic terminals are identified by country and state/department/province of origin, locality number (in parentheses; see Appendix 2), and an alphanumeric specimen identifier (see Appendix 1). Nodal support values are provided for all nodes recovered in common by MP, ML, and BI analyses.

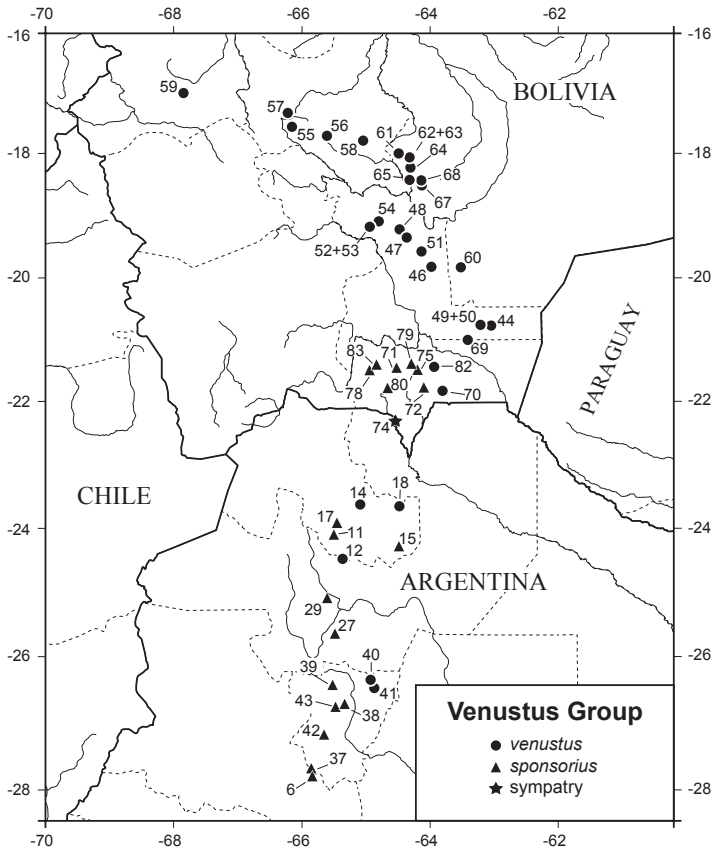

Figure 8. Map of collecting localities for specimens of the Venustus Group sequenced for this study. Numbers are keyed to entries in the gazetteer (Appendix 2).

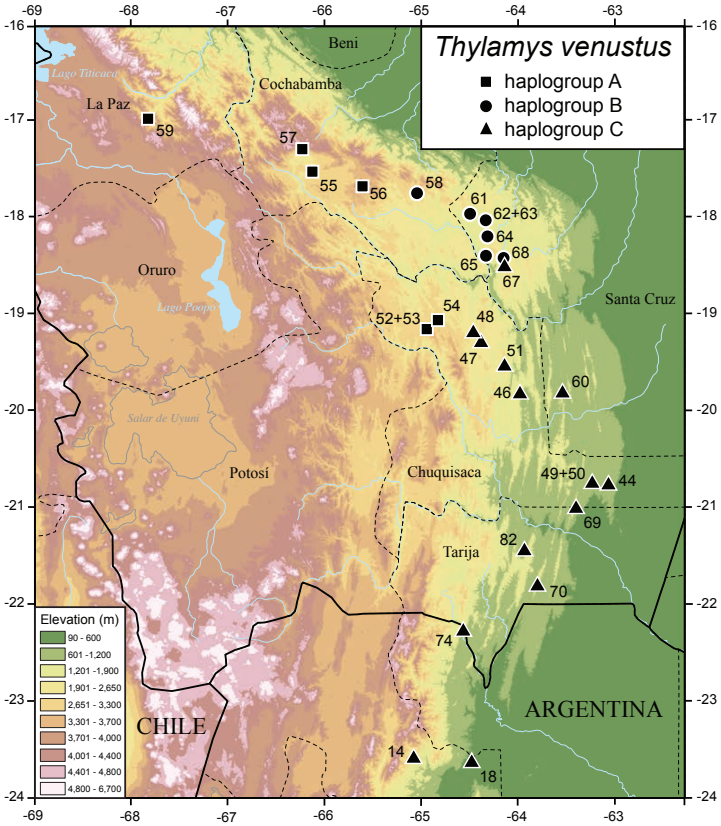

Figure 9. Map of collecting localities for specimens from each of the three *Thylamys venustus* haplogroups from Bolivia and northern Argentina (not all Argentine collecting localities of haplogroup C are shown). Numbers are keyed to entries in the gazetteer (Appendix 2).

Phylogenetic relationships based on cytochrome b. In addition to resolving species-level clades and intraspecific phylogeography as described above, our analyses of cytochrome-*b* sequence data provide noteworthy support for some deeper nodes. Within the Elegans Group, for example, there is strong support for a sister-group relationship between *T. elegans* and *T. pallidior*, and for a group that includes *T. tatei* and two unnamed Peruvian lineages (**Figure 3**). Additionally, there is strong support for the monophyly of the genus *Thylamys*, and moderate to strong support for the monophyly of a group that includes all of the species of *Thylamys* except *T. velutinus* and *T. karimii* (**Figure 2**). However, other relationships of key interest are only moderately to weakly supported by these data. Among such equivocally resolved details are the problematic relationships of three Brazilian species (*T. velutinus, T. karimii,* and *T. macrurus*) and the sister-group relationship between the Elegans and Pusillus clusters.

Phylogenetic analyses of concatenated mitochondrial sequences. Because the cytochrome-*b* dataset was unable to convincingly resolve several key interspecific relationships, we obtained ND2 and COX2 sequences from exemplars of each of the species and haplogroups described above, choosing specimens that spanned the basal split recovered for each clade in the CYTB tree. Separate analyses of the ND2 and COX2 datasets using MP, ML, and BI methods recovered topologies that were similar to those obtained from CYTB in most respects. Importantly, relationships that were strongly supported by one gene never conflicted with relationships that were strongly supported by another. This lack of hard incongruence among datasets suggests that combining these datasets is appropriate.

Maximum parsimony, maximum likelihood, and Bayesian analyses of the combined-gene (CYTB + ND2 + COX2) dataset recovered most of the same relationships obtained from cytochrome *b* and provide more robust support for clades that were only weakly or moderately supported by individual-gene analyses (**Figure 10**). Novel features of this topology include (1) a weakly supported basal clade that unites the Brazilian species *Thylamys velutinus* and *T. karimii*, and (2) a strongly supported sister-group relationship between *T. macrurus* and the Venustus Group (*T. sponsorius* + *T. venustus*). Additionally, most of the supraspecific nodes with moderate to strong support from CYTB (**Figure 2**) are consistently strongly supported by the combined-gene analyses. Among the previously described supraspecific relationships that seem conclusively resolved in these results include the monophyly of the Elegans and Venustus groups, a sister-group relationship between the Elegans Group and *Thylamys pusillus*, and a clade that includes all of the species of *Thylamys* except *T. velutinus* and *T. karimii*. Additionally, all of the species and haplogroups delineated using CYTB were recovered with very strong support in the combined-gene analyses. The single noteworthy example of a CYTB clade with diminished support in these results is *pusillus* B + *pusillus* C, which still receives strong support from MP bootstrapping, but for which ML bootstrap support and BI posterior probabilities have substantially eroded.

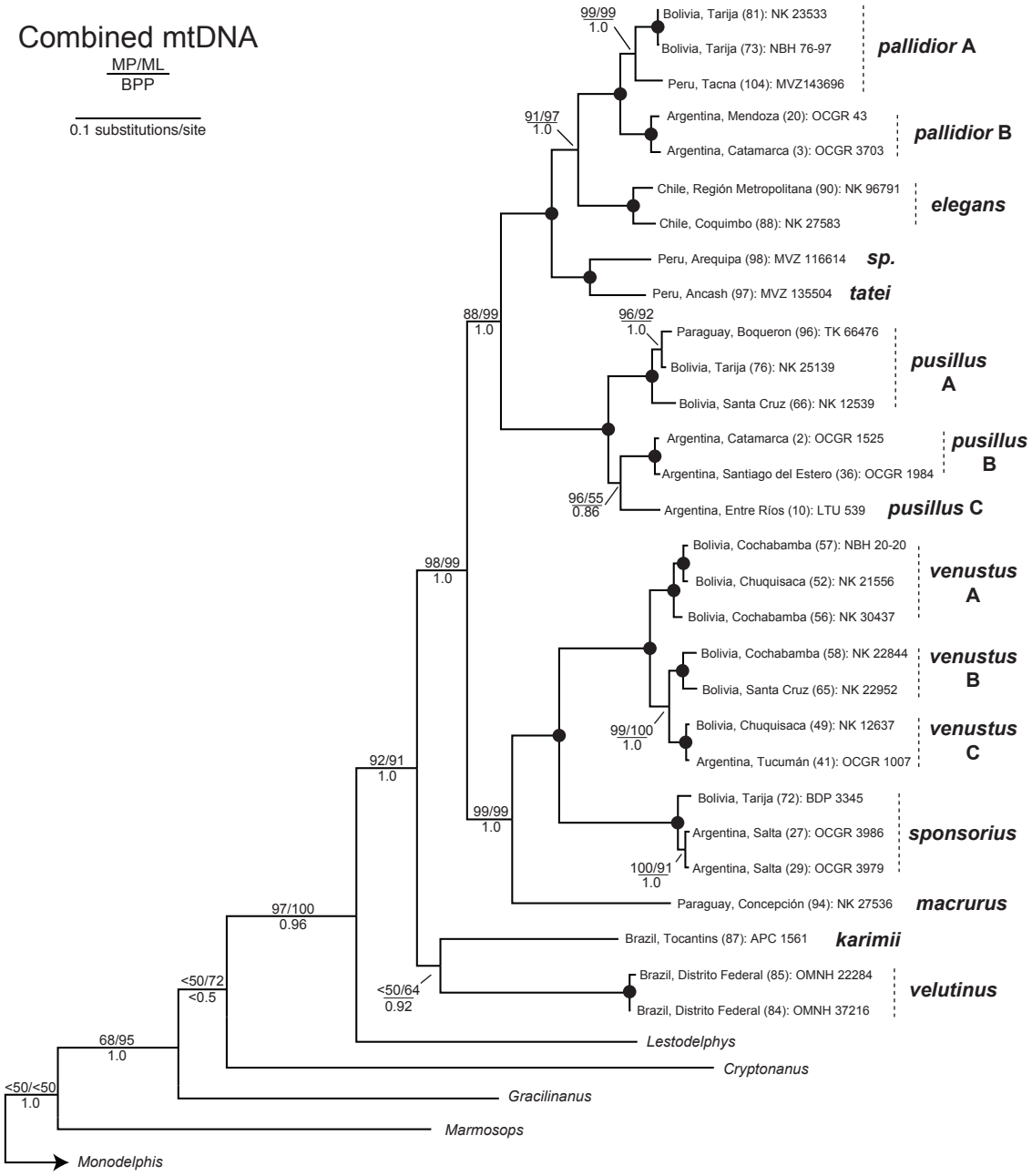

Figure 10. The maximum likelihood tree based on concatenated sequence data from cytochrome b, cytochrome c oxidase subunit II, and NADH dehydrogenase 2 from 29 specimens representing nine species of Thylamys and six individuals representing five outgroup genera (GTR+I+□, ln-likelihood = -22234.43). Phylogenetic terminals are identified by country and state/department/province of origin, locality number (in parentheses; see Appendix 2), and an alphanumeric specimen identifier (see Appendix 1). Black circles indicate nodes with 100% ML and MP bootstrap support and Bayesian posterior probabilities equal to 1.0.

Topology tests. Our combined-gene dataset provides a large sample of characters (2877 bp) for evaluating alternative hypotheses of phylogenetic relationships. Of these, only three are sufficiently explicit to merit statistical testing: (1) Palma et al. (2002) analyzed cytochrome-*b* sequence data from five species of *Thylamys* and recovered the topology (*macrurus* (*pusillus* (*venustus* (*elegans* + *pallidior*)))), (2) Braun et al. (2005) analyzed cytochrome-*b* sequence data from seven species and recovered the topology (*macrurus* ((*venustus* + *sponsorius*) (*pusillus* (*pallidior* (*elegans* + *tatei*))))), and (3) Carvalho et al. (2009) analyzed cytochrome-*b* sequence data from 8 species and recovered the topology (*macrurus* ((*sponsorius* + *venustus*) (*karimii* (*pusillus* (*pallidior* (*tatei* + *elegans*)))))). All of these hypotheses can be confidently rejected using Shimodaira and Hasegawa's (1999) likelihood-ratio test with our combined-gene data ($P = 0.00$ for all alternatives).

Although several possible explanations might account for the rejected topologies of Palma (2002), Braun et al. (2005), and Carvalho et al. (2009), we noticed that the most striking discrepancies between their results and ours involve only two sequences. One sequence has GenBank accession number AF431926 and was said to have been obtained from a Paraguayan specimen of *Thylamys macrurus* (field number NK 27536). The second is AF431927, said to have been obtained from a Bolivian specimen of *T. pusillus* (NK 25139). Both sequences were originally obtained and analyzed by Palma et al. (2002), but Braun et al. (2005) and Carvalho et al. (2009) subsequently used AF431926 (downloaded from GenBank) in their analyses. Because we independently extracted and amplified cytochrome *b* from NK 27536 (*T. macrurus*) and NK 25139 (*T. pusillus*), we were able to compare our sequences with those obtained by Palma et al. (2002). Sequence

alignment indicates that both of our *T. macrurus* sequences (from specimens NK 27536 and APC 932) are much more similar to each other (p-distance = 0.7%) than either is to AF431926 (average p-distance = 21.23%). A BLAST comparison of AF431926 against other sequences in GenBank reveals that its closest match to previously published data (87%) is a cytochrome-*b* sequence from *Marmosops impavidus* (U34670). A BLAST search on the purported *T. pusillus* sequence (AF431927) clearly demonstrates that it does not belong to a marsupial; instead, it is 95% identical to several GenBank sequences from the cricetid rodent *Phyllotis xanthopygus* (e.g., AY431053).

Parsimony ancestral state reconstruction. Ancestral areas for lineages that comprise the Elegans Group (the only clade that occurs west of the Andes) were inferred across the combined-gene ML topology using the parsimony criterion (**Figure 11**). Results indicate three Andean dispersal events: (1) the ancestral Elegans Group lineage dispersed from east to west and diversified along the western coast of Chile and Peru, (2) the lineage leading to *Thylamys pallidior* B later dispersed from west to east, forming a wide-ranging clade distributed in a variety of habitats in Argentina, and (3) a subclade of *T. pallidior* A invaded the altiplano of Bolivia and Chile (along the crest of the Andes, and therefore of ambiguous classification with respect to our cis/trans dichotomy).

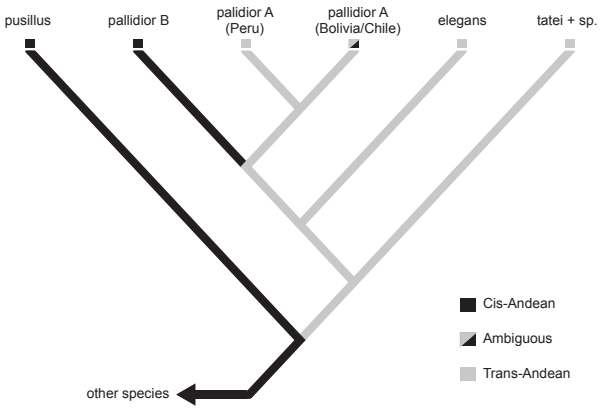

Figure 11. Parsimony reconstruction of ancestral distributions in relation to the Andean cordillera. Branch tips are coded as east of the Andes (cis-Andean), west of the Andes (trans-Andean), or both (ambiguous). The topology used for this reconstruction is derived from **Figure 10**.

DISCUSSION

This report and several other recent morphological and molecular studies (especially Solari 2003; Braun et al. 2005; Carmignotto and Monfort 2006), convincingly refute Hershkovitz's (1959) hypothesis that all of the taxa currently referred to *Thylamys* are conspecific. Contrary to his assessment (which was not based on any substantive analysis of character data), the existence of multiple species is clearly indicated by diagnostic morphological differences and high sequence divergence (up to 16% in uncorrected pairwise cytochrome-*b* comparisons; **Table 3**) among numerous robustly supported mitochondrial clades. Whereas some of the species recognized as valid in this report are allopatric, others are known to occur sympatrically in various combinations (**Table 4**), providing additional support for the notion that these are evolutionarily independent (reproductively isolated) lineages.

Table 4. Documented Cases of Sympatry Among Species of *Thylamys*

Locality	Sympatric Species
Argentina: Jujuy, 9 km NW Barcena	*T. pallidior* (OMNH 29963) and *T. sponsorius* (OMNH 29974)
Argentina: Mendoza, "Nacunan Reserve"	*T. pusillus* (UWBM 72205) and *T. pallidior* (UWBM 72195)
Bolivia: Chuquisaca, 3.8 km E Carandaytí	*T. pusillus* (AMNH 261268, MSB 55846) and *T. venustus* (AMNH 261245)
Bolivia: Tarija, 3 km SE Cuyambuyo	*T. sponsorius* (AMNH 275448–275450, MSB 67014) and *T. venustus* (AMNH 275436, 275447; NK 23762)
Bolivia: Tarija, 8 km S & 10 km E Villa Montes	*T. pusillus* (AMNH 246442–246450) and *T. venustus* (AMNH 246451)

Indeed, many of the vexing problems that previously hampered taxonomic understanding of the genus are now resolved beyond any reasonable doubt. For example, our results unambiguously support Palma and Yates' (1998) hypothesis that *T. venustus*

is a distinct species from *T. elegans*, Carmignotto and Monfort's (2006) conclusion that *T. karimii* is a distinct species from both *T. pusillus* and *T. velutinus*, and Solari's (2003) treatment of *T. tatei* as a distinct species from *T. elegans*. On the other hand, several problematic issues remain, among them the probable existence of unnamed forms of the Elegans Group in western Peru, the taxonomic status of allopatric haplogroups that we synonymize with *T. pusillus*, and the morphological diagnosability of geographically overlapping sister taxa that we recognize as *T. venustus* and *T. sponsorius* (Giarla et al. 2010). In our opinion, these enigmas are unlikely to be resolved to anyone's satisfaction until additional data (e.g., better geographic sampling, nuclear-gene sequences, new morphological characters) become available.

Despite the fact that a few key nodes remain weakly supported, phylogenetic relationships among species of *Thylamys* are completely resolved by the mitochondrial sequence data analyzed in this report (**Figure 10**). Because this is the first phylogenetic study to include all of the valid species of *Thylamys*, our results provide a uniquely appropriate framework for biogeographic inference. Although a formal analysis of ecogeographic trait evolution in the genus will be provided elsewhere, some preliminary inferences are appropriately mentioned here.

The first concerns the constraining role of the Andes, the principal topographic feature of the South American continent. All of the closest outgroups to *Thylamys* (e.g., *Chacodelphys* and *Lestodelphys*; see Voss and Jansa 2009) occur east of the Andes, as does the subgenus *Xerodelphys*, *T. (Thylamys) macrurus*, the entire Venustus Group, and *T. (T.) pusillus,* so the most recent common ancestor of *Thylamys* was almost certainly also cis-Andean. Additionally, our phylogenetic topology suggests that that the Andes

were initially traversed by the ancestral lineage of the Elegans Group, and that a single lineage of *T. pallidior* subsequently reinvaded the eastern lowlands (**Figure 11**). The almost complete absence of phylogeographic structure in haplogroup B of *T. pallidior* (**Figure 3**) suggests that its cis-Andean range expansion was both rapid and relatively recent.

Among the species of *Thylamys* that occur east of the Andes, our results do not support the east-to-west speciation sequence postulated by Palma et al. (2002). Although Palma et al.'s hypothesis was consistent with a recovered phylogenetic topology that showed a clear east-to-west progression among the five species included in their analysis (*T. macrurus* (*T. pusillus* (*T. venustus* (*T. pallidior* + *T. elegans*)))), no such progression is discernable in our more densely taxon-sampled trees (**Figure 2**, **Figure 10**) which, correspondingly, suggest more complex biogeographic scenarios. A key factor in developing new scenarios that more accurately depict relevant processes (dispersal, vicariance, speciation, and extinction) in a meaningful geographic context is the improved resolution of species distributions enabled by recent revisionary work.

Whereas most published range maps suggest that *Thylamys* species are widely distributed and eurytopic, the pattern that is now beginning to emerge from the fog of past misidentifications and erroneous synonymies is that at least some species are narrowly restricted to one or a few macrohabitats. For example, the geographic range of *T. pusillus*—formerly depicted as encompassing the Caatinga, Cerrado, Chaco, Monte Desert, Pampas, and Yungas (e.g., by Brown 2004: Figure 88)—now appears to be largely restricted to the Chaco (**Figure 12**). In effect, revised specimen-based

identifications have dramatically improved the correspondence between species' geographic ranges and macrohabitat distributions in most cases.

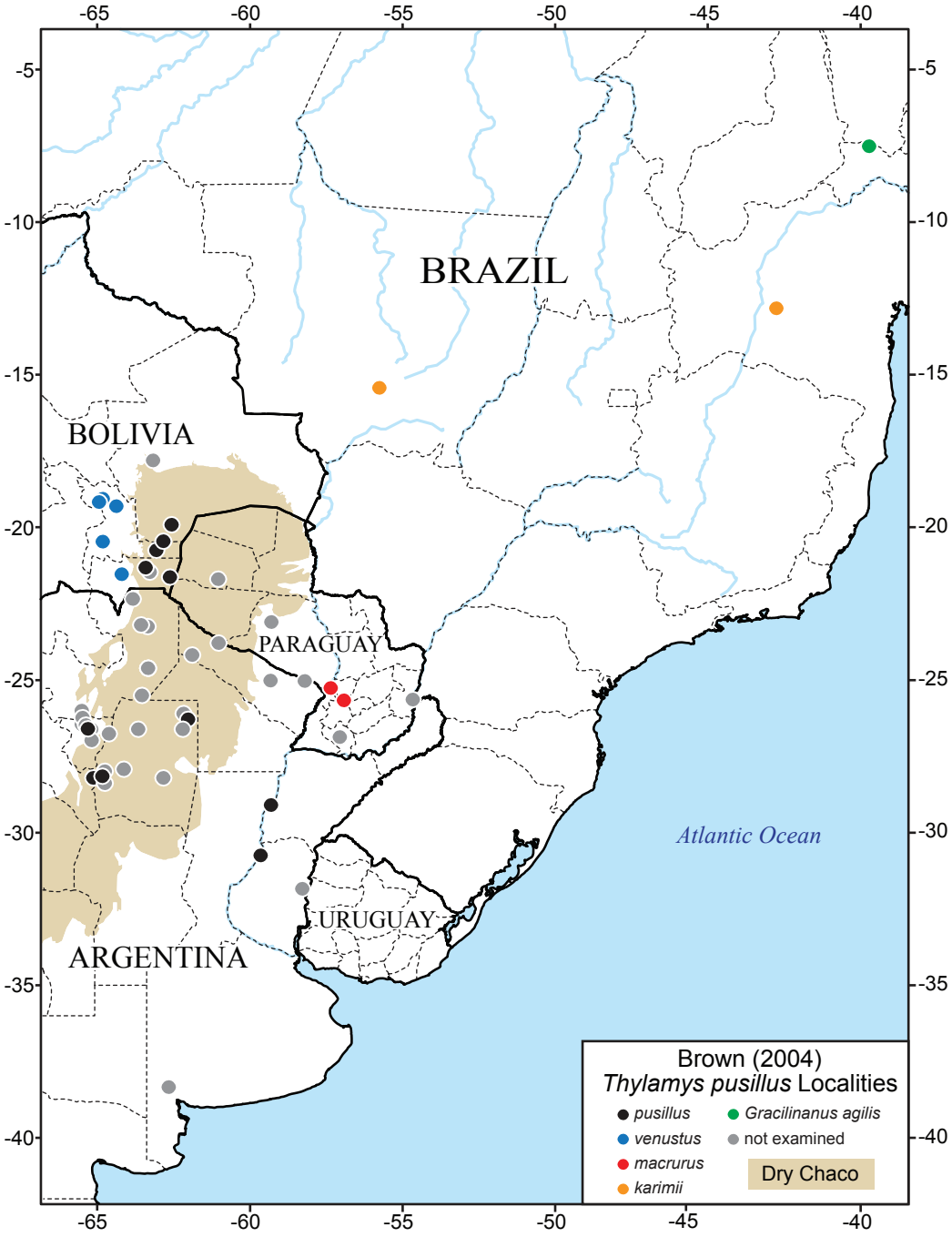

Figure 12. *Thylamys pusillus* localities mapped by Brown (2004) with reidentifications based on our examination of museum specimens. The "Dry Chaco" ecoregion is coded according to Olson et al. (2001).

CHAPTER 2

HIDDEN DIVERSITY IN THE ANDES: COMPARISON OF SPECIES DELIMITATION METHODS IN MONTANE MARSUPIALS

INTRODUCTION

The debate regarding species concepts is ongoing and shows no sign of abating (reviewed by Hausdorf 2011), but at least one uncontroversial idea has emerged: speciation is an ongoing process, and contemporary populations of organisms represent evolutionary lineages at different stages of distinctiveness (De Queiroz 2007). For some organisms, the delineation of independent evolutionary lineages seems straightforward due to the presence of diagnostic morphological characters, which can be expected to evolve either by genetic drift after a long history of isolation or by divergent selection on readily observable traits (Lande 1976). However, for many organisms—perhaps especially animals that do not rely on visual mating cues (Bickford et al. 2007)— species recognition based solely on morphological differences can be problematic if diagnostic differences are subtle, occur in commonly overlooked anatomical structures (e.g., soft tissues), or simply are not present.

A proliferation of cryptic diversity in a given area could be caused by geographical or ecological factors unique to that region. For example, mountainous areas might harbor more cryptic species than lowland areas because rugged terrain and altitudinal zonation of habitats could limit dispersal and thus create more opportunities for allopatric speciation, especially in combination with dynamic climate change regimes (Roy 1997; Weir 2006; Kozak and Wiens 2006; Brumfield and Edwards 2007; Ribas et

al. 2007). In particular, if the ecological niches of allopatric montane populations remain similar (Wiens and Graham 2005), reliable phenotypic indicators of lineage divergence may not evolve. In situations like this, morphological traits may not be sufficient to detect recently diverged (or incipiently diverging) species.

Advances in DNA sequencing technologies over the past twenty years have allowed systematists to uncover cryptic genetic diversity at a rapid pace, challenging all researchers who study biological processes at or above the species level to contend with unexpected numbers of putative lineages and candidate species (Beheregaray and Caccone 2007; Bickford et al. 2007). Mitochondrial DNA (mtDNA) markers have been especially effective for revealing cryptic genetic diversity within animal systems because of the mitochondrial genome's high mutation rate and rapid coalescence time (Moore 1995). However, the exclusive use of mtDNA in phylogenetic and phylogeographic work has been criticized extensively (Ballard and Whitlock 2004; Edwards and Bensch 2009; Galtier et al. 2009), primarily because single gene trees can deviate from historical patterns of population branching due to the stochastic nature of coalescence (gene-tree species-tree incongruence; Degnan and Rosenberg 2009). Nonetheless, studies that include mtDNA are still common because mtDNA is typically quite variable for most animal taxa, is easy to amplify and sequence, and may be a "leading indicator" of isolation (Zink and Barrowclough 2008). Many studies that included both mtDNA and multiple unlinked nuclear DNA (nDNA) markers have shown strong cytonuclear discordance (reviewed by Toews and Brelsford 2012). The discordance observed between trees inferred from mtDNA and nDNA could be caused by phylogenetic error, incomplete lineage sorting, or lateral gene transfer (Funk and Omland 2003), so studies

based on both types of data are necessary to disentangle the evolutionary processes that resulted in observed patterns.

The purpose of this study is to evaluate the evolutionary independence of morphologically cryptic mtDNA haplogroups within a genus of Neotropical marsupials based on data from 15 nuclear loci and one mitochondrial locus. Didelphid marsupials of the genus *Thylamys* are found in a variety of ecoregions in central and southern South America, but they primarily inhabit arid and semiarid open habitats. Three species—*T. pallidior*, *T. sponsorius*, and *T. venustus*—are principally montane (with some populations inhabiting areas up to 4000 meters above sea level in the Andes) and occur in Peru, Bolivia, Chile, and Argentina. In a prior study, we identified multiple allopatric mtDNA haplogroups within each of these species, but we were unable to find morphological characters that could consistently distinguish haplogroups (Giarla et al. 2010). Nevertheless, the deep mtDNA divergence we observed (2.5% to 5.35% sequence divergence among haplogroups at the cytochrome *b* locus) suggests that up to seven independent lineages might be present within this complex (Giarla et al. 2010).

A variety of methods have recently been proposed to assess the evolutionary independence of putative lineages using DNA sequence data; here, we focus on two of the more widely used approaches. Bayesian Phylogenetics and Phylogeography (BPP; Yang and Rannala 2010) provides a Bayesian approach to "species delimitation" in which both phylogenetic uncertainty and stochastic lineage coalescence are taken into account when testing predefined splits on a single proposed species tree. SpeDeSTEM (Ence and Carstens 2011) is a software pipeline based on the multilocus species tree method STEM (Kubatko et al. 2009), in which various permutations and combinations of

subpopulations within putative species are assessed for species status using previously estimated gene trees. Tests of "species limits" (in effect, genetic isolation) using both of these coalescent-based approaches have the potential to validate or refute cryptic lineage diversity that only receives weak support from traditional morphological approaches or that depends on single-gene phylogenetic analyses. Here, we use both approaches to examine the coalescent history of putative lineages within *T. pallidior*, *T. sponsorius,* and *T. venustus* in order to assess their evolutionary independence. In addition to the evolutionary implications of our results, various aspects of the lineage recognition process were considered and tested in this study, including: (1) Do different coalescent-based "species validation" approaches yield similar results based on the same data? (2) How does the choice of priors affect Bayesian lineage recognition? (3) How many loci are necessary to confidently recognize distinct lineages? And (4), to what extent do partitions based on mtDNA haplotype membership correspond to real evolutionary units?

MATERIALS AND METHODS

Nuclear marker design

Primer pairs for nearly 40 anonymous nuclear loci were developed for testing within *Thylamys* species. A genomic library of DNA fragments ranging in length from 500 to 1500 bp was developed following a modified version of the genomic library generation protocol of Glenn and Schabel (2005). First, whole genomic DNA was extracted from tissue sample NK22949 (*Thylamys venustus*) using a DNeasy Blood and Tissue Kit (Qiagen Inc.). Genomic DNA was digested with restriction enzymes *Xmn1* and *Rsa* and run on an agarose gel. Portions of the gel corresponding to fragment lengths between 500 and 1500 bp were excised and purified using a QIAquick Gel Extraction Kit

(Qiagen Inc.). Double-stranded linkers were ligated to the size-selected fragments, and PCR was used to amplify the fragments. The resulting amplified fragment library was cloned into *E. coli* cells using a pGEM-T Vector System (Promega Inc.). After growing overnight, dozens of bacterial colonies containing inserts were picked with sterile toothpicks and immediately added to a PCR reaction mixture (12.5 μL GoTaq Green Master Mix [Promega Inc.], 1.0 μL of 10 μM M13F primer solution, 1.0 μL of 10 μM M13R primer solution, and 10.5 μL water) for colony PCR (5 min. of initial melting at 95°; followed by 35 cycles of melting at 95° for 30 sec., annealing at 55° for 30 sec., and extension at 72° for 1.5 min.; and a final extension for 3 min. at 72°). PCR products were run on an agarose gel, with 700–1000 bp fragments preferentially selected for further development. Selected PCR products were cleaned using Exonuclease I and Shrimp Alkaline Phosphatase (Hanke and Wink 1994) and sequenced in both directions on an ABI 3730 at the University of Minnesota's Biomedical Genomics Center.

Sequences were assembled in Geneious (Drummond et al. 2011) and cloning vector regions were trimmed. For 35 genomic regions, forward and reverse primers were designed using the Primer3 (Rozen and Skaletsky 2000) software plug-in in Geneious with the goal of selecting primers that would amplify products between 500 and 800 bp long. All primers were designed to be approximately 30 bp long in order to achieve a higher amplification success rate and fewer instances of non-specific amplification (Belfiore 2011). Primers were tested on individuals from each of the three taxonomic species under consideration in this study: *Thylamys pallidior*, *T. sponsorius*, and *T. venustus*. If amplification was successful, PCR products were sequenced and aligned to determine if variation was present across the individuals tested. Exemplar sequences from

each of the markers developed were used as queries in a BLAST search against GenBank's nucleotide database to determine if the sequence contained protein-coding regions. In addition, a BLAT search (BLAST-Like Alignment Tool; Kent 2002) was used to identify the number and location of hits within the whole-genome sequence of the didelphid marsupial *Monodelphis domestica* (Mikkelsen et al. 2007). Some BLAT searches closely matched to multiple regions in the *M. domestica* genome, suggesting that the sequences could have one or more paralogs that could ultimately be co-amplified if the marker were to be developed further. BLAST searches revealed that a small number of sequences appeared to contain protein-coding regions. Both protein-coding sequences and sequences that might have closely related paralogs were eliminated from the pool of potential markers. Of the 35 primer pairs designed, 14 sets were ultimately chosen for inclusion in this study based on product length, ease of amplification, and presence of variation (Appendix 4).

Two other markers with smaller effective population sizes (and thus, on average, shorter coalescence times) than autosomal nDNA markers were included in analyses: the mitochondrial protein-coding gene cytochrome *b* (CYTB) and an intron within the X-linked gene OGT. CYTB sequences for the individuals of interest were compiled from a previous study (Giarla et al. 2010). The X-linked intron marker OGT was developed by downloading a subset of single-copy gene sequences from the Ensembl Genome Project (www.ensembl.org) for the *Monodelphis domestica* X chromosome and sorting by intron size. After identifying the ideally sized ~600 bp intron between exons 8 and 9 of the *M. domestica* OGT gene, primers were designed from conserved portions of the flanking exon sequences that aligned across *M. domestica* and *Homo sapiens* (Appendix 4).

Sampling of loci and individuals

Fifteen nuclear markers (14 anonymous loci and the X-linked intron OGT) and one mitochondrial marker (CYTB) were sampled for multiple individuals within *Thylamys sponsorius*, *T. pallidior*, and *T. venustus*, three species that occur in adjacent Andean biomes of northern Argentina, Bolivia, northern Chile, and southern Peru (**Figure 13**; Appendix 5). In a previous analysis of mitochondrial DNA sequence variation, Giarla et al. (2010) observed multiple morphologically undifferentiated mtDNA haplogroups within each of these Andean species. Two haplogroups were identified within *T. pallidior* and *T. sponsorius* (designated as "A" and "B") and three were identified within *T. venustus* (designated as "A", "B", and "C"). In all but one case (*T. sponsorius* A, but see Results), mitochondrial haplogroups were monophyletic and well-supported by all nodal metrics (Giarla et al. 2010). Without exception, all haplogroups assigned to the same species are allopatric, although the geographic ranges of haplogroups *T. venustus* B and *T. venustus* C are closely juxtaposed near Vallegrande in western Santa Cruz department, Bolivia (Giarla et al. 2010). Limited by the availability of high-quality tissue samples, between 4 and 10 individuals were chosen for inclusion in this study from within each haplogroup, for a total of 60 individuals (Appendix 5).

PCR conditions were optimized for each set of primers, resulting in variation among annealing temperatures and reagent concentrations. A typical PCR mixture contained 7.5 μL GoTaq Green Master Mix (Promega Inc.), 0.5 μL of each 10 μM primer solution, and 5.5 μL water. A typical thermocycler protocol consisted of 2 min. of initial melting at 95° followed by 35 cycles of melting at 95° for 30 sec., annealing at an optimized temperature for 30 sec., and extension at 72° for 1 min., and ending with a

single final extension for 7 min. at 72°. Problematic amplifications were re-run using Platinum Taq DNA polymerase (Life Technologies Corp.) with varying Taq concentrations. PCR products were cleaned and sequenced following protocols described above. Chromatograms were compiled and edited in Sequencher 4.7 (Gene Codes Inc.).

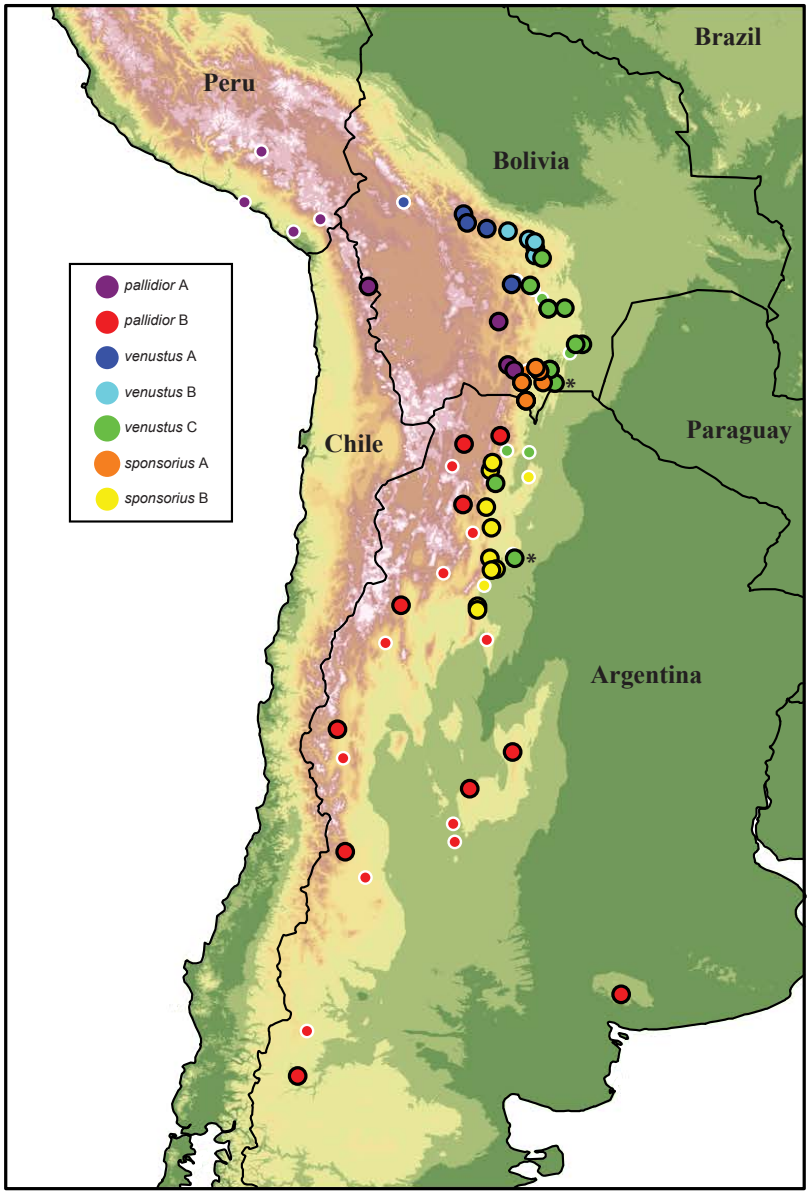

Figure 13. Map of collecting localities for specimens used in this study. Larger circles with black outlines indicate specimens for which we sequenced both CYTB and nuclear markers. Smaller circles with white outlines indicate specimens for which we only have CYTB sequences. Asterisks are placed by two collecting localities at which putative hybrid specimens were collected. Map shading reflects elevation, with dark green indicating lowlands, yellow indicating middle-elevations, and brown/white shading illustrating high/maximum elevations.

Haplotype phasing, sequence alignment, and neutrality tests

Heterozygous nuclear loci were phased in a three-step process. First, all length-variant heterozygotes were phased using the software Champuru 1.0 (Flot 2007). Champuru exploits the haplotypic information contained in the overlapping chromatograms of length-variant heterozygotes (Seroussi and Seroussi 2007; Sorenson and DaCosta 2011) and parses the haplotypes of any heterozygous single-nucleotide polymorphisms (SNPs) present in the sequence mixture. After running Champuru, remaining unresolved haplotypes were phased using the software package PHASE (Stephens et al. 2001). Input files for PHASE were created using the SeqPhase webserver (Flot, 2010). A "known haplotype" file containing all of the haplotypes resolved by Champuru was included in the PHASE analysis. PHASE was run using the default settings, except the threshold for accepting a haplotype was reduced to 0.7, an appropriate value based on simulation studies (Garrick et al. 2010). Finally, allele-specific sequencing primers were developed for heterozygous sequences that remained unresolved after the first two steps. PCR products from the original amplification attempts were re-sequenced with allele-specific primers designed following the recommendations of Scheen et al. (2012). Successfully primed haplotype sequences were added to the "known haplotype" file, and PHASE was re-run with the same settings as before. The four sequences that remained unresolved were removed from all subsequent analyses.

Sequences were aligned using Clustal 2.0 (Larkin et al. 2007). Most alignments were trivial and only required minor adjustments by eye. All alignments were tested for recombination using the DSS method (McGuire and Wright 2000) as implemented in the

software package TOPALi v2 (Milne et al. 2009). Neutrality of each locus was tested using Tajima's D in DnaSP v5 (Librado and Rozas 2009).

Gene trees and "species" trees

Gene trees sampled across *Thylamys pallidior, T. sponsorius*, and *T. venustus* were inferred for fourteen anonymous loci and one X-linked intron. For each locus, nucleotide substitution models were fit to phased sequence alignments using jModelTest 2.0 (Darriba et al. 2012) and ranked according to the Bayesian Information Criterion (BIC). The best-fitting model was chosen for each dataset, and phylogenetic trees were inferred in Garli 2.0 (Zwickl 2006). All of the default settings in Garli were used, and Garli runs were replicated five times for each locus to ensure consistency. For each locus, we report only the tree that received the highest likelihood.

We used BEAST v. 1.7.4 (Drummond et al. 2012) to infer an ultrametric tree of 102 *Thylamys pallidior*, *T. sponsorius,* and *T. venustus* CYTB sequences used in a prior study (Giarla et al. 2010). In that prior phylogenetic analysis of CYTB, which included all recognized *Thylamys* species, we found that a partitioning scheme in which each codon position received its own partition was the best fit (Giarla et al. 2010). Here, with a dataset that only included *Thylamys pallidior*, *T. sponsorius,* and *T. venustus*, we used PartitionFinder (Lanfear et al. 2012) to simultaneously evaluate nucleotide substitution models and codon-partitioning schemes. For the BEAST analysis, we assigned the best fitting model to each codon position and selected a lognormal clock model, with the prior on the ucld.mean parameter set to an exponential distribution with a mean equal to 1.0. Default settings were used for all other priors. BEAST was run for 50 million

generations, sampling every 50,000, and convergence of parameter estimates was assessed using the program Tracer 1.5 (Rambaut and Drummond 2007).

We estimated a "species" (lineage) tree from the 15 phased nDNA sequence datasets described above using BEAST 1.7.4 (Drummond et al. 2012). Based on results from the individual gene trees, sequences from two individuals with putative hybrid ancestry were removed from the datasets used to estimate the species tree (see Results). The *BEAST algorithm (Heled and Drummond, 2010) was implemented, and each sequence was assigned to a putative lineage based on the results from analysis of CYTB alone. The same nucleotide substitution model from the Garli analyses were used in BEAST. Each gene tree was tested for rate heterogeneity with likelihood-ratio tests by comparing likelihood scores of trees with and without molecular clock constraints. A strict molecular clock model was used for each locus that did not exhibit significant rate heterogeneity, whereas a lognormal relaxed-clock model was used for each locus that did. All nucleotide substitution and clock models were unlinked across loci, and all clock models were set to "estimate" so that the resultant tree would be scaled to substitutions per site. Default settings were used for all priors except for the lognormal clock model's ucld.mean priors, which were set to exponential distributions with a mean of 1.0. We ran the MCMC chain for 200 million generations, sampling every 20,000th generation, and assessed mixing in Tracer 1.5.

Bayesian "species" delimitation in BPP using nDNA

"Species" limits (lineage membership) based on mtDNA haplogroups were tested with fifteen nuclear loci (mtDNA excluded) using BPP v 2.1 (Yang and Rannala 2010). The BPP model assumes no exchange of genes between species (Mayr 1942), that

recombination does not exist within sampled loci, that all loci are independent, and that all loci are evolving neutrally. BPP implements a Jukes-Cantor model of nucleotide evolution because all sampled sequences are expected to be closely related. The posterior distribution of two population genetic parameters are sampled: contemporary and ancestral mutation-rate-scaled effective population sizes (θ) and mutation-rate-scaled species divergence times (τ). Priors for all θ parameters and the τ parameter for the root of the species tree are modeled as gamma distributions $\Gamma(\alpha, \beta)$, where the prior mean = α/β and its variance = α/β^2. All other τ parameters in the model were assigned a Dirichlet prior.

In the species delimitation model used by BPP, the variation inherent to the coalescent process is explicitly incorporated, and gene trees are sampled under the constraints of a user-defined guide tree. Given a guide tree, a reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm moves between different species delimitation models by collapsing and resolving nodes throughout the tree, from a fully resolved model wherein each tip of the guide tree is considered a species, to a fully collapsed model wherein all sequences are considered part of the same species. The rjMCMC chain samples a posterior distribution of speciation probabilities for each split in the pre-specified guide tree. Three separate datasets were initially constructed for species delimitation in BPP: (1) *Thylamys pallidior* A and B; (2) *T. sponsorius* A and B; and (3) *T. venustus* A, B, and C. Sequences were assigned to putative species based on mtDNA haplogroups, and the guide tree for *T. venustus* was based on the topology of the *BEAST species tree analysis (which also matched our previously studied CYTB topology, **Figure 10**).

The rjMCMC algorithm exhibits mixing problems for datasets with a large number of loci and sequences (Yang and Rannala 2010), as is the case for the three datasets considered here. Poor mixing is manifested by different runs giving substantially different results, or the chain getting "stuck" on the fully resolved or fully collapsed guide tree. During trial runs, these behaviors were observed for all three fully phased *Thylamys* datasets. Considerably better mixing was observed when a smaller, pruned dataset was used. Pruned datasets were constructed by randomly removing one haplotype sequence from each pair of phased sequences for a given diploid individual for all individuals (mostly eliminating duplicate sequences from homozygous individuals). For the X-linked intron OGT, males were already represented by one sequence alone, so only sequences from females were pruned by half. Pilot trials using the final pruned datasets reached the same results for different starting trees, suggesting that proper mixing of the rjMCMC algorithm was occurring.

Effect of priors on species delimitation. Misspecified priors can have a strong influence on the accuracy of species delimitation (Zhang et al. 2011), so different values for the gamma shape parameters α and β were tested for each dataset. In total, seven different sets of priors were tested (**Table 5**). The first set of priors considered (Scheme 1: *Thylamys*-specific) were derived based on *Thylamys*-specific estimates of means for τ and θ. Jansa et al. (*in review*) inferred a time-scaled phylogeny for Didelphidae based on five protein-coding nuclear genes and five external fossil calibrations; this dated phylogeny included *T. pallidior*, *T. venustus*, and two other congeneric species. The average age of the most recent split among these *Thylamys* species was estimated to be 0.8 Ma, and we assume that all splits among the haplogroups analyzed in this report (but

not analyzed by Jansa et al.) would have occurred before at least 0.7 Ma. Given this assumption and a nuclear gene mutation rate of ~10^{-9} substitutions per site (Kumar and Subramanian, 2002), a prior distribution for τ could be approximated with a mean of 0.0007. DNAsp v5 (Librado and Rozas 2009) was used to identify an initial mean value for Watterson's estimator of θ (θ_W; Watterson, 1975) for each of the three final datasets. All three datasets had similar average θ_W's across all loci (**Table 6**), and the average across the three datasets is 0.004. Diffuse gamma priors based on these estimates of θ and τ were $\theta \sim \Gamma(2, 500)$ and $\tau \sim \Gamma(2, 3000)$. Because the gamma distribution for the θ prior is based on the same data as used in the BPP runs themselves, Scheme 1 is not a strictly Bayesian approach to fixing priors. To deal with potential prior misspecification, independent prior schemes were considered. Similar to an approach suggested by Leaché and Fujita (Leaché and Fujita, 2010), six additional sets of priors were tested, in which three different divergence depths (~0.1 Ma, 1.0 Ma, and 10.0 Ma, assuming a nuclear gene mutation rate of ~10^{-9}) and two different effective population sizes (large, $\theta \sim 0.1$; and small, $\theta \sim 0.001$) were chosen (**Table 5**).

Table 5. Prior schemes tested in BPP. Assuming a mutation rate of 10^{-9} substitutions per site per year (Kumar and Subramanian 2002) divergence depths indicated as "Shallow" imply a prior with a mean ~0.1 Ma, "Moderate" implies a prior with a mean ~1.0 Ma, and "Deep" implies a prior with a mean ~10.0 Ma. For effective population sizes within haplogroups, "Large" assumes $\theta \sim 0.1$ and small assumes $\theta \sim 0.001$. Scheme 1 τ's were derived from Jansa et al.'s (*in review*) divergence time estimates. Priors were modeled as diffuse gamma distributions $G(\alpha, \beta)$, where the prior mean = α / β and its variance = α / β^2

Scheme Name	Diverg. Depth (τ)	Effective Pop. Size (θ)	Gamma Dist. for prior
Scheme 1	Est. from other data	Est. from Anon. Locus Data	$\theta \sim G(2, 500)$ and $\tau \sim G(2, 3000)$
Scheme 2	Shallow	Large	$\theta \sim G(1, 10)$ and $\tau \sim G(2, 20000)$
Scheme 3	Shallow	Small	$\theta \sim G(2, 2000)$ and $\tau \sim G(2, 20000)$
Scheme 4	Moderate	Large	$\theta \sim G(1, 10)$ and $\tau \sim G(2, 2000)$
Scheme 5	Moderate	Small	$\theta \sim G(2, 2000)$ and $\tau \sim G(2, 2000)$
Scheme 6	Deep	Large	$\theta \sim G(1, 10)$ and $\tau \sim G(1, 10)$
Scheme 7	Deep	Small	$\theta \sim G(2, 2000)$ and $\tau \sim G(1, 10)$

Table 6. Characteristics of nuclear loci used in this study. Exemplar sequences from each locus were mapped to the published genome sequence of the distantly related didelphid *Monodelphis domestica* using the best hit from a BLAT search (Kent 2002).

Locus Name	*M. domestica* Genome	Locus Size (bp)	Mol. Clock?[2]	Substitution Model[3]	Tajima's D[3,4]	θ_W [3]
Anon-61	Chromosome 5	546	Yes	HKY	-0.1912	0.00785
Anon-72	Chromosome 8	675	No	K80+G	-0.0266	0.01315
Anon-78	Chromosome 2	656	Yes	HKY	0.3668	0.00544
Anon-85	Chromosome 3	549	Yes	HKY	0.6177	0.00423
Anon-86	Unknown	716	Yes	HKY	-0.8977	0.00529
Anon-89	Unknown	650	Yes	TrN	1.3097	0.00410
Anon-90	Unknown	704	Yes	TrN+I	-0.7208	0.00775
Anon-94	Chromosome 3	752	Yes	HKY+I+G	0.361	0.01031
Anon-98	Chromosome 6	523	Yes	TrNef+I	-0.2023	0.01554
Anon-101	Chromosome 1	457	Yes	HKY+I+G	-0.2983	0.02127
Anon-115	Chromosome 1	550	Yes	HKY+I	0.6816	0.01284
Anon-121	Chromosome 1	543	Yes	K80	1.4407	0.00593
Anon-122	Chromosome 5	723	No	TPM1+I+G	0.008	0.01970
Anon-128	Chromosome 4	724	Yes	TPM1+I	-0.2977	0.00555
OGT	X-Linked Intron	653	Yes	HKY	0.222	0.00325

In order to ensure convergence and proper mixing of the rjMCMC algorithm, a total of four BPP runs (together referred to as a "set") were initialized for each of the three datasets by varying the starting tree (fully resolved or collapsed) and the rjMCMC algorithm (algorithm 0 or 1 from Yang and Rannala 2010). During pilot trials, the best mixing was observed for algorithm 0 with fine-tune parameter $\varepsilon = 20$ and for algorithm 1 with fine tune parameters $\alpha = 2$ and $m = 0.5$. As such, those tuning values were implemented in the corresponding final runs. All runs used the cleandata=1 setting to eliminate gaps; inter-locus rate heterogeneity was modeled using a Dirichlet distribution $D(\alpha)$, where $\alpha = 10$; and the heredity scalar for the X-linked OGT gene was set to 0.75.

[2] "Yes" denotes that the molecular clock cannot be rejected.
[3] Measured or fitted across all sequences sampled within *T. pallidior*, *T. sponsorius*, and *T. venustus* combined.
[4] All values are non-significant.

Each run consisted of a burn-in period of 50,000 steps and a sampling period of 500,000 steps (logged every 5), for a total of 100,000 samples.

Putative hybrids and confirmation of taxonomic species. Sequences from two individuals assigned to *Thylamys venustus* C (vouchered by OMNH 29966 and MSB 67392; Appendix 5) exhibited strikingly discordant signals between nuclear and mitochondrial trees. For nearly all of the nuclear trees, one or both of the alleles from these individuals sort with *T. sponsorius* sequences. Such strong and consistent cytonuclear conflict suggests introgression as opposed to incomplete lineage sorting. Consistent with this interpretation, both specimens are from a region of known geographic range overlap between *T. sponsorius* and *T. venustus*, and one of them (OMNH 29966) was previously recognized as a phenotypic intermediate based on craniodental measurements (Giarla et al. 2010: **Figure 18**). Because these putative hybrids likely represent gene flow between sympatric *T. sponsorius* and *T. venustus*, and not gene flow between allopatric haplogroups within *T. venustus* or *T. sponsorius*, these individuals were removed from all of the species delimitation analyses that concerned differentiation of haplogroups. To assess the taxonomic distinctiveness of *T. sponsorius* and *T. venustus*, a final set of BPP runs was completed on a dataset that grouped together haplogroups within these morphologically diagnosable species and included sequences from the putative hybrids.

Although the model assumes no gene flow between putative lineages, simulation studies have found that low levels of gene flow do not hinder species delimitation using BPP (Zhang et al. 2011; Camargo et al. 2012). A large dataset that included all individuals (including putative hybrids) and all loci for *T. sponsorius* and *T. venustus*

exhibited inconsistent results. To attain better mixing, four sequences from each haplogroup were randomly chosen for inclusion in a pruned dataset, along with the sequences from the two putative hybrids. Following the same procedure as described above, seven different prior schemes were tested to account for different depths of divergence and effective population size. In order to ensure consistent results, four rjMCMC runs were initialized for each of the prior schemes following the same procedure as for the haplogroup tests described in the previous section.

Impact of number of loci on species delimitation. In order to determine what effect the number of loci used in each BPP run might have on its ability to delimit independent evolutionary units, multilocus power analyses were completed (Roe et al. 2010). For each of the three taxonomic species, loci were randomly removed and BPP was re-run using rjMCMC algorithm 1 (with fine tune parameters $\alpha = 2$ and $m = 0.5$), Prior Scheme 1, the same run parameters as the species delimitation tests described above, and a randomly collapsed or resolved guide tree. Six dataset size classes were considered: 15 loci, 12 loci, 8 loci, 4 loci, 2 loci, and 1 locus. Within the 12-, 8-, 4-, and 2-locus size classes, four randomly rarefied datasets were constructed as replicates. For the single-locus datasets, all 15 loci were considered separately as single-locus replicates. For the complete, 15-locus dataset, the same dataset was used for four replicated BPP runs. The number of sequences per locus was the same across all trials within a given taxonomic species.

Assessment of mixing. For some datasets it can be difficult to distinguish between a run with poor mixing "stuck" on a fully resolved species model and a run in which the posterior probability of a given split in the tree is so high that the algorithm

should not be expected to visit the collapsed trees at an appreciable rate. To further verify that the rjMCMC algorithm was working effectively, nuclear sequences were randomly assigned to putative species within a given dataset, and BPP trials were repeated with Prior Scheme 1. Under this tip-randomization scheme, it is expected that a one-species model (where all of the nodes in the guide tree are collapsed) will be favored over a multi-species model. If the rjMCMC algorithm were to get stuck on a fully resolved guide tree despite randomly assigned sequences, the algorithm could not be expected to display appropriate mixing on the non-randomized datasets. However, if the algorithm is able to move between different levels of guide tree resolution and consistently results in a one-species model (i.e., the expectation for a random mixture of sequences sampled from two reproductively isolated species), the algorithm is behaving properly.

Species delimitation in SpeDeSTEM

SpeDeSTEM (Ence and Carstens 2011) is a computationally efficient approach to species validation based on the STEM method of species-tree estimation (Kubatko et al. 2009). Sequences are first assigned to putative species, then gene trees are estimated in PAUP* (Swofford 2002) and input files for STEM are generated for each possible permutation of lineage grouping. SpeDeSTEM extracts the log-likelihood scores from STEM for each permutation and ranks the models according to AIC score. This approach relies on the assumption that gene trees are estimated accurately and that the alignment for each locus does not deviate from a molecular clock model. After testing clock models for BEAST (described above), we removed two loci (Anon72 and Anon122) for which we were able to reject a molecular clock model and prepared input files for SpeDeSTEM for the remaining 13 loci. STEM requires an estimate of θ in order to scale the branch

lengths in the species trees it produces. Using DNAsp v5 (Librado and Rozas 2009), we computed the average θ across the 13 included loci (**Table 6**) but ultimately tested several different θ values in SpeDeSTEM. We ran SpeDeSTEM for 100 replicates, with each replicate including 4 randomly subsampled alleles per putative taxon as suggested by Hird et al. (2010), and tested all 20 possible permutations for haplogroup clustering within taxonomic species.

RESULTS

Gene tree and species tree results

Of the 14 anonymous nDNA markers developed for this study, 11 could be mapped to *Monodelphis domestica* autosomes (**Table 6**). The remaining three markers that could not be mapped to the *M. domestica* genome also did not receive any strong hits to the non-redundant "nr" nucleotide BLAST database, suggesting that these genomic regions do not contain protein-coding genes. Neutrality tests across all loci using Tajima's D statistic revealed that none of the loci exhibited significant signs of selection (**Table 6**), and tests for recombination using the DSS method revealed that only the alignment for Anon-94 exhibited signs of a recombination breakpoint. The shorter of the two regions on either side of the breakpoint in Anon-94 was removed from the alignment and from all subsequent analyses. Measured across all sampled sequences and species, θ_W ranges from 0.003 (the X-linked intron OGT) to 0.021 (Anon-101), and the average over all of the loci is 0.009.

For the BEAST analysis of CYTB, the best fitting data partitioning scheme included a separate partition and substitution model for each codon position (Position 1: TrNef+I; Position 2: HKY+I; Position 3: TrN). The ultrametric CYTB tree (**Figure 14**)

illustrates the relative divergence times of each haplogroup. *Thylamys sponsorius* haplogroups diverged most recently, whereas the deepest splits within *T. venustus* and *T. pallidior* appear to have occurred in the more distant past. Each of the 15 nDNA locus alignments was analyzed separately, and the resultant phylogenetic trees exhibit a wide range of relative substitution rates and topologies (Online Supplementary File 1). Notably, the haplogroup designations apparent in the CYTB tree (**Figure 14**) are not consistently resolved in many of the gene trees. In fact, for some of the nuclear loci, the taxonomic species *T. pallidior*, *T. sponsorius*, and *T. venustus* do not sort into monophyletic groups. Two individuals initially identified as *Thylamys venustus* C based on mtDNA sequences and morphology (tissue nos. Arg 1108 and NK 23992 in Appendix 5; collecting localities marked with asterisks on **Figure 13**) have nuclear sequences that cluster with *T. sponsorius*, providing evidence for limited introgression between *T. sponsorius* and *T. venustus*. Sequences from these individuals were removed from BPP analyses comparing haplogroups within taxonomic species but were retained in the analysis that lumped haplogroups in order to validate the species status of *T. venustus* versus *T. sponsorius*.

Results from the species-tree analysis of nuclear data in *BEAST (**Figure 15**) support the same tree topology as the CYTB tree alone (**Figure 14**), with posterior probabilities ≥0.95 at all nodes. For *Thylamys venustus*, haplogroups B and C together formed a clade to the exclusion of haplogroup A, and we used this topology for the *T. venustus* guide tree in BPP.

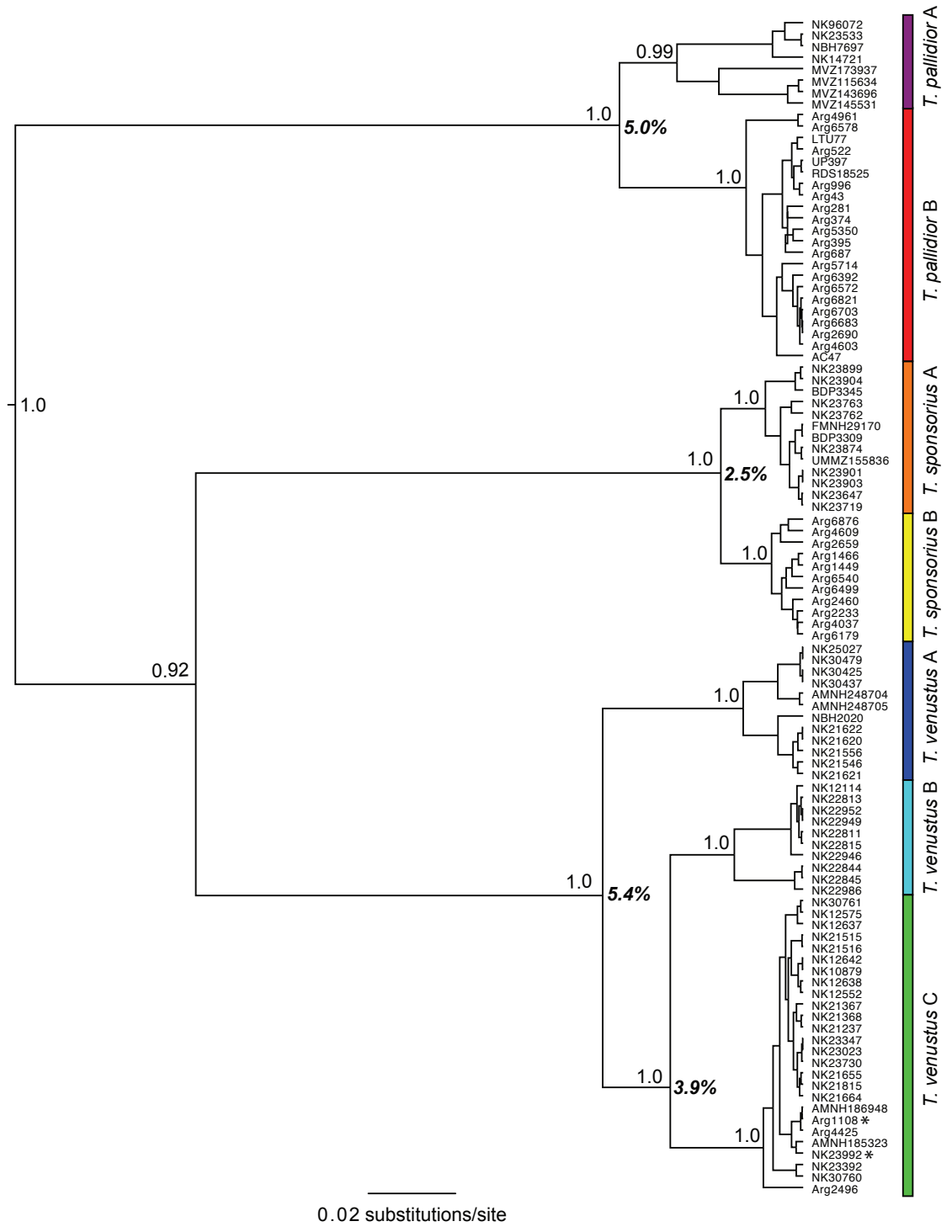

Figure 14. Ultrametric phylogenetic tree inferred in BEAST using the complete set of CYTB sequences from Giarla et al. (2010). Tip labels correspond to tissue voucher numbers (Appendix 1). Only posterior probability values for interspecific divergences and the deepest split within each haplogroup are displayed. Percentages near some nodes reflect the uncorrected average pairwise divergence between the node's daughter lineages. Asterisks mark two specimens that cluster with *T. venustus* C for CYTB but cluster with *T. sponsorius* for nearly all nuclear loci (Online Supplementary File 1).

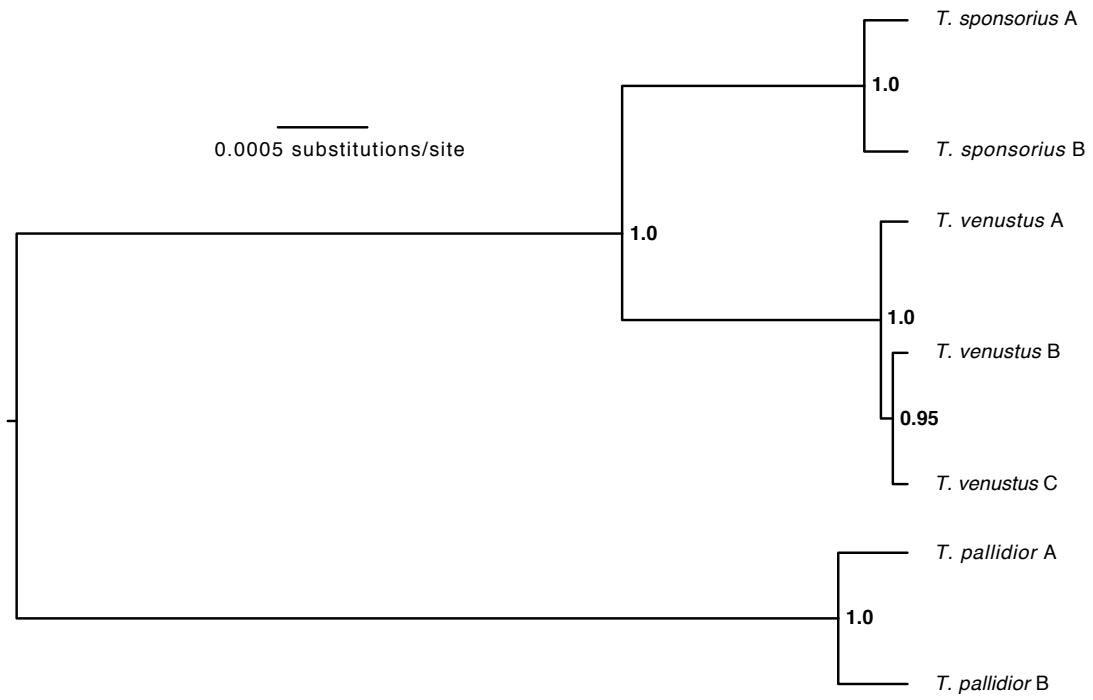

Figure 15. Species tree inferred in *BEAST from 14 anonymous loci and the X-linked intron OGT, with posterior probabilities at each node. Individuals were assigned to haplogroups based on the CYTB results. Two individuals of suspected hybrid origin were not included in this analysis.

Species Delimitation using BPP with Fifteen Nuclear Loci

Performance of the rjMCMC algorithm. As expected, BPP runs in which sequences were randomly assigned to haplogroups within a taxonomic species never recovered significant support for the defined haplogroups (**Table 7**), indicating that the dataset size and parameters used were tractable for BPP's rjMCMC algorithm, at least for Prior Scheme 1. In total, 35 sets of analyses on the non-randomized datasets were conducted (**Table 7**), with each set comprising 4 replicates that varied by rjMCMC algorithm (0 or 1 from Yang and Rannala 2010) and collapsed/resolved starting tree. We considered a set inconsistent—and thus mixing poorly—if at least one run strongly supported a different species delimitation scenario from the other runs. Of the 35 sets,

three sets using Prior Scheme 6 ("Deep-Large") exhibited inconsistent results across replicates (**Table 7**). The remaining 32 sets converged on similar results across replicates, which indicates efficient mixing of the rjMCMC chain.

Table 7. Results from species validation analysis in BPP of haplogroups within *T. pallidior, T. sponsorius*, and *T. venustus*. All numbers indicate the posterior probability that the pair under consideration can each be considered distinct species, with bolded numbers indicating support over 0.95 PP. The first row ("Randomized Tips") contains the results of the BPP trials in which sequences were randomly assigned to species in order to test performance of the rjMCMC algorithm.

Prior Scheme	*pallidior* A - *pallidior* B	*sponsorius* A - *sponsorius* B	*venustus* A - *venustus* B/C	*venustus* B - *venustus* C	*venustus* - *sponsorius*[5]
Randomized Tips[6]	0.02	0.0	0.14	0.03	0.0
1, *Thylamys*-Specific	**1.0**	**1.0**	**0.99**	0.59	**1.0**
2, Shallow-Large	**1.0**	**1.0**	**0.98**	0.77	**1.0**
3, Shallow-Small	**1.0**	**1.0**	**1.0**	0.54	**1.0**
4, Moderate-Large	**1.0**	**1.0**	0.91	0.41	**1.0**
5, Moderate-Small	**1.0**	**1.0**	0.17	0.15	**1.0**
6, Deep-Large	n/a[7]	**1.0**	n/a[7]	n/a[7]	**1.0**
7, Deep-Small	**1.0**	**1.0**	0.02	0.017	**1.0**

The effect of priors on species delimitation. The effect of applying different prior schemes on species delimitation could be studied for the 32 sets of analyses that exhibited proper mixing. For *Thylamys pallidior* and *T. sponsorius*, both sets of haplogroups receive strong support as distinct species, irrespective of prior scheme. Within *T. venustus*, the sets that exhibited proper mixing consistently reject the model that splits haplogroups A, B, and C into three distinct species. For the deeper split within *T. venustus*—between haplogroup A and the clade that unites haplogroups B+C—the choice of prior dramatically affects the results. Results based on Prior Schemes 1, 2, and 3 agree that A should be distinct from B+C, but results based on the remaining prior

[5] Sequences from putative hybrid individuals were included.
[6] Sequences were randomly assigned to haplotype groups within each taxonomic species and run under Prior Scheme 1 as defined in **Table 5**.
[7] Replicated runs within this set exhibited poor mixing.

schemes support uniting all of *T. venustus* haplogroups into one species. In order to confirm that *T. sponsorius* and *T. venustus* were, in fact, genetically isolated taxa, separate analyses were conducted on a dataset that pooled all of the haplogroups within each of these taxonomic species. Unlike the previous tests involving *T. sponsorius* and *T. venustus* haplogroups, sequences from two individuals with suspected hybrid ancestry were not removed before running BPP. Across all of the prior schemes, BPP consistently resolved *T. sponsorius* and *T. venustus* as distinct species at the highest level of support (**Table 7**).

Multilocus power analysis. The effects of randomly removing loci from BPP analyses varied among the three taxonomic species (**Figure 16**). At one extreme, the *Thylamys sponsorius* A and *T. sponsorius* B were consistently differentiated with only two loci, and seven single-locus replicates received posterior probabilities (PP) for species delimitation >0.95 (**Figure 16b**). The distinction between *T. pallidior* A and *T. pallidior* B received consistently high support (1.0 PP) with eight or more loci, but two four-locus replicates recovered these groups with >0.95 PP (**Figure 16a**). The differentiation of *T. venustus* haplogroup A from B+C received consistent support (≥0.95 PP) with eight loci (**Figure 16c**), but support varied (between 0.78 and 1.0 PP) for replicates of 12 loci; for 15 loci, the average support for this split was 0.99 PP. Across all trials, support for differentiation between *T. venustus* B and *T. venustus* C never exceeded 0.80 PP, and the maximum number of loci (15) supports differentiation between these two lineages with only 0.59 PP (**Figure 16d**). Across all species, the single-locus replicates were especially variable, with some single-locus replicates accurately

delimiting species with posterior probabilities above 0.95 but most coming nowhere near this threshold.

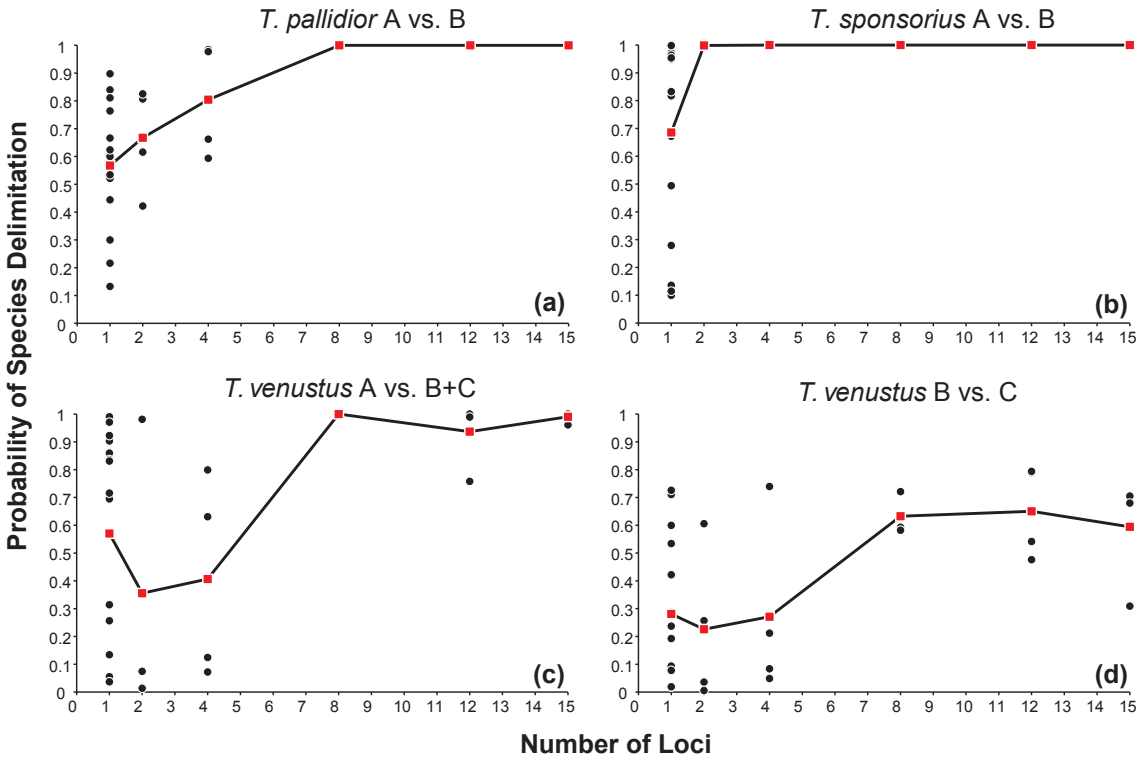

Figure 16. Multilocus power analysis in BPP for each split of interest under Prior Scheme 1. Four randomly rarefied datasets were constructed for the 12-, 8-, 4-, and 2-locus size classes; all 15 single-locus datasets were analyzed. A single complete dataset including all 15 loci together was independently analyzed four times. Each black circle represents the posterior probability of species delimitation based on a single random replicate dataset. Squares denote the average posterior probability value across all of the random trials for each dataset size class.

Species Limit Validation using SpeDeSTEM

When we used Watterson's estimator of genetic diversity across all species and haplogroups (0.009) as the estimate for θ in SpeDeSTEM, the program would not work properly, presumably due to a recognized issue with scaling factors in STEM (described in the SpeDeSTEM User Manual). After increasing θ incrementally, we were able to use SpeDeSTEM only with θ values higher than 0.05.

Table 8. Results from the species validation test in SpeDeSTEM. For each model of lineage splitting, haplogroup combinations are clustered with "+" symbols and separated from other such clusters with "|" symbols. The models are ranked from top to bottom according to AICc score. The absolute difference between the AICc score for the given model and the best-fitting one is listed under the column labeled "Δ_i" and the model weighting is listed under the column labeled "w_i." Only results from the analysis in which θ was set to 0.05 are shown. Results from larger θ values were very similar, whereas smaller θ values caused the program to fail.

Species Models	*Avg. –lnL (100 reps)*	*AICc*	Δ_i	*Model-likelihood*	w_i
*pal*A+*pal*B | *spo*A+*spo*B | *ven*A+*ven*C | *ven*B	666.58	666.58	0.00	1.00	0.52
*pal*A+*pal*B | *spo*A | *spo*B | *ven*A+*ven*C | *ven*B	668.47	668.47	1.89	0.39	0.20
*pal*A | *pal*B | *spo*A+*spo*B | *ven*A+*ven*C | *ven*B	668.48	668.48	1.90	0.39	0.20
*pal*A | *pal*B | *spo*A | *spo*B | *ven*A+*ven*C | *ven*B	670.37	670.37	3.79	0.15	0.08
*pal*A+*pal*B | *spo*A+*spo*B | *ven*A+*ven*B+*ven*C	707.60	707.60	41.02	0.00	0.00
*pal*A+*pal*B | *spo*A+*spo*B | *ven*A+*ven*B | *ven*C	709.42	709.42	42.84	0.00	0.00
*pal*A | *pal*B | *spo*A+*spo*B | *ven*A+*ven*B+*ven*C	709.50	709.50	42.92	0.00	0.00
*pal*A+*pal*B | *spo*A | *spo*B | *ven*A+*ven*B+*ven*C	709.51	709.51	42.92	0.00	0.00
*pal*A+*pal*B | *spo*A+*spo*B | *ven*A | *ven*B+*ven*C	709.57	709.57	42.99	0.00	0.00
*pal*A | *pal*B | *spo*A+*spo*B | *ven*A+*ven*B | *ven*C	711.32	711.32	44.74	0.00	0.00
*pal*A+*pal*B | *spo*A | *spo*B | *ven*A+*ven*B | *ven*C	711.35	711.35	44.77	0.00	0.00
*pal*A+*pal*B | *spo*A+*spo*B | *ven*A | *ven*B | *ven*C	711.40	711.40	44.82	0.00	0.00
*pal*A | *pal*B | *spo*A | *spo*B | *ven*A+*ven*B+*ven*C	711.41	711.41	44.82	0.00	0.00
*pal*A | *pal*B | *spo*A+*spo*B | *ven*A | *ven*B+*ven*C	711.47	711.47	44.89	0.00	0.00
*pal*A+*pal*B | *spo*A | *spo*B | *ven*A | *ven*B+*ven*C	711.47	711.47	44.89	0.00	0.00
*pal*A | *pal*B | *spo*A | *spo*B | *ven*A+*ven*B | *ven*C	713.25	713.25	46.66	0.00	0.00
*pal*A | *pal*B | *spo*A+*spo*B | *ven*A | *ven*B | *ven*C	713.30	713.30	46.72	0.00	0.00
*pal*A+*pal*B | *spo*A | *spo*B | *ven*A | *ven*B | *ven*C	713.33	713.33	46.75	0.00	0.00
*pal*A | *pal*B | *spo*A | *spo*B | *ven*A | *ven*B+*ven*C[a]	713.37	713.37	46.79	0.00	0.00
*pal*A | *pal*B | *spo*A | *spo*B | *ven*A | *ven*B | *ven*C	715.23	715.23	48.65	0.00	0.00

Results from our SpeDeSTEM analysis of the 13 nDNA loci that fit the assumption of a molecular clock do not support our hypothesis that all of the mtDNA haplogroups within *Thylamys pallidior*, *T. sponsorius*, and *T. venustus* should be considered separate species (**Table 8**; only results from the analysis in which θ was set to 0.05 are shown). Instead, the model that receives the highest support (52% of the model weighting) unites the haplogroups within *T. pallidior* and *T. sponsorius*, and separates out *T. venustus* B from *T. venustus* A+C. The latter result contrasts with both our mtDNA topology and species tree analysis, in which *T. venustus* A is sister to haplogroups B+C.

The next two best-supported models (each receiving about 20% of the overall model weighting) are similar to the top choice, but either *T. pallidior* A and B or *T. sponsorius* A and B are considered distinct in the models. The model that receives the lowest weighting is the one in which all of the haplogroups are considered distinct, and the model supported by BPP is ranked 19th of the 20 possibilities.

DISCUSSION

The main goal of this study was to test the evolutionary independence (genetic isolation) of mitochondrial lineages within three montane *Thylamys* species that were identified in a previous phylogenetic study of this genus (Giarla et al. 2010). To do so, we analyzed 15 nuclear DNA markers using two commonly used coalescent-based approaches: BPP (Yang and Rannala 2010) and SpeDeSTEM (Ence and Carstens 2011). This study design facilitates an empirical evaluation of analytical approaches to "species" delimitation, and we consider this topic first before addressing the biological and evolutionary implications of our results.

Comparison of "species delimitation" methods

Results from BPP and SpeDeSTEM support different conclusions regarding number of evolutionarily independent units within the species studied. Whereas BPP supports the recognition of six independent lineages (two each within each of the species), SpeDeSTEM supports the recognition of only four (one each for *T. pallidior* and *T. sponsorius* plus two within *T. venustus*). Moreover, the recognition of lineages within *T. venustus* differs between the two programs: BPP supports the basal split

between the mitochondrial lineages A versus B+C, whereas SpeDeSTEM supports a split between B and A+C that was not apparent in analysis of the mitochondrial data or the nDNA species tree analysis. The difference between the lineage-splitting models supported by the two analysis is significant: the top four models in the SpeDeSTEM analysis encompass 100% of the total model probability, but the model supported by BPP is ranked 19th out of 20 possible models. Such a large difference suggests that assumptions inherent in the two different approaches to lineage recognition are strongly influencing our results.

One significant issue with BPP (or any Bayesian) analysis concerns the appropriate specification of priors (Zhang et al. 2011). In our case, different priors affect both the probability of delimiting multiple species (**Table 4**) and the performance of the algorithm (e.g., consistency of the rjMCMC algorithm across multiple runs). One reasonable approach to prior specification suggested by Leaché and Fujita (2010) is to analyze data under a range of different prior schemes that encompass possible speciation scenarios. Although this approach is gaining in popularity and is far superior to assuming a single set of priors, these different prior schemes cannot be expected to work in all cases and results should be interpreted carefully. Notably, Leaché and Fujita (2010) imply that setting priors for large population size and recent divergence times will be the "most difficult" scenario in which to test species delimitation, which is not necessarily the case. Any prior that is not concordant with a strong signal from the data has the potential to cause problematic results. Even more troubling, an incorrect prior in conjunction with weak signal in the data could lead to a result strongly biased by the prior (Zhang et al. 2011).

For our data, a prior scheme that coupled deep divergence times with a large population size (Scheme 6) seemed especially problematic. For three of our five cases, analyses under this set of priors failed to exhibit adequate mixing of the rjMCMC chain, and results from these analyses are therefore unreliable (**Table 4**). Setting this apparently unreasonable prior scheme aside, posterior probability estimates of lineage independence (or lack thereof) were reassuringly consistent across a wide range of reasonable prior schemes for most of our cases. The single exception concerns whether to recognize one or two independent lineages within *T. venustus*. The decision here rests on which prior scheme is more reasonable for these data: one that assumes a recent (~0.1 Ma) basal divergence within the taxonomic species versus ones that assume moderate (~1.0 Ma) or deep (~10.0 Ma) divergence, irrespective of effective population size (**Table 4**).

In an attempt to evaluate this question, we estimated priors for our sample by using existing data. To establish a prior for the splitting-time parameter τ, we used a phylogeny of Didelphidae based on five independent nuclear protein-coding genes that was time-calibrated with four fossils (Jansa et al., *in review*). For θ, we used the average value of θ_W from the 15 nuclear loci used in this study. This "*Thylamys*-specific" prior exhibited good chain mixing in all cases, and recovered consistent posterior probability values for the four cases that were not sensitive to prior specification (**Table 4**). In the one case where prior specification affected our results (whether to recognize one vs. two lineages of *T. venustus)*, posterior probability estimates derived under this prior scheme were consistent with those derived under prior schemes that assumed a recent divergence time, but differed from those assuming moderate or deep divergences. In this case, the population-size prior seemed to exert relatively little influence, as posterior probability

estimates for the two shallow-divergence schemes (large vs. small population size, schemes 2 and 3, respectively) were similar. Although using such empirically derived parameter estimates for construction of priors might not be possible for some systems, this approach can help distinguish among reasonable and unreasonable prior schemes. In our case, the deeper divergence priors apparently conflict with the signal in the data, a behavior that was also observed by Zhang et al. (2011) in their simulation study.

For SpeDeSTEM, a significant analytical issue concerns the reliability of the method for inferring very recent lineage divergence. Two recent simulation studies testing the efficacy of SpeDeSTEM have addressed this issue. Ence and Carstens (2011) found that the method was effective at delimiting species that have diverged as recently as 0.25Ne generations ago, as long as more than 10 loci were used. Camargo et al. (2012), however, found that SpeDeSTEM performed quite poorly in scenarios in which putative species diverged less than $0.75N_e$ generations ago, even when 10 loci were used. They note that the discrepancy between their simulations and Ence and Carsten's (2011) may be due to the relative amount of genetic diversity (as approximated by θ) in the gene sequences, with low-information alignments leading to increased uncertainty in gene tree estimation (Camargo et al. 2012). Our estimates of θ and τ (0.004 and 0.0007, respectively) suggest that all of the putative divergences considered here occurred approximately $0.175N_e$ generations ago, putting these *Thylamys* species well below the range in which SpeDeSTEM has been demonstrated to perform acceptably. Moreover, Camargo et al. (2012) found that BPP performed very well across a range of different simulated datasets, even when putative species were modeled to have diverged from one another only very recently. As such, we think it is appropriate to rely on the BPP results

over the SpeDeSTEM results, and this study further supports the conclusions about the efficacy of SpeDeSTEM for recent divergences noted by Camargo et al. (2012). It is likely that error in estimation of the gene trees is high enough to obscure the signal of recent divergence and that it is necessary to incorporate this error directly into species validation methods (as done in BPP) in order to recover the signal.

Lineage recognition using mitochondrial vs. nuclear loci

Our multilocus analysis clearly indicates that six of the seven mitochondrial haplotype groups we tested correspond to genetically isolated lineages. The concordance between mitochondrial and nuclear gene analyses in these six cases provides additional evidence that the inferred mitochondrial structure reflects true lineage splits rather than stochastic lineage sorting. However, the two most recently formed mitochondrial haplogroups within *T. venustus*—even though they are each strongly supported and exhibit relatively high sequence divergence—appear simply to reflect stochastic processes. It would be difficult to predict this latter result from our analysis of the mitochondrial data alone. For example, the most recent split in our mitochondrial phylogeny defines the two haplogroups within *T. sponsorius* (**Figure 14**). If the timing of this split reflects the relative recency of population splitting, then we might expect nuclear loci to have had insufficient time to coalesce and define these two lineages. Perhaps surprisingly then, analyses of the nuclear gene data provide very robust support for recognition of these two lineages: the 15-locus BPP analysis supports two lineages of *T. sponsorius* across the full range of prior schemes (**Table 7**), and our multilocus power analysis suggests that these two lineages could be identified with datasets comprising as few as two nuclear loci (**Figure 16**). Conversely, the split between the mitochondrial

lineages B and C of *T. venustus* apparently predates the split within *T. sponsorius* (**Figure 14**), implying a higher probability that nuclear genes would have coalesced for these lineages. Nevertheless, none of the BPP analyses supports this split, even when 15 loci are included. Thus, although it has been argued that mitochondrial data—due to its relatively small effective population size and rapid coalescence time—is a leading marker of within-species genetic structure (Zink and Barrowclough 2008), our results suggest that it is impossible to distinguish population divergence from stochastic lineage sorting based solely on analyses of mitochondrial markers.

Nuclear data are also critically important for addressing the relative timing of lineage splitting events. Notably, for montane species such as these, coincident splits across multiple species could reflect a synchronized response to a single or very closely spaced geological or climatic events (Arbogast and Kenagy 2001; Hickerson et al. 2006; McCulloch et al. 2010; Bell et al. 2012). Point estimates of relative divergence times derived from our mitochondrial DNA tree suggest that the basal divergence events within each of the three taxonomic species were asynchronous, with simultaneous splits in *T. pallidior* and *T. venustus* preceding the split within *T. sponsorius* by a factor of two (**Figure 14**). However, the error surrounding these single-locus estimates is so large that the possibility of synchronous splitting cannot be rejected (**Figure 17**). In contrast, mean estimates of relative lineage divergence time from the 15-locus nuclear gene dataset are all more recent than those derived from the single-locus mitochondrial dataset; moreover, the precision of these nuclear gene estimates appears to be far higher. The error estimates surrounding basal lineage splitting times derived from the 15-locus nuclear dataset are at least five times smaller than the error surrounding estimates derived from the single

mitochondrial locus. With these reduced error estimates in hand, rather than failing to reject synchronous splitting, our confidence in the conclusion that these splits were either synchronous or very closely spaced in time is markedly improved.

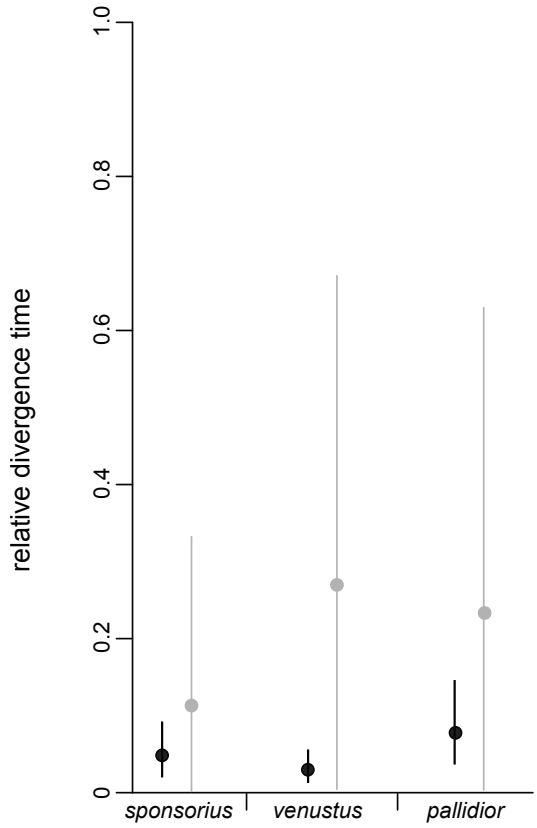

Figure 17. Point estimates (circles) and surrounding error estimates (lines) for the relative divergence time for the basal-most split within each species. Black dots and lines derive from the *BEAST analysis of the 15 nuclear loci; gray dots and lines are from the BEAST analysis of the cytochrome *b* data. In both analyses, the root node (the split between *T. pallidior* and *T. venustus*) was set at 1.0.

A large number of loci might not be needed to identify cryptic lineages

At the shallow level of divergence considered here, individual gene trees based on a single nuclear locus are not well resolved and reveal little about potential species limits (Online Supplementary File 1), a result that is in accordance with the expectations of coalescent theory for recently diverged species (Hudson and Coyne 2002). When considered in combination and as part of a coalescent framework, nuclear genes have much greater power to detect phylogenetic units; by integrating over phylogenetic error,

the BPP algorithm leverages this power to weigh alternative species delimitation scenarios (Yang and Rannala 2010). Coalescent-based species delimitation methods—including BPP and SpeDeSTEM—may offer an objective, quantitative method by which to weigh alternative hypotheses of lineage independence (Fujita et al. 2012). However, one potential drawback to these methods is that putative species boundaries and guide trees must be defined in advance, and incorrect assignments may lead to erroneous conclusions. Given these issues, it is evident that preliminary studies using mitochondrial sequences or morphology are still important steps in detecting cryptic diversity. In addition, other multilocus methods for species delimitation that do not rely on *a priori* species assignments are available to facilitate lineage recognition in potentially cryptic taxa (Pons et al. 2006; O'Meara 2010; Reid and Carstens 2012; Rittmeyer and Austin 2012). Thus, whereas morphology might not detect recently diverged lineages, and mitochondrial sequences might over-split species, multi-locus nuclear data and associated analytical approaches offer a powerful way to interrogate cryptic diversity.

With this in mind, the question of how many independent loci—and how many alleles for a given locus—to sample becomes an important analytical concern (Pluzhnikov and Donnelly 1996; Felsenstein 2006; Carling and Brumfield 2007; Brito and Edwards 2008; Dupuis et al. 2012). With massively parallel sequencing platforms becoming more widely available and cheaper to use, it is easier than ever before to generate large quantities of sequence data. A fundamental question remains, however: is it really necessary to use thousands of loci to delimit species when traditional approaches may be cheaper (at least for the time being) and just as effective? Also, if loci are limiting, and incorporating gene tree uncertainty is critical, next-generation sequencing

data (which might have only one or two SNPs per locus; McCormack et al. 2013) may not be the best choice given current analytical tools. Our results indicate that BPP can consistently detect lineage distinctness with sequences from as few as two (and possibly only one) nuclear loci, a finding that concurs with simulation studies. Zhang et al. (2011) found that a single locus might be enough to identify some recent speciation events, as long as a large number (ca. 15+) of sequences per putative species are used. Camargo et al. (2012) identified a similar pattern, and further found that BPP performed better than other coalescent-based methods like SpeDeSTEM in situations with post-divergence gene flow and recent splits. SpeDeSTEM, with its reliance on point-estimates for gene trees, would likely perform better with more information-rich alignments (i.e., higher θ values) because they would garner more robustly supported gene trees. Regardless, our data suggest that a moderate number (rather than thousands) of loci can provide relevant information for identification of cryptic lineages, even when divergence is relatively recent.

Introgression detected between two species

Thylamys sponsorius and *T. venustus* are found in similar habitats on the eastern Andean slopes and foothills of Bolivia and northern Argentina, and they are known to occur sympatrically in at least one sampling area (Giarla et al. 2010). Despite deep divergence in the mtDNA tree, these species are very similar morphologically, with no single character consistently distinguishing them. Previously, a multivariate morphometric analysis of craniodental characters demonstrated that samples of sequenced specimens could be distinguished by linear combinations of measurement variables; these factors were then used to assign taxonomic names to molecular clades by

computing scores of unsequenced holotypes (Giarla et al. 2010). Two sequenced specimens, however, could not be convincingly assigned to species based on morphometrics alone (Giarla et al. 2010), hinting at the possibility that they might be hybrids. Genetic data reported here provides limited support for that possibility. One of these morphometric intermediates (Arg 1108) is also one of the two individuals that clusters with the *T. venustus* C clade in the CYTB tree but with *T. sponsorius* in the nDNA gene trees (Online Supplementary File 1). In contrast, NK 23992 is morphometrically assigned to *T. venustus* with high confidence, but exhibits just as much cytonuclear discordance as Arg 1108. Because Arg 1108 exhibits both morphometric intermediacy and strong cytonuclear discordance, we feel confident in concluding that it comes from a population that has experienced introgression. Although NK 23992 is confidently assigned to *T. venustus* based on morphology alone, the strong degree of discord between the CYTB results and the nDNA results also support the notion that *T. sponsorius* and *T. venustus* have exchanged genes.

A similar pattern could be observed if both populations shared ancestral polymorphisms, but this explanation seems unlikely given the high degree of concordance observed across all of the independent nuclear loci we sequenced. If incomplete lineage sorting were the source of cytonuclear discordance, we would expect less consistent sorting across loci. Given our results, it seems likely that the mitochondrial genome of *T. venustus* C has introgressed through at least part of the *T. sponsorius* population. Before further conclusions can be drawn, more focused sampling at putative contact zones in northern Argentina and additional genetic studies are needed.

Nonetheless, this is the first reported case of introgression between species in the family Didelphidae.

Are these cryptic lineages species?

Although BPP and other coalescence-based approaches are often described as methods for "species delimitation" or "species validation," they effectively test the hypothesis that populations (or metapopulations) of related organisms have ceased to exchange migrants and are therefore isolated genetically (Yang and Rannala 2010; Ence and Carstens 2011). Genetic isolation, however, is a necessary but not sufficient property for species recognition because gene flow can be prevented by extrinsic factors such as geographic separation (allopatry), rather than by intrinsic factors that maintain lineage separation. As Yang and Rannala (2010: 9269) remarked, their method is expected to be maximally useful for identifying cryptic species that occur in sympatry (where genetic isolation is most likely due to intrinsic mechanisms), and that species delimitation "should rely on many [other] kinds of data, such as morphological, behavioral, and geographic evidence."

All of the mtDNA haplotypes confirmed as genetically isolated lineages in this report are allopatric, and inference about their status as full species is correspondingly problematic. Because speciation is a process, the discovery of lineages in various stages of divergence is to be expected. Although such discoveries pose challenges for taxonomic interpretation, they are also opportunities for productive research on the sequence of evolutionary intermediates associated with speciation in different plant and animal clades.

Names are available for all of the distinct lineages of *Thylamys* identified by our BPP results. As we have previously discussed (Giarla et al. 2010), the name *venustus* is based on a type specimen collected within the known geographic range of the lineage we designate *T. venustus* A, whereas the type of *cinderella* (currently regarded as a junior synonym of *venustus*), was collected within the combined geographic ranges of *T. venustus* B and C. Although sequence data from topotypic specimens would be preferable, taxonomic usage could plausibly be based on these map criteria. Similarly (see Giarla et al. [2010] for type localities), the name *janetta* could be used for *T. sponsorius* A, the name *sponsorius* could be restricted for *T. sponsorius* B, the name *pallidior* could be restricted for *T. pallidior* A, and the name *fenestrae* could be used for *T. pallidior* B.

For the reasons explained earlier, we do not think that a revised binomial nomenclature based on such assignments are warranted until the hypothesis that these lineages are full species is better corroborated (sensu De Queiroz 2007). Logical extensions of our results that might provide such corroboration could consist of ecological niche modeling to explore the possibility that each lineage is associated with a distinct range of environmental conditions (as in cryptic salamanders; Rissler and Apodaca 2007), or more sophisticated phenotypic analyses that might uncover subtle divergence in morphology overlooked in previous revisionary work (as by Berendzen et al. 2009). The latter would be particularly welcome as a basis for more securely associating type material with taxa than the geographic criteria suggested above. Lastly, fresh collections along Andean transects carefully sited to intersect contact zones (if any)

between sister lineages could provide crucial evidence for genetic isolation in sympatry or parapatry. Absent such corroboration, taxonomic restraint seems only prudent.

CHAPTER 3

THE ROLE OF PHYSICAL GEOGRAPHY AND HABITAT TYPE IN THE BIOGEOGRAPHIC HISTORY OF AN ECOLOGICALLY VARIED AND RECENT RADIATION OF NEOTROPICAL MAMMALS (*THYLAMYS*: DIDELPHIDAE)

INTRODUCTION

One of the first tasks for any historical biogeographic analysis is the selection of a biogeographic regionalization scheme appropriate for the question at hand (Harold and Mooi 1994; Kreft and Jetz 2010). When choosing among different schemes for dividing up a geographic space into meaningful elements, researchers can take into account various aspects of both physical geography (e.g., the margins of an island or a mountain ridge) and ecology (e.g., an ecotone that defines the margins of two plant communities). Although these biogeographic factors are not independent of one another, it is useful to consider the relative contribution of each as dispersal barriers. For example, the ecotone between open and forested habitats might prove to be a more effective dispersal barrier to a small mammal than the Amazon River—even over evolutionary time scales—but both must be considered to test the hypothesis. Typically, a single biogeographic regionalization that relies on expert opinions about areas of endemism and habitat types is used to test these questions, but more quantitative approaches are also available (reviewed by Kreft and Jetz 2010). For taxa spread across a wide geographic range, a single regionalization scheme can lose its generality if the constituent areas are too fine-grained. In situations like these, it might be useful to isolate the effects of different biogeographic factors by considering two broadly defined regionalization schemes based

on different concepts of what might block dispersal, limit gene flow, or promote vicariance.

Within continents, the significance of rivers and mountains as barriers to gene flow has been debated extensively, especially in the Neotropics. In his study of primate distributions in the Amazon Basin, Wallace (1852) was the first to suggest that the Amazon River and its tributaries might limit the ranges of vertebrate taxa and thus promote allopatric speciation. Several subsequent phylogeographic and ecological studies found evidence consistent with this hypothesis (Ayres and Clutton-Brock 1992; Aleixo 2004; Ribas et al. 2012), whereas other studies have not (Patton et al. 1994; Gascon et al. 2000). The Andes have been shown to act as dispersal barriers and potential vicariant forces for lowland taxa on both sides of the mountain chain (Brumfield and Capparella 1996; Miller et al. 2008). Moreover, the topographically complex and ecologically varied habitats within the Andean cordillera itself are potential drivers of diversification (Brumfield and Edwards 2007; Ribas et al. 2007; Chaves et al. 2011). Indeed, the tropical Andes contains the highest endemic terrestrial vertebrate biodiversity in the world (Myers et al. 2000), and the drivers of this remarkable diversity has been a critical topic in Neotropical biogeography. A major unresolved question is the extent to which Andean taxa evolved *in situ* or dispersed to higher elevations from lowland areas (Lynch 1986; Hall 2005; Ribas et al. 2007).

Although physical geography is often a focus of historical biogeographic hypotheses, changes in habitat in the absence of physical barriers to dispersal are also worth considering. Habitat type varies depending on physical aspects of the earth's environment, such as elevation, soil chemistry, rainfall patterns, and temperature.

However, different habitats offer emergent properties to the species that occupy them, and these properties (e.g., shade, protection from predators, food) are not captured by the properties of the abiotic environment alone. For small mammals, vegetation communities in particular may be especially important in limiting dispersal due to differences in food availability and predation risk between adjacent habitat types (Krohne 1997; Lidicker 1999; Andruskiw et al. 2008; Abu Baker and Brown 2010).

Given the potential importance of both habitat type and physical barriers to dispersal, and potential differences in the effectiveness of both type of barriers, we argue that a dual approach to the historical biogeography of widespread taxa could be especially illuminating. In this study, we consider the role of physical geography and habitat type on the historical biogeography of *Thylamys* mouse opossums, a genus within the New World marsupial family Didelphidae that, until recently, had been difficult to study because species boundaries and evolutionary relationships were not well understood (Giarla et al. 2010). Most *Thylamys* species occupy open habitats in South America, where they are distributed among a wide range of arid and seasonably dry ecoregions from sea level to nearly 4000 m a.s.l. in the Andes (**Figure 18**). The genus is an ideal Neotropical mammalian study system for historical biogeographic analysis due to the detailed distributional records, rigorous taxonomy, and well-supported phylogenetic evidence provided in recent revisionary work (Carmignotto and Monfort 2006; Voss et al. 2009; Giarla et al. 2010).

The southern half of South America is home to a variety of arid and seasonably dry habitats. These primarily non-forested areas extend from the high-elevation grasslands of the Andes to the lowland savannas of the east, and include areas with both

exceptionally high biodiversity (e.g., the Brazilian cerrado; Myers et al. 2000) and very low biodiversity (e.g. the Atacama desert of Chile). Nonetheless, the historical biogeography of Neotropical open habitats has been relatively neglected compared to other Neotropical regions (Werneck 2011). Understanding the biogeographic history of taxa distributed in these regions has the potential to reveal the forces that influence diversification and the relative frequency of vicariance and dispersal.

Here, we present a biogeographic analysis of *Thylamys* and its monotypic sister genus *Lestodelphys* that is based on the first fully sampled species-level, multilocus, time-calibrated tree for this clade. We use recently developed methods in historical biogeography (Ree et al. 2005; Ree and Smith 2008) to test the roles that physical geography (rivers and mountains) and habitat type (terrestrial biomes) may have played in the diversification and range evolution of *Thylamys* species. Results from this analysis reveal scenarios for geographic range evolution and cladogenesis within *Thylamys* and allow us to assess the extent to which transitions between biomes affected diversification in *Thylamys*, the role of rivers as vicariant forces, and the extent to which Andean species evolved *in situ* or dispersed into the Andes multiple times.

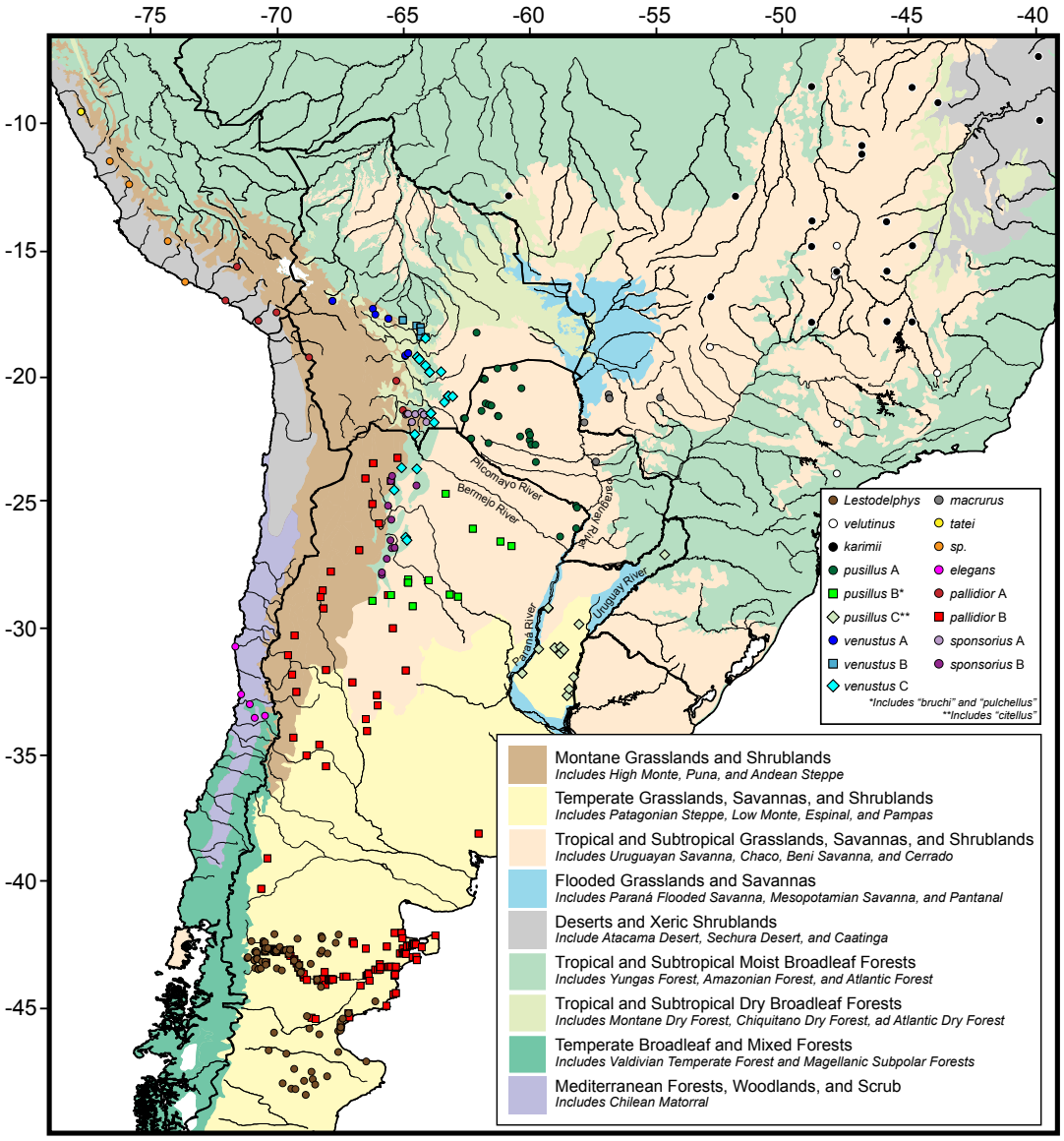

Figure 18. Terrestrial biomes of central and southern South America from Olson et al. (2001), with *Thylamys* and *Lestodelphys* collecting localities. Each species receives its own symbol color, with different shades of that color and symbols (circle, square, diamond) defining intraspecific haplogroups.

MATERIALS AND METHODS

Inferring a Time-Scaled Phylogenetic Tree

A robust phylogeny is a prerequisite to considering the historical biogeography of a clade. Here, we use the marsupial fossil record to calibrate a phylogeny of *Thylamys* to

absolute time under a relaxed molecular clock in BEAST 1.7.5 (Drummond et al. 2012). Although we are not aware of any fossils that can be reliably dated and assigned to *Thylamys*, there are at least three marsupial fossils outside of *Thylamys* that can be used as external calibrations in BEAST: two from within Didelphidae and one from Australidelphia. We included the minimum age of the split between *Didelphis* and *Philander* (3.3 Ma), the minimum age of the clade that comprises *Monodelphis*, *Marmosa*, and *Tlacuatzin* (12.1 Ma), and the minimum age of the split between Dasyuromorphia and Peramelemorphia (24.7 Ma). Details on the dating of these fossils can be found in Jansa et al. (*in review*).

We compiled a supermatrix of DNA sequences from across Marsupialia, with particularly dense sampling within *Thylamys.* In total, 50 terminal taxa were included in the DNA alignment: seven non-didelphid metatherians and 43 didelphids (Online Supplementary File 2). Included among the didelphids are sequences from 15 *Thylamys* individuals representing each major haplogroup recovered by Giarla et al. (2010) or confirmed in Chapter 2. Because coalescent-based species delimitation could not distinguish the mitochondrial haplogroups *T. venustus* B and *T. venustus* C (Chapter 2), we will refer to the two mitochondrial clades together as *T. venustus* B. We sampled three mitochondrial genes (CYTB, COX2, and ND2), four protein-coding, autosomal genes (BRCA, IRBP, DMP1, and vWF), five nuclear introns (KREMEN, PPIC, P4HB, SLC38, and the X-linked OGT), and 14 anonymous autosomal loci, for a total alignment length of over 22,000 base pairs. This collection of loci was primarily compiled from Voss and Jansa (2009), Giarla et al. (2010), and Chapter 2, but sequences from other studies were also downloaded from GenBank. The four autosomal introns were developed for this

study following the same design, amplification, and sequencing procedure as described for the X-linked intron OGT in Chapter 2. Museum voucher numbers, GenBank accession numbers, and gene sampling per taxon are recorded in Online Supplementary File 2.

We assessed the optimal partitioning strategy for the whole dataset using PartitionFinder (Lanfear et al. 2012). We linked and assigned site models in BEAST according to the best partitioning scheme, unlinked all of the clock models to allow for rate heterogeneity among loci, and set each clock model to a lognormal relaxed clock. In order to obtain a single, concatenated tree, we linked all of the tree models for each locus and assigned it a single Yule tree prior. We enforced monophyly of three nodes and fossil-calibrated the analysis by assigning lognormal priors to the age of the nodes, offset by the minimum estimated age of the fossil assigned to that clade. We assigned exponential priors (mean = 1.0) for all of the ucld.mean clock rate priors, and kept all other priors at their default values. We ran this analysis for 200 million generations, sampling every 20,000th generation, to obtain a posterior distribution of 10,000 trees. Mixing was assessed by examining effective sample sizes (ESS) and trace diagrams in Tracer 1.5 (Rambaut and Drummond 2007) with a 10% burn-in, and a maximum clade credibility tree was constructed from the posterior sample of 9,000 trees using TreeAnnotator 1.7.5 (Drummond et al. 2012).

Biogeographic Modeling

Choice of biogeographic units. To better understand the role of habitat-type versus the role of physical isolating barriers such as mountains and rivers, two biogeographic area-coding schemes were considered: Terrestrial biomes (**Figure 19a**:

Biome-Based Areas), largely following Olson et al. (2001) and (2) a classification scheme that emphasizes the importance of the Andes along with three major river systems in southern South America (**Figure 19b**: Barrier-Based Areas). The Biome-Based Areas are modified from Olson et al.'s (2001) global terrestrial ecoregion designations. We defined the Barrier-Based Areas by the natural boundaries formed by the large rivers of the Río de la Plata basin (the Río Paraguay, Río Paraná, and Río Uruguay) and, further west, the western cordillera of the Andes. The Río de la Plata basin is among the largest river basins in the world, and the three rivers we chose to focus on are the largest constituent rivers in the basin when measured by overall discharge rate (Garcia and Vargas 1996). The northeastern part of Argentina bounded by these three rivers on all sides is commonly referred to as the "Mesopotamian Region," and we use that term here as well. Finally, we note that the boundary between "West of the Río Paraguay" and "East of the Río Paraguay" becomes irrelevant north of the Paraguay-Bolivia border, where only *T. karimii* occurs north of the source of the Río Paraguay in Brazil.

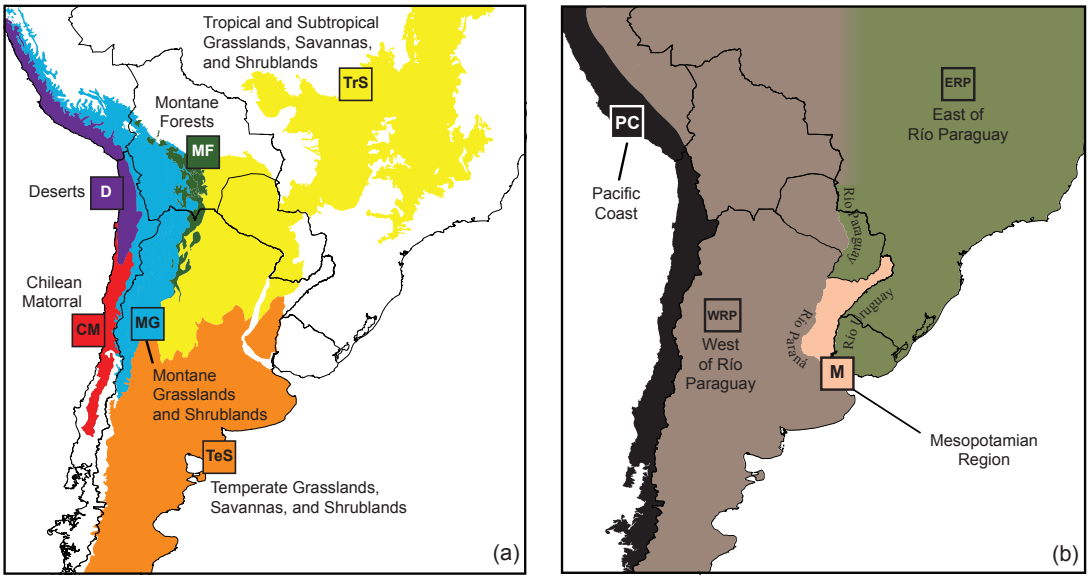

Figure 19. Biome-Based (a) and Barrier-Based (b) area-coding schemes. For (a), ecoregions where *Thylamys* has never been trapped (e.g. lowland Amazonian rainforest or Uruguayan savanna) are not colored.

Determining area assignments for *Thylamys* taxa. The geographic range of each taxon included in this study (including morphologically cryptic haplogroups within the named species *T. pallidior*, *T. pusillus*, *T. sponsorius*, and *T. venustus*) was based on georeferenced *Thylamys* and *Lestodelphys* collecting localities recorded in recent taxonomic and phylogenetic studies (Braun et al. 2005; Carmignotto and Monfort 2006; Teta et al. 2009; Giarla et al. 2010; Formoso et al. 2011), and our assignments are recorded in **Table 9**. In order to assign previously sequenced individuals to intraspecific haplogroups defined by Giarla et al. (2010), all CYTB Sequences from Braun et al. (2005) and Teta et al. (2009) were downloaded from GenBank and combined in a single dataset with CYTB sequences from Giarla et al. (2010) for phylogenetic analysis.

This dataset of 181 sequences (including five outgroup sequences from Giarla et al. [2010]) was analyzed in a Bayesian framework using MrBayes 3.2 (Ronquist et al. 2012). The GTR+I+G nucleotide substitution model fit the data best according to

jModelTest 2.1.3 (Darriba et al. 2012), and this model was specified for the full CYTB dataset in MrBayes for a single MCMC run of 20 million generations with four Metropolis-coupled chains. Previously unassigned sequences from Braun et al. (2005) and Teta et al. (2009) clustered unambiguously with the haplogroups defined in Giarla et al. (2010) and were assigned accordingly (results not shown). Once all sequences were assigned to species or haplogroups, the georeferenced localities were added to the range map (**Figure 18**).

Table 9. Taxon Area Assignments

Taxon	**Reference(s)**[8]	**Biomes**[9]	**Barriers**[10]
Lestodelphys halli	F2011	TeS	WRP
Thylamys elegans	G2010	CM	PC
Thylamys karimii	CM2006	TrS	ERP
Thylamys macrurus	CM2006; C2007; V2009	TrS	ERP
Thylamys pallidior A	G2010	MG, D	WRP, PC
Thylamys pallidior B	B2005; G2010; F2011	MG, TeS	WRP
Thylamys pusillus A	T2009; G2010	TrS	WRP
Thylamys pusillus B	T2009; G2010	TrS	WRP
Thylamys pusillus C	T2009; G2010	TrS, TeS	M
Thylamys sp.	G2010	D	PC
Thylamys sponsorius A	G2010	MG, MF	WRP
Thylamys sponsorius B	G2010	MF	WRP
Thylamys tatei	G2010	D	PC
Thylamys velutinus	CM2006	TrS	ERP
Thylamys venustus A	G2010	MG	WRP
Thylamys venustus B	G2010	MG, MF	WRP

[8] Reference abbreviations are as follows: Braun et al. 2005 (B2005), Carmignotto and Monfort 2006 (CM2006), Cáceres et al. 2007 (C2007), Voss et al. 2009 (V2009), Teta et al. 2009 (T2009), Giarla et al. 2010 (G2010), Formoso et al. 2011 (F2011).
[9] Area abbreviations are as follows: Chilean matorral (CM), deserts and xeric shrubland (D), montane forests (MF), montane grassland (MG), temperate savanna (TeS), tropical savanna (TrS).
[10] Area abbreviations are as follows: east of Río Paraguay (ERP), west of Río Paraguay (ERP), Pacific coast (PC), Mesopotamian region (M).

Several taxa discussed in Giarla et al. (2010) are based on specimen records from a limited portion of the suspected total range of the haplogroup (e.g., *Thylamys pusillus* A, B, and C). In order to expand the number of localities for these poorly sampled haplogroups, unsequenced material from Teta et al. (2009) and Formoso et al. (2011) was considered. Teta et al. include locality records of specimens from several taxa considered synonyms of *T. pusillus* by Giarla et al.: *T. citellus, T. bruchi*, and *T. pulchellus*. As discussed in Giarla et al. (2010), *T. citellus* corresponds to *T. pusillus* C and both *T. bruchi* and *T. pulchellus* correspond to *T. pusillus* B. Specimens referred to simply as "*T. pusillus*" by Teta et al. correspond to *T. pusillus* A as defined by Giarla et al. (2010). Although the assignment of these specimens to particular haplogroups would best be accomplished by phylogenetic analysis of DNA sequences, the geographic distributions of unsequenced specimens concord well with sequenced specimens and most occur within a single defined area (e.g. tropical savanna). "*Thylamys pusillus"* localities from Teta et al. all occur north of the Río Bermejo and west of the Río Paraguay, in the same region as the sequenced specimens assigned to *T. pusillus* A by Giarla et al. Teta et al.'s *T. bruchi* and *T. pulchellus* locales are all recorded in central Argentina south of the Río Bermejo and west of the Río Paraguay, the same broad area inhabited by Giarla et al.'s *T. pusillus* B. Finally, Teta et al.'s *T. citellus* localities all occur between the Ríos Paraná and Uruguay in northeastern Argentina, the same isolated area as the sequenced specimens assigned to *T. pusillus* C by Giarla et al. Given these distributions, we assigned Teta et al.'s unsequenced material to the corresponding Giarla et al. haplogroup and add the georeferenced localities to the range map (**Figure 18**).

Although Giarla et al. (2010) sequenced a broadly distributed sample of 30 *T. pallidior* individuals, the southern and eastern extent of *T. pallidior* haplogroup B's range was not robustly sampled. As such, all *T. pallidior* localities from Formoso et al. (2011) were assigned to haplogroup B based on their geographic distribution in the far south of *T. pallidior* B's range. We assign Formoso et al.'s *T. pallidior* samples to *T. pallidior* B because these individuals could not plausibly be assigned to the more northerly haplogroup A. Finally, we also included all records of *Lestodelphys halli* reported in Formoso et al. (2011) to bolster our sample size for that taxon.

Verification and modification of biogeographic area maps. Satellite imagery from Google Earth was used to verify and refine ecoregion assignments for localities near ecoregion boundaries and in regions where Olson et al.'s (2001) line maps were not sufficiently precise (Särkinen et al. 2011). Biome assignments for *Thylamys pallidior* B were revised based on satellite imagery. Six *T. pallidior* B collecting locales occur within the boundaries of Olson et al.'s "Tropical and Subtropical Savanna, Grassland, and Shrubland" biome. Upon closer inspection, each record can be found to occur on isolated mountain ridges within the Sierras de Córdoba in San Luis and Córdoba provinces. Satellite imagery from these localities show that these isolated montane habitats are quite distinct from the surrounding lowlands and would be more appropriately classified in montane or temperate biomes. As such, *T. pallidior* is not assigned to the Tropical Savanna biome.

We did not follow Olson et al. (2001) and distinguish between montane elements of moist broadleaf forests and tropical dry forests in southern Bolivia and northern Argentina (Yungas forests). These two forest types distinguished by Olson et al. are

contiguous, more similar to each other than to any other adjacent biomes (e.g., montane grasslands and lowland savanna), and so different from the other more open habitats that *Thylamys* inhabits that we join them into the general category of "Montane Forest". In addition, we did not consider Olson et al.'s Atlantic forest and flooded grasslands ecoregions in our analysis, because *Thylamys* species that are purported to inhabit these areas based on collecting localities all occur near the boundaries of included biomes and likely represent marginal populations.

Implementation of the DEC model in Lagrange. The biogeographic history of *Thylamys* was modeled using the program Lagrange (Ree and Smith 2008). Lagrange is a maximum likelihood approach that implements the Dispersal-Extinction-Cladogenesis (DEC) model (Ree et al. 2005; Ree and Smith 2008). The DEC model expands the utility of other event-based biogeographic models (e.g., DIVA; Ronquist 1997) by allowing for the incorporation of additional prior information, including branch lengths, area connectivity, and dispersal constraints. Two Lagrange models were run for both coding schemes: an unconstrained model, where dispersals among all areas and any combination of ancestral areas are allowed, and a stepping-stone model, where dispersals are prohibited between non-adjacent areas and only combinations of adjacent ancestral areas are allowed. We did not explore scenarios in which an ancestral lineage can inhabit three or more ancestral areas at once because all modern *Thylamys* species and haplogroups inhabit only one or two areas. The time-scaled maximum clade credibility tree inferred in BEAST was pruned to only include *Thylamys* and its monotypic sister-genus *Lestodelphys* before being used in Lagrange.

RESULTS

Phylogenetic Results

The complete alignment is 22,684 bp long and includes 23 nuclear loci and three mitochondrial loci. The optimal locus-partitioning scheme found by PartitionFinder included 12 partitions; the model of nucleotide evolution that fits each partition best are listed in **Table 10**. Overall, 50.9% of the alignment matrix is missing data, the majority of which is anonymous locus and intron sequences missing from non-*Thylamys* taxa (Online Supplementary File 2). After removing 10% of the trees as burn-in, ESS values estimated in Tracer suggested good mixing.

Table 10. Best Fitting Partitioning Scheme and Nucleotide Substitution Models

Partition Number	Best Model	Loci Included in Partition
1	HKY+G	Anon121, Anon61, Anon78, Anon85, PPIC
2	HKY+I+G	Anon101, Anon115, Anon122, Anon72, Anon90, Anon98
3	GTR+G	Anon86, Anon94, DMP
4	TrN+G	Anon89
5	K80+G	Anon128, P4HB
6	TrN+G	BRCA
7 (mtDNA)	GTR+I+G	COX2, CYTB
8	K80+I+G	IRBP, KREMEN, VWF
9 (mtDNA)	HKY+I+G	ND2
10 (X-linked)	HKY+G	OGT
11	GTR+I+G	RAG1
12	SYM+G	SLC38

The maximum clade credibility tree estimated from the 9,000 remaining trees (**Figure 20**; pruned to only include *Thylamys* and *Lestodelphys*) recovers the same relationships previously reported by Giarla et al. (2010), except for one key difference among *T. pusillus* haplogroups (Clade 1; **Figure 20**). Here, *T. pusillus* haplogroups A and B are recovered as sister lineages; based on mitochondrial DNA alone, haplogroups B

and C were recovered as sister lineages (Giarla et al. 2010). Divergence time estimates reveal that *Thylamys* diverged from its sister genus *Lestodelphys* approximately 3.4 Ma (95% highest posterior density interval [HPDI]: 2.7 – 4.1 Ma) and that the first split within *Thylamys*, between the subgenera *Thylamys* and *Xerodelphys*, occurred approximately 2.7 Ma (95% HPDI: 2.2 – 3.3 Ma). Divergences among species within the subgenus *Thylamys* are estimated to have occurred 1.2 and 0.2 Ma.

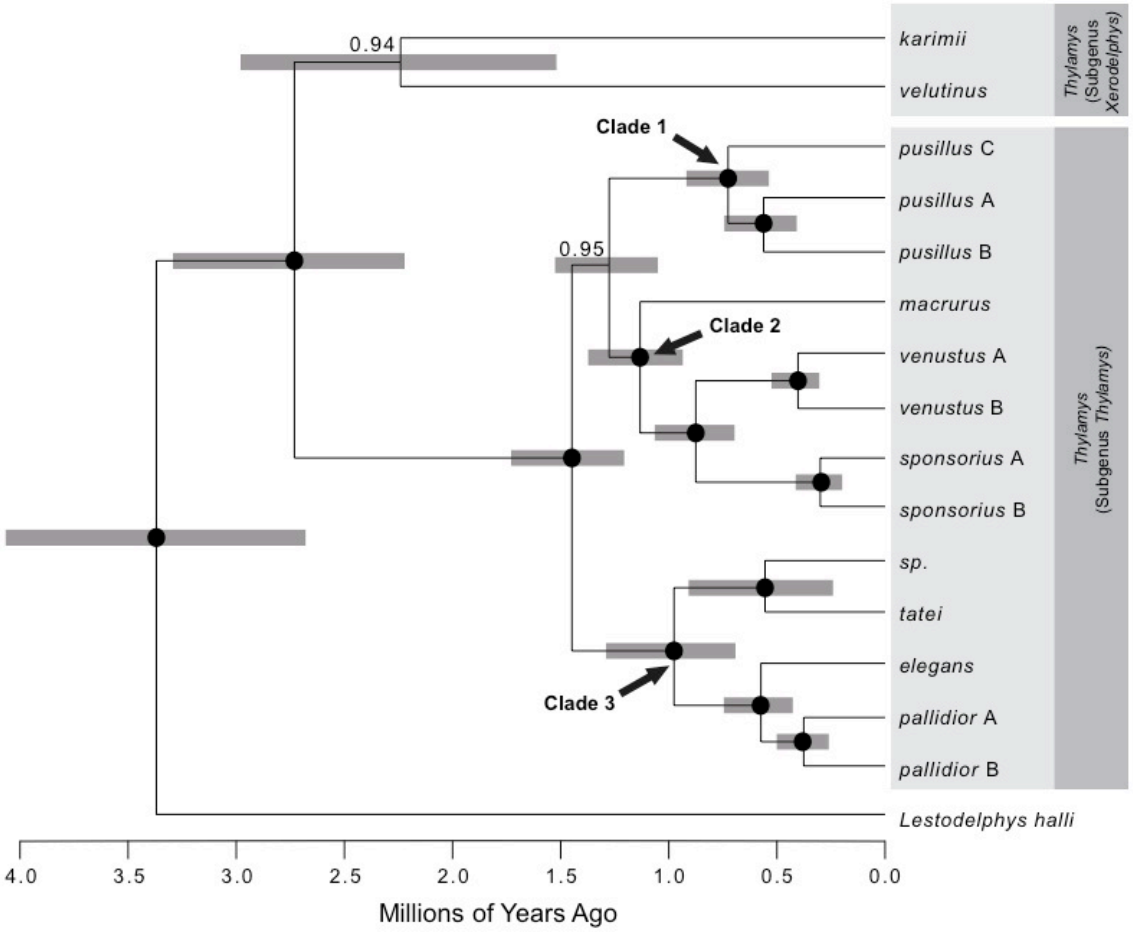

Figure 20. Time-scaled, fossil-calibrated maximum clade credibility tree based on a BEAST analysis of 23 nuclear loci and three mitochondrial loci. All outgroups have been removed from the tree, leaving only *Thylamys* and *Lestodelphys*. Bars around nodes represent the 95% highest posterior density interval for node ages. Black circles at each node signify posterior probability values of 1.0, whereas numbers at nodes indicate the values for nodes with lower support.

Biogeographic Models

Biome-Based Areas. For the Biome-Based Areas, the stepping stone model had a higher likelihood than the unconstrained model (lnL = -38.70 vs. lnL = -41.76, respectively), and only results from the best-fitting model are reported (**Figure 21**, **Table 11**). We report uncertainty in range inheritance scenarios at a given split with pie charts on each daughter branch, but all of the specific biogeographic events (e.g., range expansion events) illustrated in the figures or discussed below are based on the single scenario at each node that received the highest likelihood. Overall, four of the 15 total cladogenesis events reconstructed under the Biome-Based Areas occurred at the juncture of two areas ("cladogenesis across areas [CAA] events; indicated by asterisks in **Figure 21**). The remaining 11 cladogenesis events occurred within a single area. Range expansions into new habitat types occurred along eight of 30 total branches in the tree (excluding the root). No local extinctions of habitat types along lineages inhabiting multiple areas were reconstructed; lineages only "lost" an area due to vicariance.

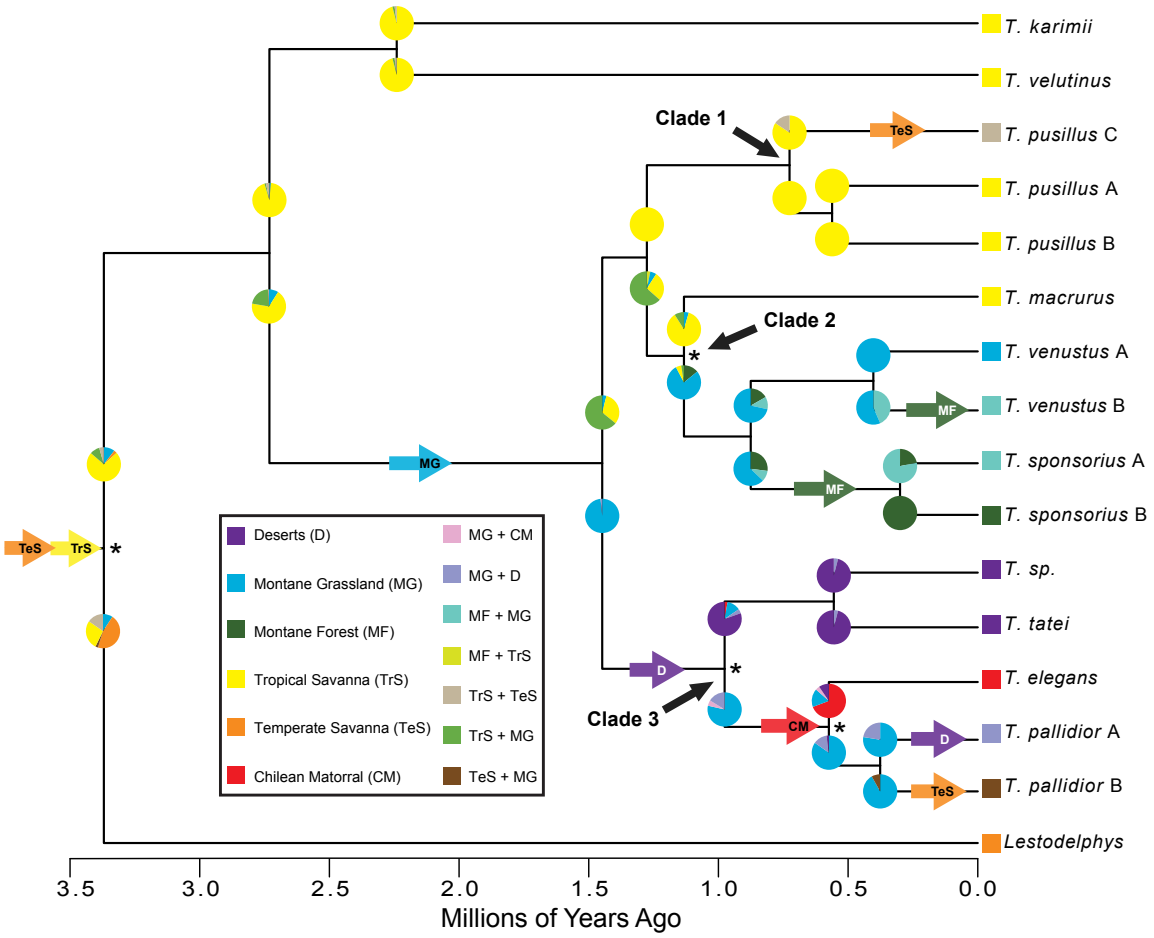

Figure 21. Range subdivision and inheritance scenarios from the stepping-stone Biome-Based Lagrange reconstruction. Events illustrated at nodes or along branches represent the range inheritance scenario that received the highest likelihood. Asterisks at nodes indicate cladogenesis across area (CAA) events, and colored arrows along branches indicate implied range expansion events. Pie charts along branches illustrate the relative proportion of the overall likelihood received for each range inheritance scenario within 2 lnL likelihood units of the most likely reconstruction for the adjacent node.

Table 11. Summary of Lagrange Model Results

Biogeographic Event[11]	**Biome-Based Model**	**Barrier-Based Model**
Within-area cladogenesis events	11	13
Between-area cladogenesis events	4	2
Range expansions	8	2
Local range extinctions	0	0

[11] The total number of cladogenesis events across the tree is 15, and the total number of branches over which range expansions and extinctions could occur is 30.

For the Biome ancestral area reconstruction, a combination of the tropical and temperate savannas comprises the most likely ancestral area for *Thylamys* + *Lestodelphys* (TrS + TeS; **Figure 21**). The split between *Thylamys* and *Lestodelphys* is reconstructed as a CAA event across the temperate/tropical savanna divide, with all of the earliest branching lineages persisting within the temperate savannas (*Lestodelphys*) or the tropical savannas (*Thylamys karimii, T. velutinus*, and the lineage leading to the remaining species). The ancestral lineage leading to the remaining *Thylamys* species expanded into montane grasslands in the early-to-mid Pleistocene, but still persisted in the tropical savannas. The montane forest biome is reconstructed to have been colonized first along the branch leading to *T. sponsorius* A+B and then again, independently, along the branch leading to *T. venustus* B. Now, *Thylamys venustus* B and *T. sponsorius* A are sympatric in at least part of their range (Giarla et al. 2010). Biomes along the Pacific coast are reconstructed to have been independently colonized three times within the clade that includes *T. elegans*, *T. pallidior*, and *T. tatei,* but no sites of sympatry among the three invaders are known (Clade 3; **Figure 21**). Finally, different portions of the temperate savanna biome are reconstructed to have been recolonized independently twice along terminal branches: from tropical savannas on the branch leading to *T. pusillus* C and from montane grasslands on the branch leading to *T. pallidior* B.

Lagrange results: Barrier-Based Areas. For the Barrier-Based Areas, the stepping stone model received a higher likelihood score than the unconstrained model (lnL = -24.45 vs. lnL = -25.10), so only results from the best-fitting model are reported (**Figure 22**, **Table 11**). Across the whole tree, two out of 15 total cladogenesis events are recovered as CAA events (coded as asterisks in **Figure 22**) with the remaining 13 splits

occurring within a single area. Range expansions occurred along 2 of 30 total branches in the tree, and no lineages experienced local extinction in a given area along any of the branches.

At the base of the tree, the ancestral lineage of *Thylamys* + *Lestodelphys* is reconstructed as inhabiting a broad swath of areas east of the Andes but is absent from the Mesopotamian region of Argentina (**Figure 22**). After *Lestodelphys* splits with *Thylamys* within the area west of the Río Paraguay, *Thylamys* maintains a broad ancestral distribution. The lineage leading to *Thylamys karimii* and *T. velutinus* splits from remaining members of the genus within the area east of the Río Paraguay, with the other *Thylamys* lineage persisting in its broad ancestral range. The first of two CAA events occurs when *T. macrurus* diverges from its sister lineage across the Río Paraguay (at the base of Clade 2; **Figure 22**). The second CAA event occurs after the lineage leading to the three *T. pusillus* haplogroups expands into the Mesopotamian region of Argentina between 1.25 and 0.75 Ma, and *T. pusillus* C diverges from its sister lineage across the Río Paraná. Pacific coastal regions are reconstructed to have been invaded only once, at the base of Clade 3, with some of the descendant lineages maintaining widespread ranges on both sides of the Andes.

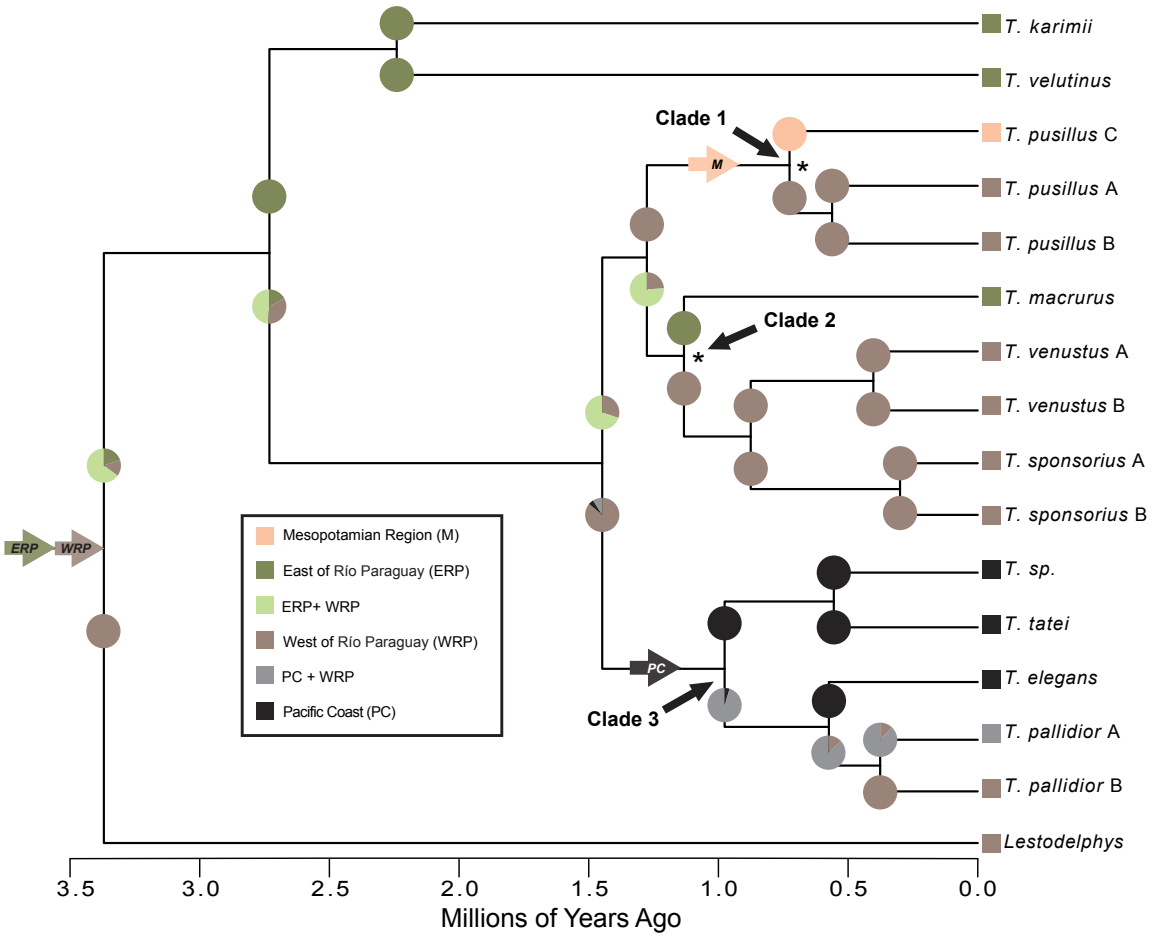

Figure 22. Range subdivision and inheritance scenarios from the stepping-stone Barrier-Based Lagrange reconstruction. Events illustrated at nodes or along branches represent the range inheritance scenario that received the highest likelihood. Asterisks at nodes indicate cladogenesis across area (CAA) events, and colored arrows along branches indicate implied range expansion events. Pie charts along branches illustrate the relative proportion of the overall likelihood received for each range inheritance scenario within 2 lnL likelihood units of the most likely reconstruction for the adjacent node.

DISCUSSION

Alternative Area-Coding Schemes

Most biogeographic studies implement only a single area-coding scheme. In some cases, the choice of areas is relatively straightforward, such as considering individual islands in an archipelago. However, for studies of continental taxa, classifying geographic areas into meaningful biogeographic regionalizations is rarely trivial (Harold and Mooi

1994; Hausdorf 2002; Kreft and Jetz 2010) and poor choices (e.g., over-splitting) can bias results (Ree and Smith 2008).

In this study, we considered two area-coding schemes based on different aspects of South American ecology and geography. First, the Biome-Based coding scheme was implemented in order to consider the role of habitat type in shaping diversification within *Thylamys*. Other historical biogeographic studies have used biomes or ecoregions as units of analysis (e.g., Schnitzler et al. 2011), but few discuss the implications and assumptions of coding areas in such a manner. The Barrier-Based areas were used to consider only the impact of potential physical isolating barriers. It is important to note, however, that these two coding schemes are not necessarily independent from one another. For example, the western cordillera of the Andes is strongly correlated with a shift in relative aridity. Indeed, the rain shadow caused by the mountain chain is one of the primary drivers of this sharp ecological transition (Houston and Hartley 2003). Because these two models are based on two different area-coding schemes, the results are not directly comparable in a statistical framework. Nonetheless, we argue that both coding schemes offer informative and complimentary perspectives on the biogeographic history of *Thylamys* when considering subclades individually.

The two reconstructions recover similar and concordant scenarios in several parts of the phylogeny. For example, both reconstructions recover the lowland region east of the Andes (TrS+TeS in **Figure 21** and WRP+ERP in **Figure 22**) as the ancestral region for *Thylamys* + *Lestodelphys*. Also, both reconstructions recover a westward range expansion event over the Andes at the base of the clade that includes *T. elegans*, *T. pallidior*, and *T. tatei*. In other parts of the phylogeny, the Biome-Based and Barrier-

Based reconstructions offer different, but largely complimentary, perspectives on the evolution of ranges within *Thylamys*, and this may be due in part to the nested nature of the area coding schemes. The Barrier-Based reconstruction identifies two cladogenesis events that occur at the boundary of two areas (CAA events), whereas the Biome-Based reconstruction recovers four such events.

The one node that includes a CAA event in both reconstructions is at the base of Clade 2, where *T. macrurus* diverges from its sister lineage. For the Biome-Based reconstruction (**Figure 21**), this divergence is modeled as occurring between tropical savanna and montane grassland biomes, whereas for the Barrier-Based reconstruction (**Figure 22**) it occurs across the Río Paraguay, with modern *T. macrurus* populations occurring only east of the Río Paraguay. These two reconstructions are not complimentary. The Barrier-Based "East of the Río Paraguay" area only includes the tropical savanna biome, so *T. macrurus* will necessarily include tropical savanna in its reconstruction. Why, then, does *T. macrurus* not occur in tropical savanna habitats west of the Paraguay River? It may be that the presence of *T. pusillus* A in intervening habitats is preventing *T. macrurus* from recolonizing tropical savanna habitats that occur west of the Río Paraguay. Or it may be that the Río Paraguay is a formidable dispersal boundary. With the data at hand, it is impossible to distinguish between these two scenarios. It is worth noting, though, that regardless of which scenario is likeliest, the Río Paraguay did not lead to vicariance between *T. macrurus* and *T. pusillus* A—which inhabit the same biome, but are divided by the river—because they are not immediate sister taxa.

Overall, a higher proportion of CAA events associated with biome range boundaries suggests that vicariance across biome boundaries may have played a marginally more important role than physical barriers in the biogeographic history of this genus, but it is interesting to note that most of these cladogenesis events do not conflict across the two reconstructions. These non-overlapping reconstructions can, in a sense, indicate the relative importance of particular area junctures, regardless of the coding scheme. In the Biome-Based reconstruction (**Figure 21**), the split at the base of Clade 1 is reconstructed to have occurred somewhere within the tropical savanna biome. The Barrier-Based reconstruction is more precise (**Figure 22**). The Río Paraná, which forms the western border of the Mesopotamian area and cuts through the tropical savanna biome, is implicated in a CAA event at the base of Clade 1 in the Barrier-Based reconstruction. The importance of the Río Paraná is lost in the Biome-Based reconstruction, highlighting the value of considering physical barriers in addition to habitat type. Of course, one potential way to accommodate both components is to devise a single, more fine-grained area-coding scheme. We chose not to do this, though, because we are particularly interested in the types of inferences made possible by choosing areas based on only one criterion (e.g., biomes).

Because the Biome-Based areas are more fine-grained than the Barrier-Based areas, they might be more useful for identifying important range evolution scenarios at a smaller scale. For example, the Biome-Based reconstruction recovers two independent range expansion events into the montane forest biome in *T. venustus* and *T. sponsorius* haplogroups. Because the Barrier-Based areas group together a wide swath of disparate habitats that include montane grasslands, montane forests, and lowland savannas into the

category “West of the Río Paraguay,” this notable expansion into novel habitats is not captured in the reconstruction.

Range Evolution and Cladogenesis Among Biomes

The results of our Biome-Based reconstruction (**Figure 21**) suggest that all of the earliest branching lineages within *Thylamys* inhabited tropical savanna habitats east of the Andes. Although most didelphid marsupials inhabit lowland rainforest (Voss and Jansa 2009), nearly all *Thylamys* species are distributed in open and dry habitats (Giarla et al. 2010). Jansa et al. (*in review*) found the same pattern in their reconstruction of macrohabitat occupancy across a phylogenetic tree that included all recognized genera in Didelphidae. Their results indicate that the ancestor of the tribe Thylamyini (which includes *Thylamys* + *Lestodelphys* and the related genera *Chacodelphys, Cryptonanus*, *Gracilinanus*, and *Marmosops*) inhabited closed-canopy forests and that colonization of open habitats east of the Andes occurred before *Lestodelphys* and *Thylamys* split.

The geology of lowland savanna habitats in central and southern South America reveals a dynamic series of events that led to the emergence of the major ecoregions of the area: the Chaco and Cerrado. Much of lowland Argentina and Bolivia was inundated by a marine incursion during the Miocene (Hernandez et al. 2005; Hulka et al. 2006; Werneck 2011). After the incursion receded in the late Miocene, lowland soils were laden with salt and the major floristic elements of the Chaco began to appear in the Pliocene (Werneck 2011). The Cerrado has been estimated to be of a similar age, with time-scaled plant phylogenies suggesting that savanna grass lineages characteristic of the Cerrado first started to diverge from their sister lineages approximately 4 Ma (Simon et al. 2009).

Paleodistribution modeling and palynological records suggests that montane forests of the central Andes would have existed slightly further north than they are currently distributed, instead being found along the north- and east-facing Andean slopes of central Bolivia and southern Peru during cool/dry periods of the Pleistocene (Werneck et al. 2011). However, plant fossils with a maximum age of 6 – 7 Ma from high-elevation sites in the central Andes (which were approximately 2000 m lower then than they are now) include taxa that are now commonly found in cloud forest (Graham 2011). This suggests that the Andean slopes were forested and somewhat mesic before the onset of the Pliocene, which is, according to our Biome-based reconstruction, several million years before the first *Thylamys* lineage would appear in that biome.

In our Biome-Based reconstruction (**Figure 21**), the first *Thylamys* lineage to colonize montane grasslands did so between ca. 2.75 and 1.5 Ma, an interval that occurs well after the first palynological evidence for those habitats appearing in the central Andes (Graham 2009). The Pacific coast of northern Chile and southern Peru has had arid to hyper-arid climates as far back as 15 Ma (Houston and Hartley 2003) well within the timeframe of *Thylamys* diversification. South of the hyper-arid deserts, the Mediterranean climate and Chilean Matorral biome arose in the late Pliocene (Armesto et al. 2007), before the appearance of *T. elegans* (the only representative of the genus to inhabit this biome). Recently, Guerrero et al. (2013) found that lizard and plant clades colonized the Pacific coastal deserts long after arid and hyper-arid conditions first appeared, suggesting that a long evolutionary lag time is necessary before desert habitats can be successfully invaded. Moreover, most colonization events took place in the past two million years, the same timeframe over which *Thylamys* first invaded these regions.

The colonization of novel habitats like the Pacific coastal deserts also has important implications for species interactions and the possibility for sympatry. Across our Biome-Based reconstruction, we find examples of lineages colonizing the same biome but not occurring in sympatry (e.g., *T. pallidior* B and *T. pusillus* C in the tropical savanna) and examples in which the two species do ultimately occur in sympatry (e.g., *T. venustus* B and *T. sponsorius* A after invading the montane forest biome). The ecological and evolutionary factors that underlie these different outcomes could be an important topic of future study.

Range Evolution in the Lowlands: The Impact of Rivers as Vicariant Forces

The Río Paraguay, Río Uruguay, and Río Paraná all likely arose in the Miocene or earlier (Potter 1997), so it is plausible that these rivers could have acted as dispersal barriers or vicariant forces over the whole evolutionary history of extant *Thylamys* lineages. The Río Paraguay is implicated in the cladogenesis event that isolates *T. macrurus* from its sister lineage at the base of Clade 2 approximately 1.25 Ma (**Figure 22**), and the Río Paraná is reconstructed as the location of the split at the base of Clade 1 approximately 0.75 Ma. As discussed earlier, these two CAA events are not captured in the Biome-Based reconstruction.

Compared to the Amazonian region, relatively few studies have considered the effect of rivers on the diversification of open-habitat species like *Thylamys*. In the open Cerrado and Caatinga biomes of central Brazil, divergent mitochondrial haplogroups within the rodent species *Calomys expulsus* were found on either side of the Río São Francisco, preliminary evidence that the river may have played a role in allopatric divergence among populations (Nascimento et al. 2011). In contrast, several studies have

found that the Río Paraguay and Río Paraná are not significant barriers to dispersal for small mammals (Myers 1982; Cáceres 2007). Nonetheless, based on our current knowledge of species distributions within *Thylamys*, both of these rivers do appear to coincide with distributional boundaries, and our Barrier-Based reconstruction suggests that they played a role in at least two cladogenic events.

Range Evolution Within and Over the Andes.

The origin of Andean fauna has been discussed extensively by Neotropical zoogeographers, with two prevalent theories emerging: that species evolved *in situ* among Andean habitats (through either allopatric or parapatric speciation) or that species colonized the Andes multiple times from the lowlands (Descimon 1986; Lynch 1986; Hall 2005; Brumfield and Edwards 2007; Ribas et al. 2007; Fjeldså et al. 2012). Our Biome-Based reconstruction does not support the idea that Andean taxa arose from multiple lowland colonists. Instead, a single lineage is reconstructed to have invaded montane grassland habitats. That lineage then diversified, with some daughter lineages invading montane forests or crossing over the Andean cordillera into the arid Pacific coastal region. The Andes may have in fact been a source for lowland diversity. The ancestral lineage that eventually gives rise to several contemporary lowland lineages (e.g., *T. macrurus* and the *T. pusillus* haplogroups) is reconstructed to have had a widespread range that includes the montane grassland biome (**Figure 21**). This finding supports the theory that the Andean regions are centers of endemism that act as "species pumps," seeding lowland areas with the descendants of populations that first diverged in the mountains (Fjeldså and Rahbek 2006; Sedano and Burns 2010).

All species within Clade 3 are primarily distributed in Andean habitats, making it a particularly interesting group to consider from a biogeographic perspective. Clade 3 originated less than 1.0 Ma, when the central Andes were ca. 100 m lower in elevation than at present (under a 0.3 mm/year uplift rate for the central Andes; Gregory-Wodzicki 2000) but still likely a significant physical obstacle to dispersal. Our two biogeographic regionalization schemes result in different implied range expansions across the Andes. Surprisingly, the Barrier-Based reconstruction downplays the role of the Andes in driving cladogenesis within *Thylamys* (**Figure 22**). Instead of divergences being reconstructed to have occurred across the Andes, all of the splits within Clade 3 occur within either the "Pacific Coast" area or the "West of the Río Paraguay" area. Clade 3 includes two distinct CAA events under the Biome-Based reconstruction (**Figure 21**), but none in the Barrier-Based reconstruction. Because montane grassland habitat is restricted almost entirely to the east of the western cordillera of the Andes (Olson et al. 2001; Houston and Hartley 2003), the Biome-Based reconstruction for Clade 3 also includes three independent trans-Andean dispersal events. The Barrier-Based reconstruction includes only one such dispersal, which implies that populations maintained widespread distributions on both sides of the Andes through multiple within-area cladogenesis events. Although a widespread ancestral lineage that somehow maintains cohesiveness across both sides of the Andes seems unlikely given the potential dispersal impediment, one modern taxon (*T. pallidior* haplogroup A) does in fact maintain a distribution on both sides of the cordillera.

One clue as to which scenario is more likely—one Andean crossing versus three—could come from population genetic analysis of gene flow over the Andes in *T.*

pallidior A. Increased sampling in concert with tests of gene flow among populations within the haplogroup could reveal the potential for population connectivity in spite of the suspected dispersal barrier. It is possible, for example, that species already inhabiting montane grasslands might move relatively easily through mountain passes and into the western coastal region. Given the current data at hand, however, it is difficult to identify which biogeographic reconstruction is the most likely.

Far from being just a dispersal barrier, the Andes host a diverse set of habitats and are home to several *Thylamys* lineages. Of particular interest are the montane grasslands and the montane forests of the eastern slopes of central Andes. According to the Biome-Based reconstruction, the montane grassland biome was colonized from lowland tropical savanna between ca. 2.75. and 1.5 Ma by a single *Thylamys* lineage that would later diversify into at least 13 distinct mitochondrial haplogroups and species (**Figure 21**). Overall, nine of 15 possible cladogenesis events across the whole *Thylamys* + *Lestodelphys* tree were reconstructed as occurring within the montane forest and montane grassland biomes or between those biomes and adjacent biomes, further highlighting the influence of montane habitats on *Thylamys* diversification. In topographically complex areas like the Andes, dramatic changes in habitat-type over short distances could promote parapatric or ecological speciation. In addition, global climatic events like the Quaternary glacial cycles might have also had a particularly strong effect on montane species, isolating montane forests or grasslands (and the terrestrial vertebrates that inhabit them) in refugia during interglacial periods and promoting allopatric speciation (Martínez Meyer et al. 2004; Kozak and Wiens 2006; Galbreath et al. 2010).

Conclusions

Results from this study reveal a complex series of dispersal, vicariance, and range expansion events over the past four million years of diversification in the marsupial genus *Thylamys* and its sister taxon *Lestodelphys*. Included among these events are multiple independent transitions across the Andean cordillera; cladogenesis across the boundaries of both rivers and biomes; and multiple instances of range expansion into a diverse set of biomes, from the montane forests along the eastern Andean foothills to the hyper-arid coastal deserts of Peru and Chile. Considering two different area-coding schemes—one based on biomes, the other based on physical geography—allowed us to take into account both the ecological factors and the traditional biogeographic barriers that affect cladogenesis, dispersal, and range extinction. Ultimately, both coding-schemes led to largely complimentary scenarios of range inheritance, which suggests that the best area-coding scheme may perhaps be some combination of both schemes. Nevertheless, the piecemeal perspective we offer here allowed us to isolate the potential cladogenic effects of physical barriers to dispersal from habitat-based barriers. Our results suggest that in the lowlands, rivers can be an important factor limiting dispersal and driving divergence. In the montane regions, differentiation among habitat types is key for *in situ* diversification. Invasion of high-elevation grasslands may have been the impetus for further diversification in montane habitats, eventually seeding the adjacent lowland areas with even more species.

Results from this analysis open a wide variety of potential research directions. Many of the events modeled by Lagrange can be cross-validated using different data and analytical techniques. For example, *Thylamys pallidior* haplogroup B is reconstructed to

have expanded its range into lowland regions of the temperate savanna biome. Such a significant range expansion might coincide with an increase in sheer population size that would, in turn, leave a genetic signature traceable by phylogeographic analysis (e.g. Galbreath et al. 2009; Camargo et al. 2013). Further phylogeographic investigations of recent biogeographic event have the potential to reveal the forces behind diversification in *Thylamys* and other taxa from the open and arid regions of the world.

CHAPTER 4

THE ROLE OF GLACIAL CYCLES IN SHAPING PHYLOGEOGRAPHIC PATTERNS OF SIX ANDEAN MARSUPIAL LINEAGES (DIDELPHIDAE: *THYLAMYS*)

INTRODUCTION

The Earth's climatic history is punctuated by periods of climatic instability, such as the global temperature oscillations that dominated the Quaternary period (Gibbard and Cohen 2008). These climatically turbulent periods had a profound impact on life, altering resource availability, the spatial distribution of habitats, and countless other facets of an organism's ecology (Bennett 1990; Hewitt 1996; Roy et al. 1996; Dynesius and Jansson 2000; Jansson and Dynesius 2002; Hewitt 2004). When suitable environmental conditions for a population shift in space, and if the population does not become extinct, it can respond by either staying in place and adapting to its new environment or by tracking its preferred habitat into the new area. The outcome depends on the degree to which the environmental niche of the species is evolutionary conserved, and mounting evidence suggest that recently diverged species (or recently isolated allopatric populations) tend to retain conserved environmental niche characteristics (Wiens et al. 2010; Peterson 2011; Petitpierre et al. 2012; Quintero and Wiens 2013).

Over evolutionary timescales, the joint effects of shifting climates and stable niches can lead to a build-up of phylogeographic structure and, eventually, allopatric speciation (Wiens and Graham 2005), especially in regions with high habitat heterogeneity, such as mountain chains. For example, if a montane population occupies a

relatively conserved niche characterized by a cool climate, once the climate warmed, it would be expected to track its increasingly marginalized habitat into higher elevations and become fragmented among inhospitable valleys (Martínez Meyer et al. 2004). Because the Earth is presently experiencing a warm interglacial period, montane taxa are often found in fragmented populations and exhibit high genetic diversity (Galbreath et al. 2010).

The Pleistocene epoch was a time of global climate fluctuations marked by cool glacial periods and warm interglacials. Most investigations of Pleistocene habitat fragmentation and phylogeographic patterns have been set in lowland habitats, where there has been considerable debate with regard to the age of species and the effect of climate oscillations (Haffer 1974; Fjeldså 1994; Klicka and Zink 1997; Avise et al. 1998; Rull 2008; 2011). Recently, however, there has been a growing effort to examine the role of Pleistocene climate oscillations in shaping species diversity and phylogeographic structure in both temperate and tropical montane systems (Roy 1997; Hewitt 2004; Kozak and Wiens 2006; Shepard and Burbrink 2009; Galbreath et al. 2010; Mouret et al. 2011).

Among global montane regions, the tropical and subtropical Andes of South America are an ideal place to investigate how changing environments during the late Quaternary might have affected diversification, because the region is situated at the nexus of a spectacular diversity of biomes and contains the highest endemic terrestrial vertebrate biodiversity in the world (Myers et al. 2000). The extent to which this remarkable diversity was driven by climate-induced allopatric isolation, as opposed to dispersal from neighboring biomes, is not known. Only a handful of studies have examined phylogeographic patterns of central Andean terrestrial vertebrates (e.g., Marín

et al. 2007; Koscinski et al. 2008; Weir et al. 2008; Cossíos et al. 2009; Chaves et al. 2011), and none have considered more than one lineage from the same region at a time in a comparative framework.

Across much of the eastern slopes of the Andes, high elevation grassland and shrubland plant communities (e.g., páramo, puna, and altiplano) are replaced by cloud forests or seasonally dry forests at lower elevations (Olson et al. 2001). During the last glacial maximum (LGM), approximately 26,000 – 21,000 years before present (Clark et al. 2009), temperatures were 5 – 10 ºC cooler than present in the tropical Andes, and it is not clear whether or not this cooling was accompanied by drier or wetter climates (Baker et al. 2001; 2003; Mourguiart and Ledru 2003). Regardless of the relative aridity at the LGM, evidence from the palynological record suggests that the high grassland/montane forest boundary had shifted downslope at glacial maxima (van der Hammen 1985; Hooghiemstra et al. 1993; Flenley 1998; Quattrocchio et al. 2011), and this could have been driven by temperature changes alone (Baker et al. 2003). Elevation bands occupy more surface area at lower elevations in the mountains, so, at a regional scale, habitats currently restricted to small areas at high elevations would have expanded in area when pushed into lower elevations during previous glacial periods (Urrego et al. 2005). As such, montane habitats likely shrunk in total area as warm interglacial periods replaced cool glacial periods, and vice versa.

The post-LGM contraction of both montane forest and montane grassland habitats may have led to a concomitant reduction in population size among the terrestrial animal species that utilize those habitats (Galbreath et al. 2009). Given this paleoecological context, terrestrial montane animal populations are expected to show signs of population

isolation during the last interglacial period (LIG) 73,900 and 129,800 years ago (Kukla et al., 1997), demographic expansion and range expansion in the cooling period between the LIG and LGM, and finally, demographic decline as a response to post-LGM habitat shifts upslope. Recent methodological advances now allow researchers to test these predictions by estimating and comparing demographic models and phylogeographic diffusion models, a multifaceted approach with the power to disentangle demographic shifts and range shifts (Heled and Drummond 2008; Lemey et al. 2009; 2010; Baele et al. 2012; Camargo et al. 2013).

Of the terrestrial vertebrates, opossums (Metatheria: Didelphidae) are among the oldest and least-studied lineages in the tropical and subtropical Andes (Voss and Jansa 2009). Although the majority of didelphids inhabit lowland tropical rainforests, several lineages have successfully invaded montane habitats. Among these, members of the genus *Thylamys*, the fat-tailed mouse opossums, can be found from sea level to 4000 m a.s.l (Giarla et al. 2010). Overall, *Thylamys* species are widely distributed in the southern half of South America, inhabiting savannas, grasslands, deserts, and dry forests from western Peru and eastern Brazil southward to central Chile and Patagonia (Giarla et al. 2010). The montane species in particular (**Figure 23**), however, display many characteristics that make them an ideal study system for investigating the impacts of Pleistocene climate fluctuations on animal populations. A recent taxonomic revision of *Thylamys* based on morphology and mitochondrial DNA revealed at least seven morphologically cryptic mitochondrial haplogroups within three montane species: *Thylamys pallidior*, *T. sponsorius*, and *T. venustus* (Giarla et al. 2010).

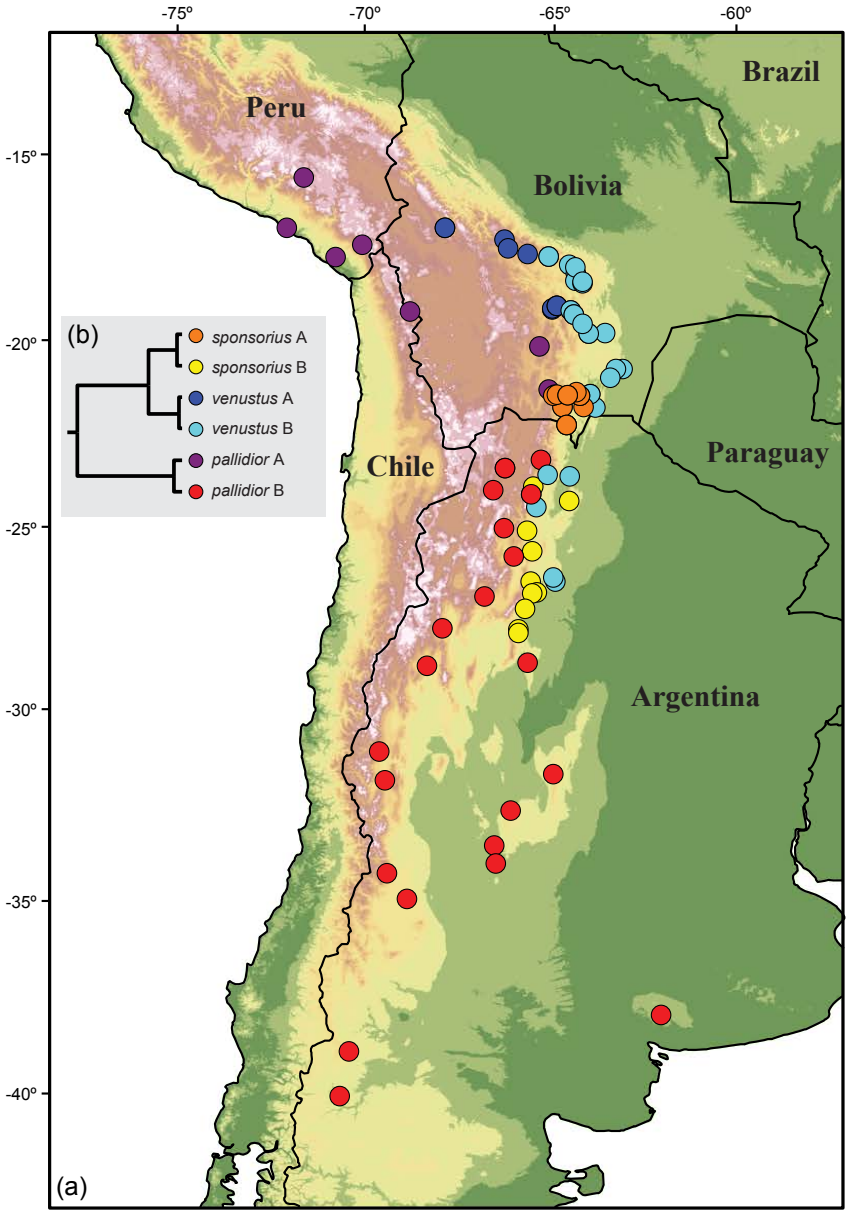

Figure 23. (a) Collecting localities for *Thylamys pallidior, T. sponsorius*, and *T. venustus* haplogroups. Specific locality information for each site is recorded in Giarla et al. (2010), and details on mitochondrial versus nuclear gene sampling per site can be found in Chapter 2. (b) Relationships among the six montane *Thylamys* haplogroups studied here (modified from **Figure 15**. Species tree inferred in *BEAST from 14 anonymous loci and the X-linked intron OGT, with posterior probabilities at each node. Individuals were assigned to haplogroups based on the CYTB results. Two individuals of suspected hybrid origin were not included in this analysis.). Note that here *T. venustus* B comprises the mitochondrial haplogroups *T. venustus* B and *T. venustus* C (Chapter 2).

Coalescent-based species delimitation tests based on 15 nuclear markers supported the long-term persistence of at least six of these haplogroups: two pairs of

genetically isolated cryptic lineages within each of the three focal montane species (Chapter 2). Divergence times were estimated between cryptic sister species as part of a broader phylogenetic analysis (Chapter 3), but error surrounding those estimates is substantial, inhibiting our ability to connect population history to geological or climatic events. Furthermore, there has there been no investigation of the post-LGM population dynamics of these populations or very many co-distributed species (but see Marín et al. 2007; Koscinski et al. 2008; Cossíos et al. 2009). A more thorough understanding of the phylogeography of these three pairs of lineages will illustrate from multiple perspectives the impact of climatic shifts on phylogeographic structure in a montane system.

We hypothesize that these three pairs of *Thylamys* cryptic lineages diverged due to changes in the environment of the central Andes during the Pleistocene. Assuming niche conservatism, montane species such as *T. pallidior*, *T. sponsorius*, and *T. venustus* would be expected to shift their ranges up and down mountain slopes, tracking their preferred habitat, in response to the climate cycles of the Pleistocene (Wiens 2004). The goal of this study is to test two core predictions: (1) that pairs of cryptic montane lineages in *Thylamys* arose in the late Pleistocene, likely during warm interglacial periods, and (2), that presently montane lineages experienced population size and range size expansions during the cooling period before the LGM and population size and range size declines after the LGM. By testing these predictions using comparative phylogeographic data from three pairs of *Thylamys* cryptic lineages, we hope to gain a more thorough understanding of climate-driven evolution across a suite of ecologically similar montane lineages.

MATERIALS AND METHODS

Estimating Divergence Times

We employ three analytical approaches to estimating divergence times across cryptic lineage pairs in *Thylamys*: (1) a fossil-calibrated relaxed-clock phylogenetic approach in BEAST 1.7.5 (Drummond et al. 2012), (2) a mutation-rate-scaled species tree analysis for the six montane lineages using the multispecies coalescent in BEAST (the *BEAST algorithm; Heled and Drummond 2010), and (3) a mutation-rate scaled estimation of divergence times for each of the three pairs of lineages using BPP's two-species model (Yang and Rannala 2010). For the fossil-calibrated approach, we rely on the results from a 26-locus phylogenetic analysis of Didelphidae, which included all *Thylamys* lineages and three marsupial fossil calibration points (Chapter 3). The concatenated alignment contained a combination of mitochondrial, autosomal, and sex-linked loci drawn primarily from Voss and Jansa (2009), Giarla et al. (2010), and Chapter 2. To assess the impact of excluding nuclear loci from divergence time estimation, we reanalyzed sequences from the mitochondrial gene cytochrome *b* (CYTB, which was included among their 26 loci) alone using the same settings as the 26-locus analysis (Chapter 3) and ran the chain for 25 million MCMC generations, sampling every 2,500 steps.

Divergence time estimation with species trees. We conducted a species tree analysis of the six focal lineages (*Thylamys pallidior* A and B, *T. sponsorius* A and B, and *T. venustus* A and B) in BEAST using the same 15 nuclear loci and sampling as in Chapter 2, but with the addition of CYTB sequences from Giarla et al. (2010) for the relevant taxa. Sequences were assigned to "species" in BEAST according to

mitochondrial haplogroup. *Thylamys venustus* haplogroups B and C were combined into a single "species" because they were not recognized as distinct according to results from multilocus species delimitation tests (Chapter 2). We fit nucleotide substitution models in jModelTest 2.1.3 (Darriba et al. 2012) to all 16 alignments, and assigned the best fitting model to each alignment. We used the multispecies coalescent with a Yule process species tree prior and assigned the appropriate ploidy levels to our X-linked and mtDNA loci. All clock models were set to "uncorrelated lognormal," and, to calibrate the species tree analysis to absolute time, the clock model for CYTB was assigned a uniform ucld.mean rate prior ranging from 0.0172 – 0.0668 substitutions per site per million-years. This estimate of the CYTB rate is based on the 95% highest posterior density interval (HPDI) estimated for the CYTB rate for the clade that includes *Thylamys* and its sister-genus *Lestodelphys* from the fossil-calibrated 26-locus analysis in Chapter 3 (**Table 12**). All other clock models received exponential rate priors with a mean of 1.0. The remaining priors were kept at their default values, and we ran two 125 million generation MCMC chains, sampling each every 12,500 steps. We removed 10% of the trees from each run as burn-in, combined the two runs in LogCombiner 1.7.5 (Drummond et al. 2012), assessed effective sample size (ESS) values in Tracer 1.5 (Rambaut and Drummond 2007), and assembled a maximum clade credibility tree from the post-burn-in sample of species trees using TreeAnnotator 1.7.5 (Drummond et al. 2012).

Divergence time estimation with BPP. Each pair of cryptic lineages was analyzed independently using BPP's two species model with the same loci and sampling of individuals as in the species tree analysis described above. In BPP, all seven prior schemes introduced in Chapter 2 were tested for each species pair to assess the impact of

prior specification on divergence time estimates. These schemes varied in the assumed depth of the split (τ prior) and the assumed effective population size (θ prior). Inter-locus rate heterogeneity was allowed under the default Dirichlet prior, and the heredity scalars for the X-linked locus and mitochondrial locus were set to 0.75 and 0.25, respectively.

Table 12. Evolutionary Rates Per Locus for *Thylamys* + *Lestodelphys*. These rates were extracted from the maximum clade credibility tree resulting from analysis of 26-loci and three fossil calibration points (Chapter 3). Only rate estimates for the node that joins *Thylamys* with its monotypic sister genus *Lestodelphys* were used.

Locus	Substitutions/Site/Million-Years		
	95% HPDI (Low)	95% HPDI (High)	Mean
Anon-61	0.0013	0.0058	0.0036
Anon-72	0.0000	0.0012	0.0004
Anon-78	0.0020	0.0067	0.0042
Anon-85	0.0012	0.0069	0.0041
Anon-86	0.0016	0.0057	0.0035
Anon-89	0.0012	0.0072	0.0043
Anon-90	0.0008	0.0096	0.0051
Anon-94	0.0001	0.0064	0.0026
Anon-98	0.0000	0.0186	0.0062
Anon-101	0.0006	0.0208	0.0098
Anon-115	0.0013	0.0145	0.0079
Anon-121	0.0001	0.0053	0.0027
Anon-122	0.0004	0.0289	0.0118
Anon-128	0.0006	0.0026	0.0015
OGT (X-Linked Intron)	0.0025	0.0053	0.0040
CYTB (mtDNA)	0.0172	0.0668	0.0416
Average	**0.0019**	**0.0133**	**0.0071**

We ran the MCMC chain for 2.5 million generations, sampling every five steps, for a posterior distribution of 500,000 samples. 50,000 samples were removed as burn-in, and we assessed chain mixing by examining ESS values in Tracer. For each pair, the 95% HPDI BPP estimate for population splitting time (the τ parameter) was time-calibrated using an estimate for the average substitution rate across the 16 loci used here that were

also used in the 26-locus fossil calibrated analysis (Chapter 3). To do this, we extracted mean rate estimates from the maximum clade credibility tree for the clade that includes *Thylamys* and its sister-genus *Lestodelphys* for each of the 16 loci overlapping with Giarla and Jansa's larger set (**Table 12**).

Modeling Demographic Fluctuations and Range Expansions

Extended Bayesian skyline plots. We modeled demographic changes in each of the six cryptic lineages using extended Bayesian Skyline plots (EBSPs) in BEAST 1.7.5 (Drummond et al. 2012). We extracted six separate, lineage-specific datasets from the 16 alignments used in the species tree analysis described above, and tested nucleotide substitution models for each alignment using jModelTest 2.1.3 (Darriba et al. 2012). In BEAST, we excluded invariant loci, unlinked all site models and clock models, set all clock models to strict clocks, and assigned the best fitting available substitution model to each alignment. For the tree priors, we chose the extended Bayesian skyline plot model with a piecewise-linear population size function, and assigned the appropriate ploidy levels to our X-linked and mtDNA loci. We assigned exponential priors (mean = 1) to all of the ucld.mean priors, kept all other priors at their default values, and ran the MCMC chain for 400 million generations (*T. pallidior* B and *T. sponsorius* B) or 300 million generations (all other lineages).

Demographic model comparisons. We compared constant size and exponential growth coalescent models for each lineage using marginal likelihood estimates calculated via path sampling and stepping stone sampling (PS/SS), two approaches that have been demonstrated to provide accurate and reliable marginal likelihood estimates for model

comparisons (Baele et al. 2012). Overall, we set up 12 analyses in BEAST, one constant size coalescent and one exponential growth coalescent model for each of the six lineages. Site models and clock models were treated the same as for the EBSP analyses, and either the constant size or exponential growth tree prior was chosen. We used the default Laplace prior on the exponential growth rate, which allows the MCMC chain to sample both positive and negative rates. Because PS/SS marginal likelihood estimations can be biased by priors with probability distributions that do not integrate to 1 (Baele et al. 2013), we changed the default "1/*x*" priors for the population size parameters to lognormal priors with a mean of 1.0 and standard deviation of 2.0. The PS/SS calculations rely on well-mixed posterior distribution as "burn-in", so we initialized each marginal likelihood calculation with 50 million MCMC generations, sampling every 5,000 steps. For all analyses, we used Baele et al.'s (2012) XML code for the PS/SS calculations, with a chain length of 100,000 generations and 1,000 path steps.

Phylogeographic diffusion. For the southern "B" lineages that exhibit signs of population expansion (see Results), we modeled the phylogeographic history across a geographic landscape using a continuous-diffusion approach in BEAST (Lemey et al. 2010). We compared two diffusion models using PS/SS: a Brownian diffusion model (BD), in which all branches have the same diffusion rate, and a relaxed random walk model (RRW), in which the diffusion rates for all branches are drawn from a prior distribution. Because this approach is only available for single-locus analyses, we only used CYTB sequences for the lineages of interest. For *T. pallidior* B, we increased geographic sampling by including additional sequences from GenBank. Geographic

coordinates for collecting localities for all individuals were drawn from Giarla et al. (2010) and (for the GenBank sequences) from Chapter 3.

For each analysis, we used the same settings as the EBSP analyses. For the geographic diffusion component of these analyses, we compared a BD model and an RRW model using marginal likelihood estimates based on SS/PS. We ran each analysis for 100 million generations, sampling every 10,000 steps, and examined the log files in Tracer for stationarity and high ESS values. We discarded the first 10% of trees as burn-in and constructed maximum clade credibility trees from the remaining trees in TreeAnnotator. We used the software SPREAD 1.0.6 (Bielejec et al. 2011) to visualize the phylogeographic dispersion of each lineage across the landscape. We used the TimeSlicer function to summarize location parameter the posterior distribution of trees at ten times across each lineage's coalescent history.

RESULTS

Divergence Times

Sequence characteristics for the datasets used in this study can be found in Chapter 2 and Chapter 3. Stationarity and high ESS values were attained for all of the analyses across each of the three analytical approaches used to estimate divergence times (BEAST, *BEAST, and BPP). For the BEAST analysis of CYTB alone, 95% HPD intervals for divergence estimates between haplogroup pairs—*Thylamys pallidior* A vs. B, *T. sponsorius* A vs. B, and *T. venustus* A vs. B—extend across a relatively wide range between approximately 0.25 and 2.0 Ma (**Figure 24**). The concatenated analysis of 26 loci (Chapter 3) resulted in younger and tighter HPD intervals for the timing of the three splits, between approximately 200,000 and 500,000 years ago (**Figure 24**). The

coalescent-based *BEAST and BPP analysis resulted in largely congruent and considerably more recent and tighter HPD intervals around estimates for the timing of each split (**Figure 24**, **Table 13**). We report only the BPP results from Prior Scheme 1 (Chapter 2) because the results from different prior schemes were largely overlapping. For *T. pallidior* A/B and *T. sponsorius* A/B, these estimates overlap with the LIG, which extended from 73,900 and 129,800 years ago (Marine Isotope Stage 5; Kukla et al. 1997). For *T. venustus* A/B, these estimates are more recent still, with HPD intervals for the timing of the split extending between approximately 20,000 and 75,000 years ago, during the cooling period between the LIG and the LGM.

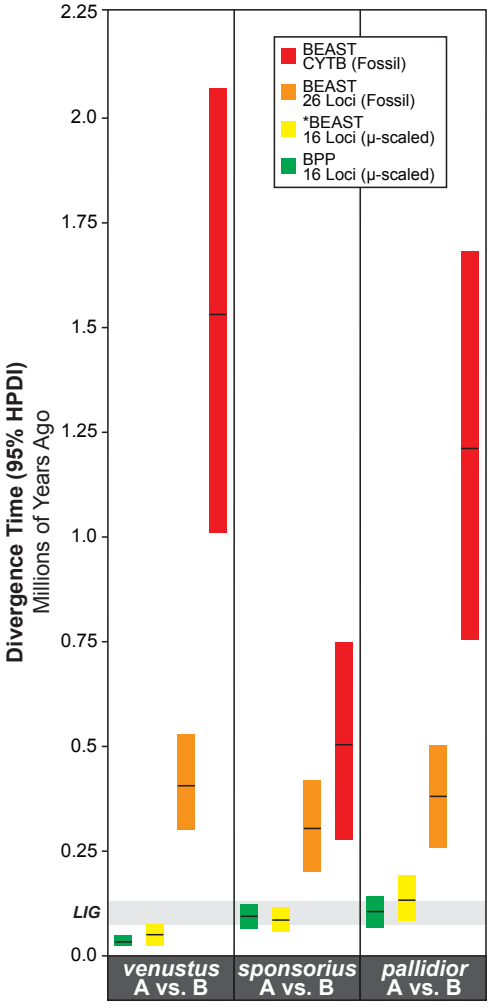

Figure 24. Divergence time estimates for three pairs of montane *Thylamys* lineages using four different datasets and three different methodologies. Colored bars represent 95% highest posterior density intervals. Black lines within the bars are mean divergence time estimates. The shaded horizontal bar illustrates the last interglacial period (Marine Isotope Stage 5; Kukla et al. 1997). In red are results from the fossil-calibrated CYTB BEAST analysis of *Thylamys* and outgroups. In orange are results from the 26-locus, concatenated, fossil-calibrated BEAST analysis from Chapter 3. In yellow are results from the multilocus species tree analysis calibrated with the CYTB mutation rate interval from **Table 12**. In green are results from three separate multilocus BPP analyses calibrated using the average substitution rate from **Table 12**.

Table 13. Divergence Time Estimations from BPP and *BEAST

	***pallidior* A vs. B**		***sponsorius* A vs. B**		***venustus* A vs. B**	
	*BEAST	BPP	*BEAST	BPP	*BEAST	BPP
Mean Divergence (years)	133,000	105,843	85,500	94,481	50,600	33,584
95% HPD Minimum (years)	83,000	67,081	56,100	62,985	27,100	19,912
95% HPD Maximum (years)	192,500	143,624	118,100	126,959	76,300	49,004

Historical Demography and Range Shifts

Each of the six EBSP runs show stationarity and high ESS values in Tracer. For *T. pallidior* A, effective population size gradually increases over the course of its coalescent history (**Figure 25a**). For *T. pallidior* B, the population size remains relatively constant until the end of the LIG, when the population undergoes a 10-fold increase (**Figure 25b**). *Thylamys sponsorius* A (**Figure 25c**), *T. venustus* A (**Figure 25e**), and *T. venustus* B (**Figure 25f**) appear to remain approximately the same size over the course of their demographic history. *Thylamys sponsorius* B has a complex demographic history (**Figure 25d**), showing a nearly ten-fold population size decline from coalescence until approximately 50,000 years ago, when it reverses and rapidly expands back to its former population size. The population-size-changes parameter for *T. sponsorius* B has a median of two, suggesting that the direction of population size variation changed twice over the course of its coalescent history, whereas all other lineages exhibit medians of one (*T. pallidior* A and B*, T. venustus* A and B) or zero changes (*T. sponsorius* A).

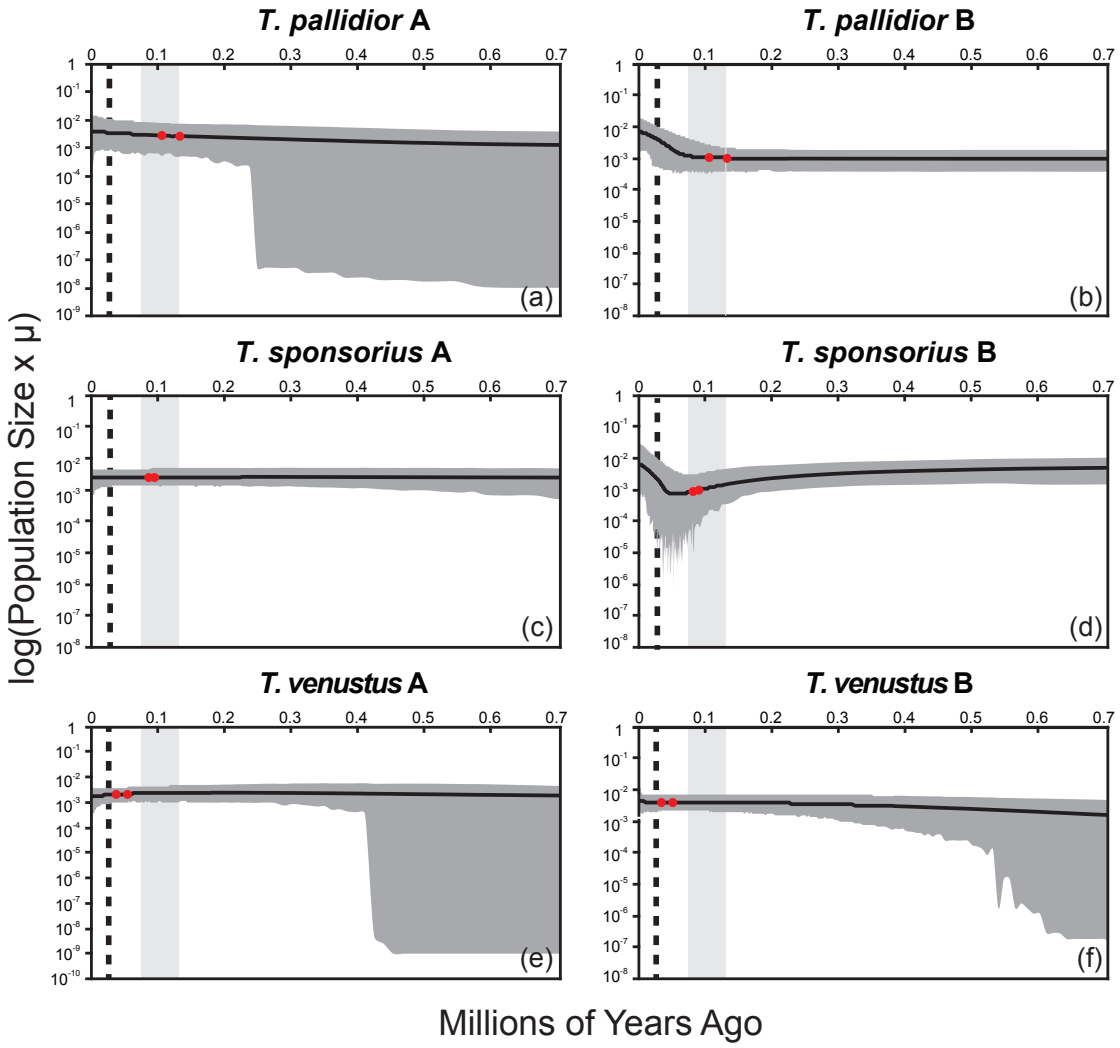

Figure 25. Extended Bayesian skyline plots for *Thylamys pallidior* A (a), *T. pallidior* B (b), *T. sponsorius* A (c), *T. sponsorius* B (d), *T. venustus* A (e), and *T. venustus* B (f). Black lines denote median estimates for population size over time. Dark gray shading represents the 95% highest posterior density interval of the population size over the coalescent history. The vertical light gray bars denote the last interglacial period (Marine Isotope Stage 5; Kukla et al. 1997), and the vertical black dashed lines denote the onset of the last glacial maximum (Clark et al. 2009). The X-axis is time-scaled from substitutions/site/million years to millions of years using the average substitution rate from **Table 12**, whereas the Y-axis remains unscaled.

Constant size and exponential growth coalescent tree priors were compared for each lineage using PS/SS marginal likelihood estimates (**Table 14**). For each analysis, we used both path sampling and stepping stone sampling to estimate the marginal likelihoods, and in each case, both estimators gave identical or nearly identical results.

The exponential growth model received a significantly higher marginal likelihood score as measured by Bayes Factors (BF; Kass and Raftery 1995) for *Thylamys pallidior* A and B, and *T. venustus* B. The constant size model received a significantly higher marginal likelihood for *T. sponsorius* A and *T. venustus* A, and the constant size model could not be rejected for *T. sponsorius* B. There was a large variation in the estimates for the exponential growth rate across runs (**Table 14**).

Table 14. Marginal Likelihood Estimates (MLE) and Exponential Growth Rates (Gr. Rate) from Constant and Exponential Coalescent Models in BEAST. MLEs are based on Stepping Stone Sampling (SS) and Path Sampling (PS). Exponential growth rates are scaled to million-years^{-1} by assuming a mutation rate of 0.0071 substitutions/site/million-years (**Table 12**). 2ln(BF) scores higher than 2 are considered positive evidence against the null hypothesis (here, constant size), and scores higher than 20 are considered very strong evidence against the null hypothesis (Kass and Raferty 1995). The best fitting model (when significant) is underlined and displayed in bold text. "Rate" refers to the estimated growth rate.

***pallidior* A**				***pallidior* B**			
	MLE (SS)	**MLE (PS)**	**Rate**		**MLE (SS)**	**MLE (PS)**	**Rate**
Constant	-10944.40	-10944.12		Constant	-15730.06	-15729.69	
Exponential	-10941.31	-10941.03	207.96%	**Exponential**	-15723.77	-15723.41	623.03%
2ln(BF)	6.18	6.18		2ln(BF)	12.58	12.56	
***sponsorius* A**				***sponsorius* B**			
	MLE (SS)	**MLE (PS)**	**Rate**		**MLE (SS)**	**MLE (PS)**	**Rate**
Constant	-14378.64	-14378.27		Constant	-11999.79	-11999.56	
Exponential	-14383.11	-14382.64	91.87%	Exponential	-12000.33	-12000.07	-89.53%
2ln(BF)	8.94	8.74		2ln(BF)	1.06	1.02	
***venustus* A**				***venustus* B**			
	MLE (SS)	**MLE (PS)**	**Rate**		**MLE (SS)**	**MLE (PS)**	**Rate**
Constant	-14957.52	-14957.20		Constant	-17833.98	-17832.79	
Exponential	-14958.55	-14958.23	89.96%	**Exponential**	-17816.29	-17815.11	143.07%
2ln(BF)	2.06	2.06		2ln(BF)	35.38	35.36	

T. pallidior B (for which the exponential growth model fit the data best) is estimated to have grown at a rate of 623.02% per million years over its coalescent history, where as *T. pallidior* A and *T. venustus* B increased at rates of 207.96% and 143.07% per million years, respectively. Although the exponential growth model did not provide a significantly better fit to the data than the constant growth model for *T.*

sponsorius B, the overall rate is negative, suggesting exponential decay of 89.53% per million years. To explore the possibility that the southern "B" lineages may have expanded southward in range after the LGM, we inferred phylogeographic diffusion models using an expanded CYTB dataset. Given their restricted modern day ranges and poorer population sampling, we did not conduct any additional phylogeographic analyses for the northern "A" lineages. We compared Brownian Diffusion (BD) models and Relaxed Random Walk (RRW) models for each "B" lineage, but only report results from the best fitting model. For *T. pallidior* B, the RRW model fits the data best, and the diffusion model shows an ancestral range in the Andes of central Argentina, with recent expansions both to the north and the south (**Figure 26**). The southward range extension includes colonization of lowland temperate regions. For *T. sponsorius* B, the BD model fits the data best, and the contemporary range is very similar to the inferred ancestral range (**Figure 27**). For *T. venustus* B, the RRW model fits the data best, and the diffusion models illustrate an ancestral Bolivian range with a recent southward expansion southward within the Andes into Argentina (**Figure 28**).

Figure 26. Relaxed random walk continuous phylogeographic model for *T. pallidior* B. Time Slices 1 – 3 are taken from equal intervals over the lineage's range expansion history, with time slice 1 representing the ancestral range. Shaded polygons on the surface of the map represents the 80% highest posterior density intervals for the inferred ancestral range at each node in the time slice, from older (black) to younger (teal). Arcs on the map represent branches in the phylogeny, with red to black shading indicating older to younger branching times, respectively.

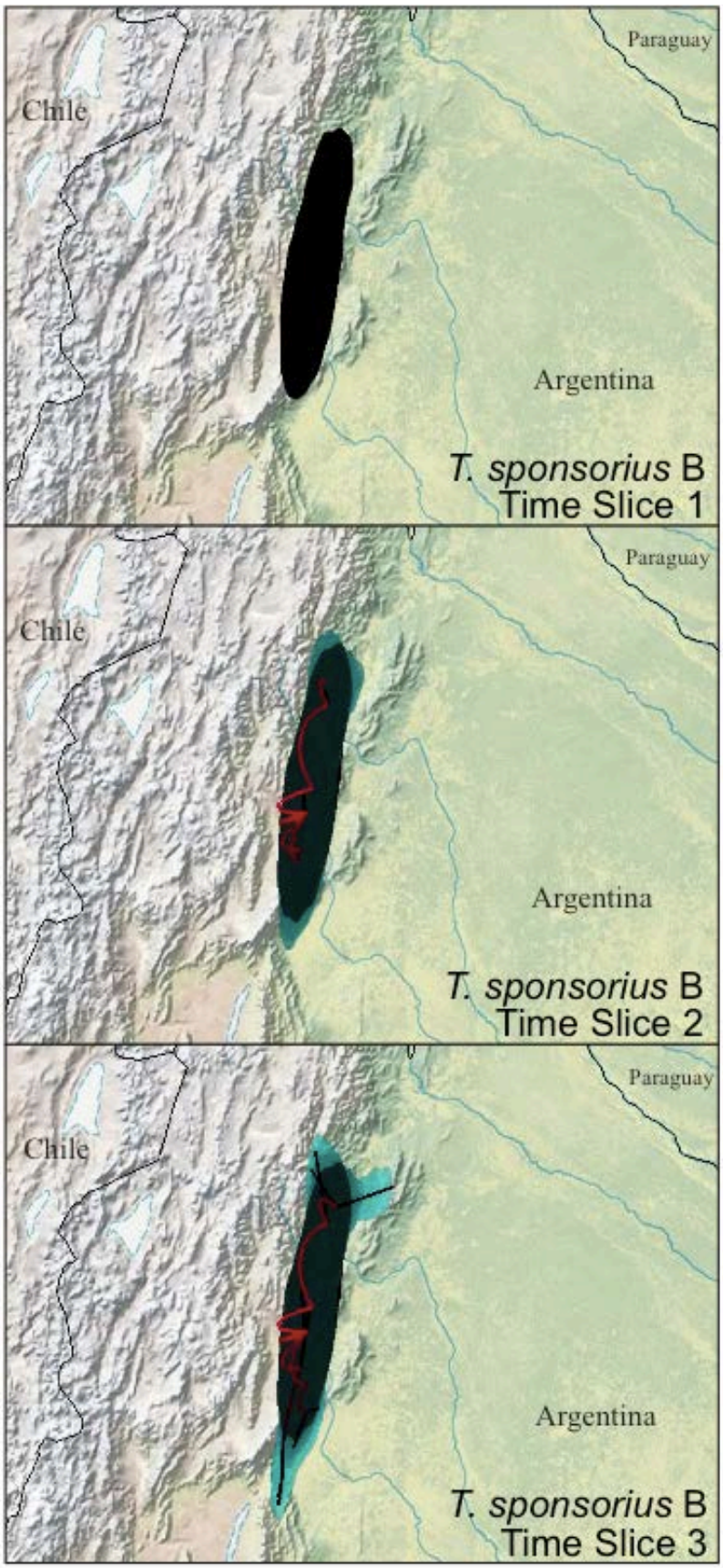

Figure 27. Brownian diffusion continuous phylogeographic model for *T. sponsorius* B. Time Slices 1 – 3 are taken from equal intervals over the lineage's range expansion history, with time slice 1 representing the ancestral range. Shaded polygons on the surface of the map represents the 80% highest posterior density intervals for the inferred ancestral range at each node in the time slice, from older (black) to younger (teal). Arcs on the map represent branches in the phylogeny, with red to black shading indicating older to younger branching times, respectively.

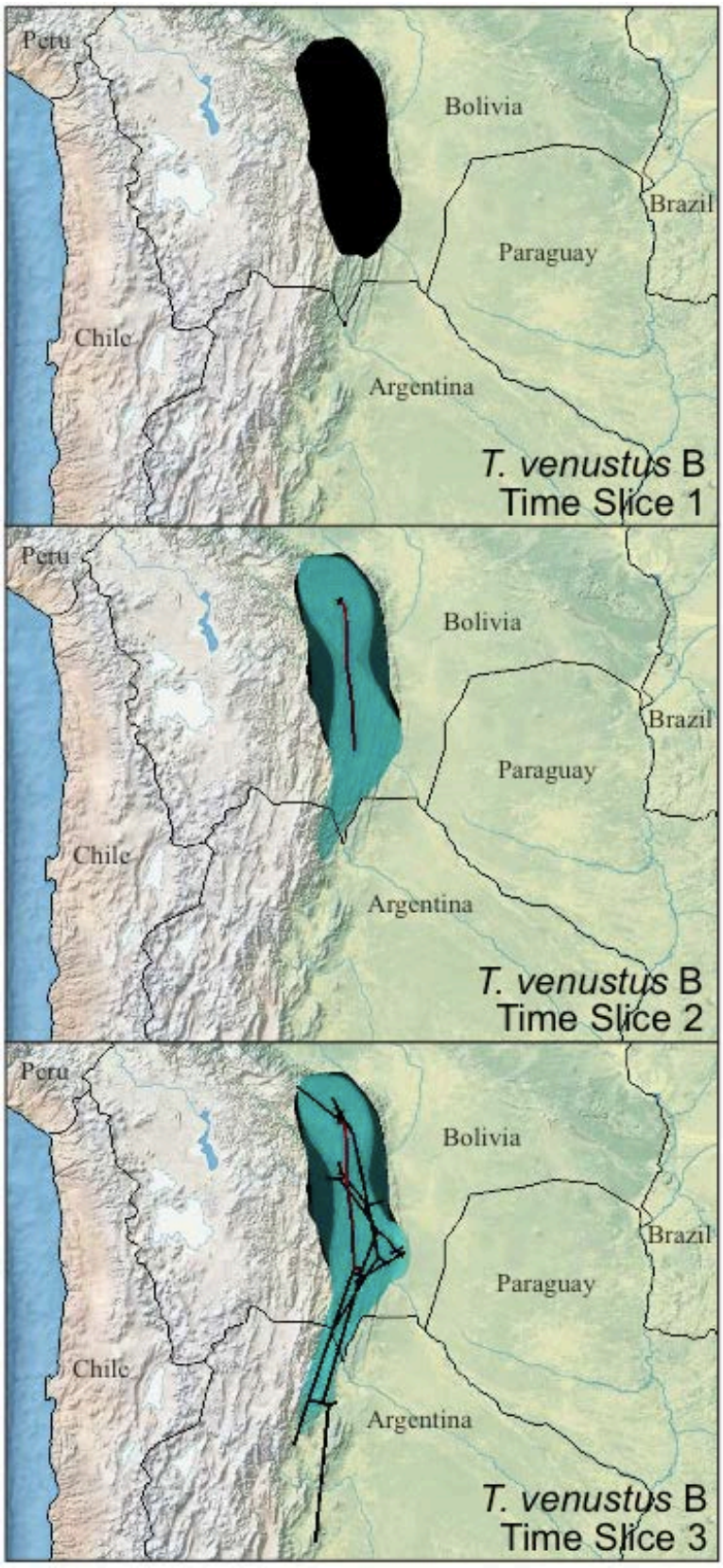

Figure 28. Relaxed random walk continuous phylogeographic model for *T. venustus* B. Time Slices 1 – 3 are taken from equal intervals over the lineage's range expansion history, with time slice 1 representing the ancestral range. Shaded polygons on the surface of the map represents the 80% highest posterior density intervals for the inferred ancestral range at each node in the time slice, from older (black) to younger (teal). Arcs on the map represent branches in the phylogeny, with red to black shading indicating older to younger branching times, respectively.

DISCUSSION

A major goal of phylogeography is to understand how ancient earth dynamics shaped contemporary species distributions and the structure of genetic diversity across a geographic landscape. In montane areas such as the central Andes, where late Quaternary climate oscillations may have repeatedly isolated and reconnected populations, the relationship between population histories and present-day phylogeographic structure is complex. Late Quaternary climate oscillations have been implicated as drivers of phylogeographic patterns in a number of Andean taxa (reviewed by Turchetto-Zolet et al. 2013), but the imprecision of parameter estimates and the lack of statistical tests have hampered the ability of researchers to tie inferred evolutionary changes to geological or climatological events in the earth's history.

Here, we used multiple analytical approaches and large, multilocus datasets to infer population splitting times and population size fluctuations in order to test predictions about the source of the remarkable level of cryptic diversity seen in montane *Thylamys* mouse opossums. For primarily montane lineages like the cryptic haplogroups within *Thylamys pallidior*, *T. sponsorius*, and *T. venustus*, we predicted that populations diverged during the last interglacial (LIG), when climates were relatively warm and similar to present day conditions. Assuming niche conservatism, populations that are presently occurring in high elevation areas were likely fragmented during the LIG. We also predicted that population size and range size of montane species increased following the cooling period after the LIG and reached their most recent maximum around the same time as the earth's last glacial maximum (LGM).

Tying Divergence Times to Historical Climate Shifts

We compared different approaches to estimating divergence times, and, depending on the method used, our results vary in the precision (as measured by 95% HPD intervals) and the mean timing of the estimates (**Figure 24**). The fossil-calibrated analyses of CYTB alone were the least precise, which is not an unexpected outcome for single-locus studies (Edwards and Beerli 2000; Arbogast et al. 2002; Carstens and Knowles 2007; Brito and Edwards 2008). The CYTB divergence time estimates are also substantially older than the multilocus estimates; for the *T. pallidior* and *T. venustus* pairs, the 95% HPD intervals do not overlap with any of the other estimates. The divergence time estimates that we extracted from the concatenated 26-locus analysis (Chapter 3) are more precise than CYTB alone (**Figure 24**). However, the increased precision is likely an artifact of the concatenation process. By stringing together disparate loci into one linked "genealogy," discordance among the genealogies and variation in coalescence times is lost, which can provide misleadingly precise results and bias phylogenetic inference, especially for recently diverged taxa (Edwards and Beerli 2000; Degnan and Rosenberg 2006; Kubatko and Degnan 2007; McCormack et al. 2011; Weisrock et al. 2012). For divergence times, methods that rely on single gene trees (including both single locus and concatenated datasets) will likely lead to significantly older estimates than the actual population splitting time. This is because gene tree divergences must pre-date species tree divergences in the absence of post-divergence gene flow (Edwards and Beerli 2000).

To explicitly incorporate gene tree variation while estimating divergence times, we conducted two sets of analyses. The results from BPP's two-species model and

BEAST's multi-species coalescent model (**Figure 24**, **Table 13**) were essentially in agreement for each species pair: *Thylamys pallidior* A and B and *T. sponsorius* A and B diverged either slightly before or during the LIG, whereas *T venustus* A and B diverged during the cooling period that follows the LIG. The 95% HPD intervals for these divergence time estimates are significantly younger and narrower than the single locus and concatenated analyses. The increased precision afforded by the coalescent-based approaches allows us to confidently link the split between haplogroups within *T. pallidior* and *T. sponsorius* to the LIG, as predicted. Intriguingly, the A and B haplogroups of both of these species are divided in the same geographic region near the political boundary between Argentina and Bolivia (**Figure 23**). The coincident timing and location of this split is evidence that the same event precipitated speciation in both species. The fact that both approaches largely agree that the split occurred during the LIG is in accordance with our predictions about the role of late Quaternary climate oscillations in driving montane speciation. The younger split between *T. venustus* A and B is unlikely to have occurred during the LIG, and therefore does not support our model.

According to a recent review, our results are broadly concordant with population divergences seen in other South American species (Turchetto-Zolet et al. 2013). Overall, most mammalian sister species diverged during the Pleistocene, but only two studies pinpoint the divergences to the period after the LIG (Turchetto-Zolet et al. 2013). The paucity of divergence times estimated to have occurred during the last part of the Pleistocene might be due to the reliance on gene tree divergences instead of species tree estimates, which, as discussed above, tend to overestimate divergence times. In order to accurately estimate divergence times and compare them to codistributed taxa, future

studies must rely on methodologies that explicitly incorporate coalescent variation among loci (Edwards et al. 2007; Yang and Rannala 2010; Heled and Drummond 2010). Otherwise, divergence estimates will almost certainly be biased towards older splitting times (McCormack et al. 2011).

Demographic and Range Shifts within Montane Lineages

Tracing population size shifts over time for the six focal lineages examined here did not reveal a consistent pattern. Instead, each lineage has an idiosyncratic demographic history. The two lineages for which the constant growth model fit best (*T. sponsorius* A and *T. venustus* A; **Table 14**) both occur exclusively in the Bolivian Andes and have the most geographically restricted contemporary ranges of any of the lineages studied here (**Figure 23**). Extended Bayesian skyline plots (**Figure 25c** and **Figure 25e**) support a constant size over the coalescent history of the individuals sampled, although *T. venustus* A appears to decrease in size to a small degree after the LGM. The minor reduction in population size for *T. venustus* A does fit our predictions for a montane population in a warming climate, but the constant population size that precedes it does not concord with our expectations.

Despite a strong pattern of recent demographic expansion in its skyline plot (**Figure 25d**), an exponential growth model does not fit *T. sponsorius* B better than a constant growth model (**Table 14**). Instead, there appear to be two phases of population size change: first, a decrease until the start of the LGM, and, second, an increase from the LGM to present day. A simple exponential growth function cannot adequately explain such a pattern; a more complex, two-epoch model would likely fit the data better (Shapiro et al. 2004). The timing of the demographic shifts does not concord with our

predictions, but it is interesting to note that the trend in population size decline stops right at the end of the LIG, when the global climate started a cooling period. Although the population did not respond to this cooling pattern by expanding immediately, it does appear to stop declining.

An exponential growth model fits *Thylamys pallidior* A, *T. pallidior* B, and *T. venustus* B best (**Table 14**). For *T. pallidior* A and *T. venustus* B, the growth is gradual across the history of both lineages according to the skyline plots (**Figure 25a** and **Figure 25f**). This consistent and gradual growth does not concord with our predictions. In contrast, the skyline plot for *T. pallidior* B shows a sudden increase in population size by at least an order of magnitude during the cooling period between the LIG and LGM (**Figure 25b**), which fits our expectations for how a montane population would behave. However, we also expected to observe a decline in the population size in the warming period that occurs after the LGM, and we see no evidence of such a response. The lack of post-LGM population decline in all of these populations (except, perhaps, for *T. venustus* A) may be because preferable habitat for these populations did not in fact diminish in extent. One possibility is that montane habitats did shrink after the LGM, but environmentally similar habitats began to become available at lower elevations further south. As these populations were pushed further upslope at the northern extent of their range, the southern populations may have been able to expand southward into lower elevation areas (e.g., Patagonia and the lowland Monte desert) that share similar environmental conditions with higher elevation areas to the north. This scenario is borne out in the phylogeographic diffusion models for *T. pallidior* B and *T. venustus* B, which show recent southward range expansions (**Figure 26** and **Figure 28**). The pattern does

not hold for *T. sponsorius* B, however, where the diffusion model suggests a relatively stable range size and no significant expansion southward (**Figure 27**). Overall, the diffusion models suggest that for wide-ranging lineages like *T. pallidior* B and *T. venustus* B, it is important to consider the effects of both latitude and elevation when considering the possibility of range shifts.

Late-Quaternary range expansions in southern South America have been shown in several other terrestrial taxa, including amphibians (Nuñez et al. 2011), squamates (Olave et al. 2011; Camargo et al. 2013), birds (Masello et al. 2011), and other mammals (Marín et al. 2007; Himes et al. 2008; Cossíos et al. 2009; Lessa et al. 2010), suggesting that this pattern might be widespread. However, phylogeographic data from the open and montane habitats of central and southern South America is scarce (Turchetto-Zolet et al. 2013), and future work should focus on clarifying demographic shifts in response to climate change in these regions. Furthermore, accurate estimates of the timing of demographic shifts are essential to testing explicit phylogeographic hypotheses. Here, interpretation of the timing of population shifts in skyline plots is complicated by the fact that lineages were considered individually. The oldest part of the plots in each of the skyline plots extend well beyond BPP and *BEAST divergence time estimates (denoted as red dots on the skyline plots in **Figure 25**). For any given lineage, gene coalescence times are expected to extend well beyond the time each population lineage split from its sister lineage (Edwards and Beerli 2000). Therefore, the ends of the tails do not represent population divergence times. To properly interpret these charts, it is important to conduct coalescent-based divergence time estimations simultaneously, as we demonstrated here,

or to consider more complex multi-lineage models like the Isolation with Migration model (Hey 2010).

Conclusions

Our comparison of divergence times, historical demography, and range shifts among six Andean *Thylamys* lineages suggest that late Quaternary climate oscillations were an important (but not the only) driver of phylogeographic patterns. For the haplogroups within *T. pallidior* and *T. sponsorius*, divergence times are tightly correlated with the LIG as predicted. However, for several lineages our initial predictions about how populations might have responded to climate shifts were not supported, especially with regard to historical demography. Instead of observing population declines following post-LGM warming, we find evidence from multiple lineages of continued population expansion. For two of these lineages (*T. pallidior* B and *T. venustus* B), we attribute this finding to latitudinal range expansions, which we did not initially consider in our predictions but which are just as likely to accompany climate change as altitudinal shifts. By considered more than just one lineage at a time, our results offer a comparative perspective on Andean phylogeography. Even though these lineages are all closely related and inhabit similar areas, the idiosyncratic phylogeographic patterns underscore the complex nature of evolution in montane habitats. Although the shifting climates of the late Pleistocene likely shaped the population structure of each lineage in some way, no single explanation is sufficient to explain the population history of these lineages.

BIBLIOGRAPHY

Abu Baker, M., and J. Brown. 2010. Islands of fear: effects of wooded patches on habitat suitability of the striped mouse in a South African grassland. Functional Ecology 24: 1313–1322.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716–723.

Aleixo, A. 2004. Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. Evolution 58: 1303–1317.

Anderson, S. 1997. Mammals of Bolivia, taxonomy and distribution. Bulletin of the American Museum of Natural History 231: 1–652.

Andruskiw, M., J. Fryxell, I. Thompson, and J. Baker. 2008. Habitat-mediated variation in predation risk by the American marten. Ecology 89: 2273–2280.

Arbogast, B., and G. Kenagy. 2001. Comparative phylogeography as an integrative approach to historical biogeography. Journal of Biogeography 28: 819–825.

Arbogast, B., S. Edwards, J. Wakeley, P. Beerli, and J. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual Review of Ecology and Systematics 707–740.

Armesto, J., M. K. Arroyo, and L. Hinojosa. 2007. The Mediterranean environment. In: T. Veblen, K. Young, & A. Orme (editors), The Physical Geography of South America: 184–199. Oxford: Oxford University Press.

Avise, J., D. Walker, and G. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proceedings of the Royal Society B: Biological Sciences 265: 1707–1712.

Ayres, J., and T. Clutton-Brock. 1992. River boundaries and species range size in Amazonian primates. The American Naturalist 140: 531–537.

Baele, G., P. Lemey, T. Bedford, A. Rambaut, M. Suchard, and A. Alekseyenko. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Molecular Biology and Evolution 29: 2157–2167.

Baele, G., W. Li, A. Drummond, M. Suchard, and P. Lemey. 2013. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Molecular Biology and Evolution 30: 239–243.

Baker, P., G. Seltzer, S. Fritz, R. Dunbar, M. Grove, P. Tapia, S. Cross, H. Rowe, and J Broda. 2001. The history of South American tropical precipitation for the past 25,000 years. Science 291: 640–643.

Baker, P., M. Bush, S. Fritz, C. Rigsby, G. Seltzer, and M. Silman. 2003. Last Glacial Maximum in an Andean cloud forest environment (Eastern Cordillera, Bolivia): Comment and Reply. Geology 31: e26–e27.

Baker, R., and R. Bradley. 2006. Speciation in mammals and the genetic species concept. Journal of Mammalogy 87: 643–662.

Ballard, J., and M. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729–744.

Beheregaray, L., and A. Caccone. 2007. Cryptic biodiversity in a changing world. Journal of Biology 6: 9.

Belfiore, N. 2011. Developing nuclear sequences for species tree estimation in nonmodel organisms: insights from a case study of Bottae's pocket gopher, *Thomomys bottae*. In: L. Knowles and L. Kubatko (editors), Estimating Species Trees: Practical and Theoretical Aspects: 175–191. Hoboken, NJ: John Wiley & Sons, Inc.

Bell, R., J. MacKenzie, M. Hickerson, K. Chavarría, M. Cunningham, S. Williams, and C. Moritz. 2012. Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. Proceedings of the Royal Society B: Biological Sciences 279: 991–999.

Bennett, K. 1990. Milankovitch cycles and their effects on species in ecological and evolutionary time. Paleobiology 11–21.

Berendzen, P., W. Olson, and S. Barron. 2009. The utility of molecular hypotheses for uncovering morphological diversity in the *Notropis rubellus* species complex (Cypriniformes: Cyprinidae). Copeia 2009: 661–673.

Bickford, D., D. Lohman, N. Sodhi, P. L. Ng, R. Meier, K. Winker, K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22: 148–155.

Bielejec, F., A. Rambaut, M. Suchard, and P. Lemey. 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27: 2910–2912.

Bradley, R., and R. Baker. 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. Journal of Mammalogy 82: 960–973.

Braun, J., R. Van Den Bussche, P. Morton, and M. Mares. 2005. Phylogenetic and biogeographic relationships of mouse opossums *Thylamys* (Didelphimorphia, Didelphidae) in southern South America. Journal of Mammalogy 86: 147–159.

Brito, P., and S. Edwards. 2008. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135: 439–455.

Brown, B. 2004. Atlas of New World marsupials. Fieldiana Zoology (new ser.) 102: i–vii, 1–308.

Brown, J., and A. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Systematic Biology 56: 643–655.

Brumfield, R., and A. Capparella. 1996. Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. Evolution 1607–1624.

Brumfield, R., and S. Edwards. 2007. Evolution into and out of the Andes: a Bayesian analysis of historical diversification in *Thamnophilus* antshrikes. Evolution 61: 346–367.

Bull, J., J. Huelsenbeck, C. Cunningham, D. Swofford, and P. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384–397.

Cáceres, N. 2007. Semideciduous Atlantic Forest mammals and the role of the Paraná River as a riverine barrier. Neotropical Biology and Conservation 2: 84–89.

Cáceres, N., R. Napoli, W. Lopes, J. Casella, G. Gazeta. 2007. Natural history of the marsupial *Thylamys macrurus* (Mammalia, Didelphidae) in fragments of savannah in southwestern Brazil. Journal of Natural History 41: 1979–1988.

Camargo, A., F. Werneck, M. Morando, J. Sites Jr, and L. Avila. 2013. Quaternary range and demographic expansion of *Liolaemus darwinii* (Squamata: Liolaemidae) in the Monte Desert of Central Argentina using Bayesian phylogeography and ecological niche modelling. Molecular Ecology 22: 4038–4054.

Camargo, A., M. Morando, L. Avila, and J. Sites Jr. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). Evolution 66: 2834–2849.

Carling, M., and R. Brumfield. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). PLoS ONE 2: e160.

Carmignotto, A., and T. Monfort. 2006. Taxonomy and distribution of the Brazilian species of *Thylamys* (Didelphimorphia: Didelphidae). Mammalia 2006: 126–144.

Carstens, B., and L. Knowles. 2007. Shifting distributions and speciation: species divergence during rapid climate change. Molecular Ecology 16: 619–627.

Carvalho, B., L.B. Oliveira, and M. Mattevi. 2009. Phylogeny of *Thylamys* (Didelphimorphia, Didelphidae) species, with special reference to *Thylamys karimii*. Iheringia (sér. zool.) 99: 419-425.

Chaves, J., J. Weir, and T. Smith. 2011. Diversification in *Adelomyia* hummingbirds follows Andean uplift. Molecular Ecology 20: 4564–4576.

Clark, P., A. Dyke, J. Shakun, A. Carlson, J. Clark, B. Wohlfarth, J. Mitrovica, S. Hostetler, and A. McCabe. 2009. The last glacial maximum. Science 325: 710–714.

Cossíos, D., M. Lucherini, M. Ruiz-García, and B. Angers. 2009. Influence of ancient glacial periods on the Andean fauna: the case of the pampas cat (*Leopardus colocolo*). BMC Evolutionary Biology 9: 68.

Costa, L., Y.R. Leite, and J. Patton. 2003. Phylogeography and systematic notes on two species of gracile mouse opossums, genus *Gracilinanus* (Marsupialia: Didelphidae) from Brazil. Proceedings of the Biological Society of Washington 116: 275–292.

Creighton, G., and A. Gardner. 2008 ("2007"). Genus *Thylamys* Gray, 1843. In A. Gardner (editor), Mammals of South America. Vol. Marsupials, Xenarthrans, Shrews, and Bats: 107–117. Chicago: Chicago University Press.

Darriba, D., G. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: D. Howard and S. Berlocher (editors), Endless Forms: Species and Speciation: 57–75. New York: Oxford University Press.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56: 879–886.

Degnan, J., and N. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics 2: e68.

Degnan, J., and N. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24: 332–340.
Descimon, H. 1986. Origins of lepidopteran faunas in the high tropical Andes. In: F. Vuilleumier & M. Monasterio (editors), High Altitude Tropical Biogeography: 478–499. New York & Oxford: Oxford University Press & American Museum of Natural History.
Drummond, A., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and A. Wilson 2011. Geneious v5 created by Biomatters. Available from http://www.geneious.com.
Drummond, A., M. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1. Molecular Biology and Evolution 29: 1969–1973.
Dupuis, J., A. Roe, and F. H. Sperling. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Molecular Ecology 21: 4422–4436.
Dynesius, M., and R. Jansson. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. Proceedings of the National Academy of Sciences 97: 9115–9120.
Edwards, S., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54: 1839–1854.
Edwards, S., and S. Bensch. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Molecular Ecology 18: 2930–2933.
Edwards, S., L. Liu, and D. Pearl. 2007. High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences 104: 5936–5941.
Ence, D., and B. Carstens. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. Molecular Ecology Resources 11: 473–480.
Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Molecular Biology and Evolution 23: 691–700.
Ferguson, J. 2002. On the use of genetic divergence for identifying species. Biological Journal of the Linnean Society 75: 509–516.
Fjeldså, J. 1994. Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. Biodiversity and Conservation 3: 207–226.
Fjeldså, J., and C. Rahbek. 2006. Diversification of tanagers, a species rich bird group, from lowlands to montane regions of South America. Integrative and Comparative Biology 46: 72–81.
Fjeldså, J., R. Bowie, and C. Rahbek. 2012. The role of mountain ranges in the diversification of birds. Annual Review of Ecology, Evolution, and Systematics 43: 249–265.
Flenley, J. 1998. Tropical forests under the climates of the last 30,000 years. Climatic Change 39: 177–197.

Flores, D., M. Díaz, and R. Barquez. 2000. Mouse opossums (Didelphimorphia, Didelphidae) of northwestern Argentina: systematics and distribution. Zeitschrift für Säugetierkunde 65: 321–339.
Flot, J. 2007. Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. Molecular Ecology Notes 7: 974–977.
Flot, J. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. Molecular Ecology Resources 10: 162–166.
Formoso, A., D. U. Sauthier, P. Teta, and U. J. Pardiñas. 2011. Dense-sampling reveals a complex distributional pattern between the southernmost marsupials *Lestodelphys* and *Thylamys* in Patagonia, Argentina. Mammalia 75: 371–379.
Frost, D., H. Crafts, L. Fitzgerald, and T. Titus. 1998. Geographic variation, species recognition, and molecular evolution of cytochrome oxidase I in the *Tropidurus spinulosus* complex (Iguania: Tropiduridae). Copeia 1998: 839–851.
Fujita, M., A. Leaché, F. Burbrink, J. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology & Evolution 27: 480–488.
Funk, D., and K. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34: 397–423.
Galbreath, K., D. Hafner, and K. Zamudio. 2009. When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American pika, *Ochotona princeps*). Evolution 63: 2848–2863.
Galbreath, K., D. Hafner, K. Zamudio, and K. Agnew. 2010. Isolation and introgression in the Intermountain West: contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). Journal of Biogeography 37: 344–362.
Galtier, N., B. Nabholz, S. Glemin, and G. D. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18: 4541–4550.
Garcia, N., and W. Vargas. 1996. The spatial variability of runoff and precipitation in the Rio de la Plata basin. Hydrological Sciences Journal 41: 279–299.
Gardner, A. 2005. Order Didelphimorphia. In D. Wilson and D. Reeder (editors), Mammal Species of the World, 3rd ed.: 3–18. Baltimore: Johns Hopkins University Press.
Gardner, A., and G. Creighton. 1989. A new generic name for Tate's (1933) Microtarsus group of South American mouse opossums (Marsupialia: Didelphidae). Proceedings of the Biological Society of Washington 102: 3–7.
Garrick, R., P. Sunnucks, and R. Dyer. 2010. Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. BMC Evolutionary Biology 10: 118.
Gascon, C., J. Malcolm, J. Patton, M. F. da Silva, J. Bogart, S. Lougheed, C. Peres, S. Neckel, and P. Boag. 2000. Riverine barriers and the geographic distribution of Amazonian species. Proceedings of the National Academy of Sciences 97: 13672–13677.

Giarla, T., R. Voss, and S. Jansa. 2010. Species limits and phylogenetic relationships in the didelphid marsupial genus *Thylamys* based on mitochondrial DNA sequences and morphology. Bulletin of the American Museum of Natural History 346: 1–67.
Gibbard, P., and K. Cohen. 2008. Global chronostratigraphical correlation table for the last 2.7 million years. Episodes 31: 243.
Glenn, T., and N. Schable. 2005. Isolating microsatellite DNA loci. Methods in Enzymology 395: 202–222.
Graham, A. 2009. The Andes: a geological overview from a biological perspective. Annals of the Missouri Botanical Garden 96: 371–385.
Graham, A. 2011. The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. American Journal of Botany 98: 336–351.
Gregory-Wodzicki, K. 2000. Uplift history of the Central and Northern Andes: a review. Geological Society of America Bulletin 112: 1091.
Guerrero, P., M. Rosas, M. K. Arroyo, and J. Wiens. 2013. Evolutionary lag times and recent origin of the biota of an ancient desert (Atacama-Sechura). Proceedings of the National Academy of Sciences of the United States of America 110: 11469–11474.
Haffer, J. 1974. Avian speciation in Tropical South America, with a systematic survey of the Toucans (Ramphastidae) and Jacamars (Galbulidae). Publications of the Nuttall Ornithological Club, no. 14.
Hall, J. W. 2005. Montane speciation patterns in *Ithomiola* butterflies (Lepidoptera: Riodinidae): are they consistently moving up in the world? Proceedings of the Royal Society B: Biological Sciences 272: 2457–2466.
Hanke, M., and M. Wink. 1994. Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs. BioTechniques 17: 858.
Harold, A., and R. Mooi. 1994. Areas of endemism: definition and recognition criteria. Systematic Biology 43: 261–266.
Hausdorf, B. 2002. Units in biogeography. Systematic Biology 51: 648–652.
Hausdorf, B. 2011. Progress toward a general species concept. Evolution 65: 923–931.
Heled, J., and A. Drummond. 2008. Bayesian inference of population size history from multiple loci. BMC Evolutionary Biology 8: 289.
Heled, J., and A. Drummond. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27: 570–580.
Hernandez, R., T. Jordan, A. Dalenz Farjat, L. Echavarr I a, B. Idleman, and J. Reynolds. 2005. Age, distribution, tectonics, and eustatic controls of the Paranense and Caribbean marine transgressions in southern Bolivia and Argentina. Journal of South American Earth Sciences 19: 495–512.
Hershkovitz, P. 1959. Nomenclature and taxonomy of the Neotropical mammals described by Olfers, 1818. Journal of Mammalogy 40: 337–353.
Hewitt, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247–276.
Hewitt, G. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society B: Biological Sciences 359: 183–195.

Hey, J. 2010. Isolation with migration models for more than two populations. Molecular Biology and Evolution 27: 905–920.
Hickerson, M., G. Dolman, and C. Moritz. 2006. Comparative phylogeographic summary statistics for testing simultaneous vicariance. Molecular Ecology 15: 209–223.
Himes, C., M. Gallardo, and G. Kenagy. 2008. Historical biogeography and post-glacial recolonization of South American temperate rain forest by the relictual marsupial *Dromiciops gliroides*. Journal of Biogeography 35: 1415–1424.
Hird, S., L. Kubatko, and B. Carstens. 2010. Rapid and accurate species tree estimation for phylogeographic investigations using replicated subsampling. Molecular Phylogenetics and Evolution 57: 888–898.
Hooghiemstra, H., J. Melice, A. Berger, and N. Shackleton. 1993. Frequency spectra and paleoclimatic variability of the high-resolution 30–1450 ka Funza I pollen record (Eastern Cordillera, Colombia). Quaternary Science Reviews 12: 141–156.
Houston, J., and A. Hartley. 2003. The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. International Journal of Climatology 23: 1453–1464.
Hudson, R., and J. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56: 1557–1565.
Hulka, C., K. Gräfe, B. Sames, C. Uba, and C. Heubeck. 2006. Depositional setting of the middle to late Miocene Yecua Formation of the Chaco Foreland Basin, southern Bolivia. Journal of South American Earth Sciences 21: 135–150.
Irwin, D., T. Kocher, and A. Wilson. 1991. Evolution of the cytochrome-b gene in mammals. Journal of Molecular Evolution 32: 128–144.
Jansson, R., and M. Dynesius. 2002. The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. Annual review of ecology and systematics 741–777.
Jansa, S.A., Barker F.K., Voss R.S. The early diversification history of didelphid marsupials: A window into South America's "splendid isolation." *Submitted to Evolution*
Kass, R., and A. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90: 773–795.
Kent, W. 2002. BLAT—the BLAST-like alignment tool. Genome Research 12: 656–664.
Klicka, J., and R. Zink. 1997. The importance of recent ice ages in speciation: a failed paradigm. Science 277: 1666–1669.
Koscinski, D., P. Handford, P. Tubaro, S. Sharp, and S. Lougheed. 2008. Pleistocene climatic cycling and diversification of the Andean treefrog, *Hypsiboas andinus*. Molecular Ecology 17: 2012–2025.
Kozak, K., and J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60: 2604–2621.
Kreft, H., and W. Jetz. 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37: 2029–2053.
Krohne, D. 1997. Dynamics of metapopulations of small mammals. Journal of Mammalogy 1014–1026.
Kubatko, L., and J. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56: 17–24.

Kubatko, L., B. Carstens, and L. Knowles. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25: 971–973.
Kukla, G., J. McManus, D.-D. Rousseau, and I. Chuine. 1997. How long and how stable was the last interglacial? Quaternary Science Reviews 16: 605–612.
Kumar, S., and S. Subramanian. 2002. Mutation rates in mammalian genomes. Proceedings of the National Academy of Sciences 99: 803–808.
Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30: 314–334.
Lanfear, R., B. Calcott, S. W. Ho, and S. Guindon. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29: 1695–1701.
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. Bioinformatics 23: 2947–2948.
Leaché, A., and M. Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proceedings of the Royal Society B: Biological Sciences 277: 3071–3077.
Lemey, P., A. Rambaut, A. Drummond, and M. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology 5: e1000520.
Lemey, P., A. Rambaut, J. Welch, and M. Suchard. 2010. Phylogeography takes a relaxed random walk in continuous space and time. Molecular Biology and Evolution 27: 1877–1885.
Lessa, E., G. D'Elía, and U. J. Pardiñas. 2010. Genetic footprints of late Quaternary climate change in the diversity of Patagonian-Fueguian rodents. Molecular Ecology 19: 3031–3037.
Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
Lidicker, W., Jr. 1999. Responses of mammals to habitat edges: an overview. Landscape Ecology 14: 333–343.
Lynch, J. 1986. Origins of the high Andean herpetological fauna. In: F. Vuilleumier & M. Monasterio (editors), High Altitude Tropical Biogeography: 478–499. New York & Oxford: Oxford University Press & American Museum of Natural History.
Maddison, D., and W. Maddison. 2000. MacClade, Version 4.08. Sunderland, Massachusetts: Sinauer Associates.
Marín, J., C. Casey, M. Kadwell, K. Yaya, D. Hoces, J. Olazabal, R. Rosadio, J. Rodriguez, A. Spotorno, M. Bruford, and J. Wheeler. 2007. Mitochondrial phylogeography and demographic history of the Vicuña: implications for conservation. Heredity 99: 70–80.
Martínez Meyer, E., A. Townsend Peterson, and W. Hargrove. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. Global Ecology and Biogeography 13: 305–314.
Masello, J., P. Quillfeldt, G. Munimanda, N. Klauke, G. Segelbacher, H. Schaefer, M. Failla, M. Cortés, and Y. Moodley. 2011. The high Andes, gene flow and a stable

hybrid zone shape the genetic structure of a wide-ranging South American parrot. Frontiers in Zoology 8: 16.
Mayr, E. 1942. Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press.
McCormack, J., J. Heled, K. Delaney, A. Peterson, and L. Knowles. 2011. Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. Evolution 65: 184–202.
McCormack, J., S. Hird, A. Zellmer, B. Carstens, and R. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66: 526–538.
McCulloch, G., G. Wallis, and J. Waters. 2010. Onset of glaciation drove simultaneous vicariant isolation of Alpine insects in New Zealand. Evolution 64: 2033–2043.
McGuire, G., and F. Wright. 2000. TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. Bioinformatics 16: 130–134.
McGuire, J., C. Witt, D. Altshuler, and J. Remsen. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Systematic Biology 56: 837–856.
Meynard, A., R. Palma, and E. Rivera-Milla. 2002. Phylogeographic relationships of the Chilean llacas of the genus *Thylamys* (Marsupialia, Didelphidae) based on sequences of the cytochrome b mitochondrial gene. Revista Chilena de Historia Natural 75: 299–306.
Mikkelsen, T., M. Wakefield, B. Aken, C. Amemiya, J. Chang, S. Duke, M. Garber, A. Gentles, L. Goodstadt, and A. Heger 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature 447: 167–177.
Miller, M., E. Bermingham, J. Klicka, P. Escalante, F. R. do Amaral, J. Weir, and K. Winker. 2008. Out of Amazonia again and again: episodic crossing of the Andes promotes diversification in a lowland forest flycatcher. Proceedings of the Royal Society B: Biological Sciences 275: 1133–1142.
Milne, I., D. Lindner, M. Bayer, D. Husmeier, G. McGuire, D. Marshall, and F. Wright. 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25: 126–127.
Moore, W. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49: 718–726.
Morton, S. 1980. Ecological correlates of caudal fat storage in small mammals. Australian Mammalogy 3: 81–86.
Mouret, V., A. Guillaumet, M. Cheylan, G. Pottier, A. Ferchaud, and P. Crochet. 2011. The legacy of ice ages in mountain species: post-glacial colonization of mountain tops rather than current range fragmentation determines mitochondrial genetic diversity in an endemic Pyrenean rock lizard. Journal of Biogeography 38: 1717–1731.
Mourguiart, P., and M.-P. Ledru. 2003. Last glacial maximum in an Andean cloud forest environment (Eastern Cordillera, Bolivia). Geology 31: 195–198.

Mustrangi, M., and J. Patton. 1997. Phylogeography and systematics of the slender mouse opossum *Marmosops* (Marsupialia, Didelphidae). University of California Publications in Zoology 130: 1–86.

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

Myers, P. 1982. Origins and affinities of the mammal fauna of Paraguay. In: M. Mares & H. Gennoway (editors), Mammalian Biology in South America (Vol. 6): 85–93. Linesville, PA: Pymatuning Laboratory of Ecology.

Nascimento, F. D., L. Pereira, L. Geise, A. R. Bezerra, P. D'Andrea, and C. Bonvicino. 2011. Colonization Process of the Brazilian Common Vesper Mouse, *Calomys expulsus* (Cricetidae, Sigmodontinae): A Biogeographic Hypothesis. Journal of Heredity 102: 260–268.

Nuñez, J., N. Wood, F. Rabanal, F. Fontanella, and J. Sites Jr. 2011. Amphibian phylogeography in the Antipodes: Refugia and postglacial colonization explain mitochondrial haplotype distribution in the Patagonian frog *Eupsophus calcaratus* (Cycloramphidae). Molecular Phylogenetics and Evolution 58: 343–352.

Nylander, J., F. Ronquist, J. Huelsenbeck, and J. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53: 47.

Nylander, J.A. 2004. MrModeltest v2. Program distributed by author, Evolutionary Biology Centre, Uppsala University.

O'Meara, B. 2010. New heuristic methods for joint species delimitation and species tree inference. Systematic Biology 59: 59–73.

Olave, M., L. Martinez, L. Avila, J. Sites Jr, and M. Morando. 2011. Evidence of hybridization in the Argentinean lizards *Liolaemus gracilis* and *Liolaemus bibronii* (Iguania: Liolaemini): an integrative approach based on genes and morphology. Molecular Phylogenetics and Evolution 61: 381–391.

Olson, D., E. Dinerstein, E. Wikramanayake, N. Burgess, G. N. Powell, E. Underwood, J. D'amico, I. Itoua, H. Strand, J. Morrison, C. Loucks, T. Allnutt, T. Ricketts, Y. Kura, J. Lamoreux, W. Wettengel, P. Hedao, and K. Kassem. 2001. Terrestrial ecoregions of the world: a new map of life on earth. BioScience 51: 933–938.

Palma, R., and T. Yates. 1998. Phylogeny of southern South American mouse opossums (*Thylamys*, Didelphidae) based on allozyme and chromosomal data. Zeitschrift für Säugetierkunde 63: 1–15.

Palma, R., E. Rivera-Milla, T. Yates, P. Marquet, and A. Meynard. 2002. Phylogenetic and biogeographic relationships of the mouse opossum *Thylamys* (Didelphimorphia, Didelphidae) in southern South America. Molecular Phylogenetics and Evolution 25: 245–253.

Palumbi, A., and F. Cipriano. 1998. Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. Journal of Heredity 89: 459–464.

Patton, J., and M.F. da Silva. 1997. Definition of species of pouched four-eyed opossums (Didelphidae, *Philander*). Journal of Mammalogy 78: 90–102.

Patton, J., M. F. da Silva, and J. Malcolm. 1994. Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon basin: a test of the riverine barrier hypothesis. Evolution 1314–1323.

Patton, J., M.F. da Silva, and J. Malcolm. 2000. Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin of the American Museum of Natural History 244: 1–306.
Patton, J., S. dos Reis, and M.F. da Silva. 1996. Relationships among didelphid marsupials based on sequence variation in the mitochondrial cytochrome b gene. Journal of Mammalian Evolution 3: 3–29.
Paynter, R., Jr. 1995. Ornithological Gazetteer of Argentina, 2nd Ed. Cambridge: Museum of Comparative Zoology (Harvard University).
Paynter, R., Jr., and M. Traylor, Jr. 1991. Ornithological Gazetteer of Brazil. Cambridge: Museum of Comparative Zoology (Harvard University).
Peterson, A. 2011. Ecological niche conservatism: a time-structured review of evidence. Journal of Biogeography 38: 817–827.
Petitpierre, B., C. Kueffer, O. Broennimann, C. Randin, C. Daehler, and A. Guisan. 2012. Climatic niche shifts are rare among terrestrial plant invaders. Science 335: 1344–1348.
Pluzhnikov, A., and P. Donnelly. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics 144: 1247–1262.
Pons, J., T. Barraclough, J. Gomez-Zurita, A. Cardoso, D. Duran, S. Hazell, S. Kamoun, W. Sumlin, and A. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55: 595–609.
Potter, P. 1997. The Mesozoic and Cenozoic paleodrainage of South America: a natural history. Journal of South American Earth Sciences 10: 331–344.
Quattrocchio, M., W. Volkheimer, A. Borrromei, and M. Martinéz. 2011. Changes of the palynobiotas in the Mesozoic and Cenozoic of Patagonia: a review. Biological Journal of the Linnean Society 103: 380–396.
Quintero, I., and J. Wiens. 2013. Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. Ecology Letters.
Rambaut, A., and A. Drummond. 2007. Tracer v1.5. Computer program available from http://beast.bio.ed.ac.uk/tracer.
Ree, R., and S. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57: 4–14.
Ree, R., B. Moore, C. Webb, and M. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59: 2299–2311.
Reid, N., and B. Carstens. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evolutionary Biology 12: 196.
Ribas, C., A. Aleixo, A. R. Nogueira, C. Miyaki, and J. Cracraft. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Proceedings of the Royal Society B: Biological Sciences 279: 681–689.
Ribas, C., R. Moyle, C. Miyaki, and J. Cracraft. 2007. The assembly of montane biotas: linking Andean tectonics and climatic oscillations to independent regimes of

diversification in *Pionus* parrots. Proceedings of the Royal Society B: Biological Sciences 274: 2399–2408.
Rissler, L., and J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Systematic Biology 56: 924–942.
Rittmeyer, E., and C. Austin. 2012. The effects of sampling on delimiting species from multi-locus sequence data. Molecular Phylogenetics and Evolution 65: 451–463.
Roe, A., A. Rice, S. Bromilow, J. K. Cooke, and F. H. Sperling. 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Molecular Ecology Resources 10: 946–959.
Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46: 195–203.
Ronquist, F., and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. Suchard, and J Huelsenbeck. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.
Roy, K., J. Valentine, D. Jablonski, and S. Kidwell. 1996. Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. Trends in Ecology & Evolution 11: 458–463.
Roy, M. 1997. Recent diversification in African greenbuls (Pycnonotidae: *Andropadus*) supports a montane speciation model. Proceedings of the Royal Society B: Biological Sciences 264: 1337–1344.
Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Methods in Molecular Biology: Bioinformatic Methods and Protocols, Vol. 132: 365–386, S. Misener and S. A. Krawetz (editors). Totowa, NJ: Humana Press Inc.
Rull, V. 2008. Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. Molecular Ecology 17: 2722–2729.
Rull, V. 2011. Neotropical biodiversity: timing and potential drivers. Trends in Ecology & Evolution 26: 508–513.
Särkinen, T., R. Pennington, M. Lavin, M. Simon, and C. Hughes. 2011. Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. Journal of Biogeography 39: 884–900.
Scheen, A.-C., B. Pfeil, A. Petri, N. Heidari, S. Nylinder, and B. Oxelman. 2012. Use of allele-specific sequencing primers is an efficient alternative to PCR subcloning of low-copy nuclear genes. Molecular Ecology Resources 12: 128–135.
Schnitzler, J., T. Barraclough, J. Boatwright, P. Goldblatt, J. Manning, M. Powell, T. Rebelo, and V. Savolainen. 2011. Causes of plant diversification in the cape biodiversity hotspot of south africa. Systematic Biology 60: 343–357.
Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics 6: 461–464.

Sedano, R., and K. Burns. 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). Journal of Biogeography 37: 325–343.

Seroussi, Y., and E. Seroussi. 2007. TraceHaplotyper: using direct sequencing to determine the phase of an indel followed by biallelic SNPs. BioTechniques 43: 452–456.

Shapiro, B., A. Drummond, A. Rambaut, M. Wilson, P. Matheus, A. Sher, O. Pybus, M. Gilbert, I. Barnes, J. Binladen, E. Willerslev, A. Hansen, G. Baryshnikov, J. Burns, S. Davydov, J. Driver, D. Froese, C. Harington, G. Keddie, P. Kosintsev, M. Kunz, L. Martin, R. Stephenson, J. Storer, R. Tedford, S. Zimov, and A. Cooper. 2004. Rise and fall of the Beringian steppe bison. Science 306: 1561–1565.

Shepard, D., and F. Burbrink. 2009. Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander, *Plethodon fourchensis*. Molecular Ecology 18: 2243–2262.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 1114–1116.

Simon, M., R. Grether, De Queiroz, L., C. Skema, R. Pennington, and C. Hughes. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. Proceedings of the National Academy of Sciences 106: 20359–20364.

Sites, J., and J. Marshall. 2003. Delimiting species: a Renaissance issue in systematic biology. Trends in Ecology & Evolution 18: 462–470.

Solari, S. 2003. Diversity and distribution of *Thylamys* (Didelphidae) in South America, with emphasis on species from the western side of the Andes. In: M. Jones, C. Dickman, and M. Archer (editors), Predators with Pouches: The Biology of Carnivorous Marsupials: 82–101. Melbourne: CSIRO Press.

Solari, S. 2007. New species of *Monodelphis* (Didelphimorphia: Didelphidae) from Peru, with notes on *M. adusta* (Thomas, 1897). Journal of Mammalogy 88: 319–329.

Sorenson, M., and J. DaCosta. 2011. Genotyping HapSTR loci: phase determination from direct sequencing of PCR products. Molecular Ecology Resources 11: 1068–1075.

Stephens, M., N. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68: 978–989.

Swofford, D. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods); 4.0 Beta for Macintosh. Sinauer Associates.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4. Molecular Biology and Evolution 24: 1596–1599.

Tate, G. 1933. A systematic revision of the marsupial genus *Marmosa*, with a discussion of the adaptive radiation of the murine opossums (*Marmosa*). Bulletin of the American Museum of Natural History 66: 1–250.

Teta, P., G. D'Elía, D. Flores, and N. de La Sancha. 2009. Diversity and distribution of the mouse opossums of the genus *Thylamys* (Didelphimorphia, Didelphidae) in north-eastern and central Argentina. Gayana 73: 180–199.

Toews, D. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21: 3907–3930.

Turchetto-Zolet, A., F. Pinheiro, F. Salgueiro, and C. Palma-Silva. 2013. Phylogeographical patterns shed light on evolutionary process in South America. Molecular Ecology 22: 1193–1213.

Urrego, D., M. Silman, and M. Bush. 2005. The Last glacial maximum: stability and change in a western Amazonian cloud forest. Journal of Quaternary Science 20: 693–701.

van der Hammen, T. 1985. The Plio-Pleistocene climatic record of the tropical Andes. Journal of the Geological Society 142: 483–489.

Voss, R., and S. Jansa. 2003. Phylogenetic studies on didelphid marsupials II. Nonmolecular data and new IRBP sequences: separate and combined analyses of didelphine relationships with denser taxon sampling. Bulletin of the American Museum of Natural History 276: 1–82.

Voss, R., and S. Jansa. 2009. Phylogenetic relationships and classification of didelphid marsupials, an extant radiation of New World metatherian mammals. Bulletin of the American Museum of Natural History 322: 1–177.

Voss, R., P. Myers, F. Catzeflis, A. Carmignotto, and J. Barreiro. 2009. The six opossums of Felix de Azara: identification, taxonomic history, neotype designations, and nomenclatural recommendations. Bulletin of the American Museum of Natural History 331: 406–433.

Wallace, A. 1852. On the monkeys of the Amazon. Proceedings of the Zoological Society of London 20: 107–110.

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. Theoretical population biology 7: 256–276.

Weir, J. 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution 60: 842–855.

Weir, J., E. Bermingham, M. Miller, J. Klicka, and M. González. 2008. Phylogeography of a morphologically diverse Neotropical montane species, the Common Bush-Tanager (*Chlorospingus ophthalmicus*). Molecular Phylogenetics and Evolution 47: 650–664.

Weisrock, D., S. Smith, L. Chan, K. Biebouw, P. Kappeler, and A. Yoder. 2012. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. Molecular Biology and Evolution 29: 1615–1630.

Werneck, F. 2011. The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. Quaternary Science Reviews 30: 1630–1648.

Werneck, F., G. Costa, G. COLLI, D. Prado, and J. Sites Jr. 2011. Revisiting the historical distribution of seasonally dry tropical forests: new insights based on palaeodistribution modelling and palynological evidence. Global Ecology and Biogeography 20: 272–288.

Wiens, J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution 58: 193–197.
Wiens, J., and C. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annual Review of Ecology, Evolution, and Systematics 36: 519–539.
Wiens, J., and T. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Systematic Biology 51: 69–91.
Wiens, J., D. Ackerly, A. Allen, B. Anacker, L. Buckley, H. Cornell, E. Damschen, T. Davies, J.-A. Grytnes, S. Harrison, B. Hawkins, R. Holt, C. McCain, and P. Stephens. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters 13: 1310–1324.
Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences 107: 9264–9269.
Zamudio, K., and H. Greene. 1997. Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for Neotropical biogeography, systematics, and conservation. Biological Journal of the Linnean Society 62: 421–442.
Zhang, C., D. Zhang, T. Zhu, and Z. Yang. 2011. Evaluation of a Bayesian coalescent method of species delimitation. Systematic Biology 60: 747–761.
Zink, R., and G. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17: 2107–2121.
Zwickl, D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation. University of Texas at Austin.

APPENDICES

APPENDIX 1: LIST OF SEQUENCED SPECIMENS

				Base Pairs Sequenced		
Species	**Tissue**	**Voucher**	**Locality**[12]	**CYTB**	**ND2**	**COX2**
elegans	NK 27583	MSB 87097	Chile: Coquimbo (88)	1149	1044	684
elegans	NK 95111	MSB 133104	Chile: Metropolitana (89)	1149		
elegans	NK 96791	MSB 133097	Chile: Metropolitana (90)	1149	1044	684
elegans	NK 27606	MSB 87098	Chile: Valparaíso (92)	1149		
elegans	NK 96763	MSB 133095	Chile: Valparaíso (93)	1149		
karimii	APC 1561	MZUSP 32094	Braz.: Tocantins (87)	1149	1044	684
macrurus	APC 932	MZUSP 32094	Braz.: M. Grosso Sul (86)	1149		
macrurus	NK 27536	MSB 70700	Para.: Concepción (94)	1149	1044	684
pallidior	LTU 77	CNP 1919	Arg.: Buenos Aires (1)	1149		
pallidior	OCGR 3703	OMNH 29964	Arg.: Catamarca (3)	1149	1044	684
pallidior	OCGR 7196	OMNH 34903	Arg.: Catamarca (5)	1149		
pallidior	OCGR 3763	OMNH 32556	Arg.: Catamarca (7)	1149		
pallidior	AC 47	MLP 24.X.01.3	Arg.: Córdoba (8)	1149		
pallidior	OCGR 3596	OMNH 29963	Arg.: Jujuy (11)	1149		
pallidior	OCGR 2153	OMNH 29957	Arg.: Jujuy (13)	1149		
pallidior	OCGR 7390	OMNH 34911	Arg.: Jujuy (16)	1149		
pallidior	OCGR 4180	[uncataloged]	Arg.: La Rioja (19)	1149		
pallidior	OCGR 43	OMNH 23482	Arg.: Mendoza (20)	1149	1044	684
pallidior	OCGR 230	OMNH 23480	Arg.: Mendoza (21)	1149		
pallidior	UP 397	CNP 1921	Arg.: Neuquén (23)	1149		
pallidior	OCGR 3120	[uncataloged]	Arg.: Neuquén (24)	1149		
pallidior	OCGR 7279	OMNH 34908	Arg.: Salta (25)	1149		
pallidior	OCGR 7294	OMNH 34909	Arg.: Salta (25)	1149		
pallidior	OCGR 3957	OMNH 32559	Arg.: Salta (26)	1149		
pallidior	OCGR 3996	OMNH 32544	Arg.: Salta (28)	1149		
pallidior	OCGR 343	OMNH 23485	Arg.: San Juan (30)	1149		
pallidior	OCGR 322	OMNH 32571	Arg.: San Juan (31)	1149		
pallidior	OCGR 624	OMNH 23489	Arg.: San Luis (32)	1149		
pallidior	OCGR 904	OMNH 23490	Arg.: San Luis (33)	1149		
pallidior	OCGR 460	OMNH 23488	Arg.: San Luis (34)	1149		
pallidior	NK 14721	MSB 57003	Bol.: Chuquisaca (45)	1149		

[12] Numbers in parentheses under "Locality" refer to listings in the gazetteer (Appendix 2).

pallidior	NBH 76-97	FMNH 162495	Bol.: Tarija (73)	1149	1044	684
pallidior	NK 23533	MSB 87099	Bol.: Tarija (81)	1149	1044	684
pallidior	NK 96072	MSB 133108	Chile: Tarapacá (91)	1149		
pallidior		MVZ 145531	Peru: Arequipa (100)	1149		
pallidior		MVZ 173937	Peru: Arequipa (99)	1149		
pallidior		MVZ 143696	Peru: Tacna (104)	1149	283	178
pallidior		MVZ 115634	Peru: Tacna (105)	1149		
pusillus	OCGR 1525	OMNH 23483	Arg.: Catamarca (2)	1149	1044	684
pusillus	OCGR 3770	OMNH 32562	Arg.: Catamarca (4)	1149		
pusillus		BMNH 98.8.19.12	Arg.: Corrientes (9)	1149		
pusillus	LTU 539	CNP 1920	Arg.: Entre Ríos (10)	1149	1044	684
pusillus		ZSM 1966/70	Arg.: Misiones (22)	139		
pusillus	OCGR 4240	[uncataloged]	Arg.: San. del Estero (35)	1149		
pusillus	OCGR 1984	CML 3198	Arg.: San. del Estero (36)	1149	1044	684
pusillus	NK 12574	MSB 55846	Bol.: Chuquisaca (44)	1149		
pusillus	NK 12539	AMNH 260025	Bol.: Santa Cruz (66)	1149	1044	684
pusillus	NK 25139	MSB 67016	Bol.: Tarija (76)	1149	1044	684
pusillus	NK 25141	AMNH 275440	Bol.: Tarija (77)	1149		
pusillus	TK 65612	TTU 109099	Para.: Alto Paraguay (95)	1149		
pusillus	TK 66476	TTU 109052	Para.: Boquerón (96)	1149	1044	684
sponsorius	OCGR 1363	OMNH 29965	Arg.: Catamarca (6)	1149		
sponsorius	OCGR 3600	OMNH 29974	Arg.: Jujuy (11)	1149		
sponsorius	OCGR 7432	OMNH 34534	Arg.: Jujuy (15)	1149		
sponsorius	OCGR 2144	OMNH 29970	Arg.: Jujuy (17)	1149		
sponsorius	OCGR 3986	OMNH 32548	Arg.: Salta (27)	1149	1044	684
sponsorius	OCGR 3979	OMNH 32545	Arg.: Salta (29)	1149	1044	684
sponsorius	OCGR 1346	OMNH 29967	Arg.: Tucumán (37)	1149		
sponsorius	OCGR 1860	IADIZA 4009	Arg.: Tucumán (38)	1149		
sponsorius	OCGR 2047	OMNH 32566	Arg.: Tucumán (39)	1149		
sponsorius	OCGR 3448	OMNH 29977	Arg.: Tucumán (42)	1149		
sponsorius	OCGR 3929	OMNH 32553	Arg.: Tucumán (43)	1149		
sponsorius		FMNH 29170d	Bol.: Tarija (71)	812		
sponsorius		UMMZ 155836	Bol.: Tarija (71)	609		
sponsorius	BDP 3345	FMNH 162505	Bol.: Tarija (72)	1149	1044	684
sponsorius	NK 23762	MSB 67014	Bol.: Tarija (74)	863		
sponsorius	NK 23763	MSB 67015	Bol.: Tarija (74)	1149		
sponsorius	NK 23901	AMNH 275437	Bol.: Tarija (75)	1149		
sponsorius	NK 23903	MSB 67010	Bol.: Tarija (75)	1149		
sponsorius	NK 23904	MSB 67012	Bol.: Tarija (75)	1149		
sponsorius	NK 23647	MSB 140295	Bol.: Tarija (78)	849		
sponsorius	BDP 3309	FMNH 162507	Bol.: Tarija (79)	1149		
sponsorius	NK 23874	[uncataloged]	Bol.: Tarija (80)	863		

sponsorius	NK 23899	[uncataloged]	Bol.: Tarija (80)	863		
sponsorius	NK 23719	MSB 67009	Bol.: Tarija (83)	863		
tatei		MVZ 135504	Peru: Ancash (97)	1149	283	178
tatei		MVZ 135503	Peru: Ancash (97)	1038		
velutinus		OMNH 37216	Braz.: D. Federal (84)	1149	283	684
velutinus		OMNH 22284	Braz.: D. Federal (85)	1149	283	
venustus	OCGR 2071	OMNH 29952	Arg.: Jujuy (12)	1149		
venustus		AMNH 185323	Arg.: Jujuy (14)	1008		
venustus		AMNH 186948	Arg.: Jujuy (18)	1008		
venustus	OCGR 3553	OMNH 29976	Arg.: Tucumán (40)	1149		
venustus	OCGR 1007	OMNH 29966	Arg.: Tucumán (41)	1149	1044	684
venustus	NK 12575	AMNH 261245	Bol.: Chuquisaca (44)	1149		
venustus	NK 21367	MSB 63261	Bol.: Chuquisaca (46)	1149		
venustus	NK 21368	MSB 63262	Bol.: Chuquisaca (46)	1149		
venustus	NK 21655	MSB 63267	Bol.: Chuquisaca (47)	1149		
venustus	NK 21664	MSB 63268	Bol.: Chuquisaca (47)	1149		
venustus	NK 21515	AMNH 263558	Bol.: Chuquisaca (48)	1149		
venustus	NK 21516	MSB 63269	Bol.: Chuquisaca (48)	1149		
venustus	NK 10879	[uncataloged]	Bol.: Chuquisaca (49)	1149		
venustus	NK 12637	AMNH 261254	Bol.: Chuquisaca (49)	1149	1044	684
venustus	NK 12638	AMNH 261260	Bol.: Chuquisaca (49)	1149		
venustus	NK 12642	AMNH 261255	Bol.: Chuquisaca (49)	1149		
venustus	NK 12552	AMNH 261264	Bol.: Chuquisaca (50)	1149		
venustus	NK 21815	MSB 63264	Bol.: Chuquisaca (51)	1149		
venustus	NK 21556	AMNH 263556	Bol.: Chuquisaca (52)	1149	1044	684
venustus	NK 21546	AMNH 263555	Bol.: Chuquisaca (53)	1149		
venustus	NK 21620	MSB 63272	Bol.: Chuquisaca (54)	1149		
venustus	NK 21621	MSB 63265	Bol.: Chuquisaca (54)	1149		
venustus	NK 21622	MSB 63273	Bol.: Chuquisaca (54)	1149		
venustus	NK 30425	AMNH 275427	Bol.: Cochabamba (55)	1149		
venustus	NK 25027	MSB 87109	Bol.: Cochabamba (56)	1149		
venustus	NK 30437	MSB 87100	Bol.: Cochabamba (56)	1149	1044	684
venustus	NK 30479	AMNH 275429	Bol.: Cochabamba (56)	1149		
venustus	NBH 2020	[uncataloged]	Bol.: Cochabamba (57)	1149	1044	684
venustus	NK 22844	AMNH 275428	Bol.: Cochabamba (58)	1149	1044	684
venustus	NK 22845	MSB 67001	Bol.: Cochabamba (58)	950		
venustus		AMNH 248704	Bol.: La Paz (59)	807		
venustus		AMNH 248705	Bol.: La Paz (59)	1008		
venustus	NK 21237	MSB 63260	Bol.: Santa Cruz (60)	1149		
venustus	NK 12114	AMNH 260030	Bol.: Santa Cruz (61)	1149		
venustus	NK 22815	MSB 87106	Bol.: Santa Cruz (62)	1149		
venustus	NK 22811	[uncataloged]	Bol.: Santa Cruz (63)	1149		

venustus	NK 22813	MSB 87107	Bol.: Santa Cruz (63)	1149		
venustus	NK 22946	MSB 67003	Bol.: Santa Cruz (64)	1149		
venustus	NK 22949	[uncataloged]	Bol.: Santa Cruz (65)	1149		
venustus	NK 22952	AMNH 275433	Bol.: Santa Cruz (65)	1149	1044	684
venustus	NK 23023	MSB 67005	Bol.: Santa Cruz (67)	1149		
venustus	NK 22986	MSB 67004	Bol.: Santa Cruz (68)	1149		
venustus	NK 23347	MSB 67007	Bol.: Tarija (69)	1149		
venustus	NK 23992	MSB 67392	Bol.: Tarija (70)	1149		
venustus	NK 23730	[uncataloged]	Bol.: Tarija (74)	1149		
venustus	NK 23392	MSB 67008	Bol.: Tarija (82)	1149		
venustus	NK 30760	MSB 140320	Bol.: Tarija (82)	1149		
venustus	NK 30761	MSB 140321	Bol.: Tarija (82)	1149		
sp.		MVZ 116614	Peru: Arequipa (98)	1149		
sp.		MVZ 137896	Peru: Ayacucho (101)	1149		
sp.		MVZ 119913	Peru: Lima (102)	493		
sp.		MVZ 137585	Peru: Lima (103)	866		

APPENDIX 2: GAZETTEER OF SEQUENCED SPECIMENS

Below we list all of the localities from which specimens of *Thylamys* were sequenced for this report. Italicized place names are those of the largest political units (states, departments, or provinces) within each country. Geographic coordinates (in decimal degrees) and elevation above sea level (if known, recorded in meters [m] or feet [ft]) are given in parentheses. Except as noted otherwise, geographic coordinates for most Argentinian, Chilean, and Peruvian localities were obtained from collector's labels or online collection databases; Bolivian coordinate data are from Anderson (1997).

ARGENTINA

1. *Buenos Aires,* Campamento Base, Sierra de la Ventana (38.069°S, 62.023°W; 50 m).
2. *Catamarca,* Chumbicha, 0.5 km E Highway 38 on Highway 60 (28.867°S, 66.233°W; 1500 ft).
3. *Catamarca,* 17 km N Barranca Larga (26.853°S, 66.755°W; 3219 m).
4. *Catamarca,* Bella Vista (28.627°S, 65.497°W; 974 m).
5. *Catamarca,* 34.6 km by road W Fiambala (27.704°S, 67.883°W; 7902 ft).
6. *Catamarca,* 5 km S Las Higuerillas on Highway 9 (27.828°S, 65.850°W; 3580 ft).
7. *Catamarca,* 7 km SW Los Morteros (28.639°S, 65.606°W; 1424 m).
8. *Córdoba,* La Tapera, Pampa de Achala (31.622°S, 64.911°W; 1959 m).
9. *Corrientes,* Goya (29.133°S, 59.267°W; 37 m [Paynter, 1995]).
10. *Entre Ríos,* Estancia Santa Ana de Carpinchorí (30.796°S, 58.644°W; 134 m).
11. *Jujuy,* 9 km NW Barcena (24.118°S, 65.500°W; 2655 m).
12. *Jujuy,* Highway 9 at border with Salta, at campground on the way to El Carmen (24.470°S, 65.366°W; 4600 ft).
13. *Jujuy,* 11 km E Humahuaca and 2 km E Pucará on road to Cianzo (23.200°S, 65.243°W; 11,500 ft).

14. *Jujuy,* Santa Bárbara (23.600°S, 65.067°W; 1800 m [Paynter, 1995]).
15. *Jujuy,* 24.8 km E Santa Clara (24.296°S, 64.485°W; 1321 m).
16. *Jujuy,* 8.2 km S Sey (24.013°S, 66.515°W; 4167 m).
17. *Jujuy,* 10 km W Tiraxi on Highway 29 (23.917°S, 65.448°W; 5800 ft).
18. *Jujuy,* Yuto (23.633°S, 64.467°W; 349 m [Paynter, 1995]).
19. *La Rioja,* 15 km N Villa San José de Vinchina (28.717°S, 68.293°W; 1681 m).
20. *Mendoza,* 3 km W Refugio Militar General Alvarado (34.270°S, 69.363°W).
21. *Mendoza,* Salinas del Diamante RR Station (34.967°S, 68.833°W).
22. *Misiones,* Dos de Mayo (27.033°S, 54.650°W; ca. 300 m [Paynter, 1995]).
23. *Neuquén,* Cerrito Piñón, Estancia Collon Cura (40.249°S, 70.632°W; 608 m).
24. *Neuquén,* 0.58 km W and 4.2 km N Cerro Mellizo Sud, Parque Nacional Laguna Blanca (39.050°S, 70.383°W; 1315 m).
25. *Salta,* 16 km S and 1.8 km W Barrancas, along Río de las Burras (23.416°S, 66.206°W; 3521 m).
26. *Salta,* 17 km NW Cachi (25.022°S, 66.238°W; 10,350 ft).
27. *Salta,* 25 km SE La Viña (25.648°S, 65.482°W; 1579 m).
28. *Salta,* Los Sauces (25.787°S, 65.966°W; 5544 ft).
29. *Salta,* 5 km NW Pulares (25.091°S, 65.614°W; 4760 ft).
30. *San Juan,* Castaño Nuevo, 9 km NW Villa Nueva (31.015°S, 69.570°W; 5040 ft).
31. *San Juan,* 8 km W Complejo Astronómico El Leoncito (31.783°S, 69.418°W).
32. *San Luis,* Río Gomes, 7 km E San Francisco del Monte de Oro (32.600°S, 66.059°W; 2800 ft).
33. *San Luis,* 12 km by road N Varela (34.008°S, 66.450°W; 2200 ft).
34. *San Luis,* 15 km E Salinas del Bebedero (33.533°S, 66.488°W; 1350 ft).
35. *Santiago del Estero,* Salinas de Ambargasta, ca. 8 km SE Cerro Rico (29.067°S, 64.633°W; 141 m).
36. *Santiago del Estero,* Virgen del Valle picnic area on Highway 64 (28.133°S, 64.833°W; 2300 ft).
37. *Tucumán,* 5 km N Las Higuerillas on Highway 308 (27.738°S, 65.850°W; 2900 ft).
38. *Tucumán,* Biological Reserve at Horco Molle (26.750°S, 65.350°W; 2400 ft).
39. *Tucumán,* Km 42 on Highway 364 S of San Pedro de Colalao (26.456°S, 65.517°W; 4700 ft).
40. *Tucumán,* Los Chorillos, 13 km NO, limite norte Estancia Los Chorillos (26.350°S, 64.917°W).
41. *Tucumán,* Piedra Tendida, 12 km WNW Burruyacú along Río Cajón (26.459°S, 64.861°W; 2500 ft).
42. *Tucumán,* Reserva La Florida, 7 km W Ibatín, Río Pueblo Viejo (27.192°S, 65.668°W; 515 m).
43. *Tucumán,* 5 km SW Siambón (26.769°S, 65.471°W; 3100 ft).

BOLIVIA

44. *Chuquisaca,* 3.8 km by road E Carandaytí (20.767°S, 63.050°W; 460–480 m).
45. *Chuquisaca,* 68 km by road N Comargo (20.150°S, 65.283°W; 3400 m).
46. *Chuquisaca,* 2 km SW Monteagudo (19.817°S, 63.967°W; 1130 m).
47. *Chuquisaca,* 9 km by road N Padilla (19.300°S, 64.367°W; 2000–2100 m).

48. *Chuquisaca,* 11 km N and 16 km W Padilla (19.200°S, 64.450°W; 2050 m).
49. *Chuquisaca,* Porvenir (20.750°S, 63.217°W; 675 m).
50. *Chuquisaca,* 1.3 km SW Porvenir (20.760°S, 63.229°W; 675 m).
51. *Chuquisaca,* Río Limón (19.550°S, 64.133°W; 1300 m).
52. *Chuquisaca,* 2 km N Tarabuco (19.167°S, 64.933°W; 3250 m).
53. *Chuquisaca,* 4 km N Tarabuco (19.133°S, 64.933°W; 3250 m).
54. *Chuquisaca,* 12 km N and 11 km E Tarabuco (19.067°S, 64.817°W; 2450 m).
55. *Cochabamba,* 1.3 km W Jamachuma (17.533°S, 66.117°W; 2800 m).
56. *Cochabamba,* 7.5 km SE Rodeo (17.683°S, 65.600°W; 3800–4000 m).
57. *Cochabamba,* Tholapujru, camino a Laph'ia (17.300°S, 66.217°W; 11,108 ft).
58. *Cochabamba,* Tinkusiri, 17 km E Totora (17.750°S, 65.033°W; 2950 m).
59. *La Paz,* Caracato (16.983°S, 67.817°W; 2900 m).
60. *Santa Cruz,* Cerro Itahuaticua (19.800°S, 63.517°W; 810 m).
61. *Santa Cruz,* 5 km by road SE Comarapa (17.967°S, 64.483°W; 1695 m).
62. *Santa Cruz,* 15 km NE Quiñe (18.050°S, 64.317°W; 1900 m).
63. *Santa Cruz,* 6 km NNE Quiñe (18.033°S, 64.317°W; 1975 m).
64. *Santa Cruz,* 17 km S Quiñe (18.200°S, 64.300°W; 2100 m).
65. *Santa Cruz,* Río Ariruma, 7 km by road SE Ariruma (18.400°S, 64.317°W; 1750 m).
66. *Santa Cruz,* Tita (18.250°S, 62.100°W; 300 m).
67. *Santa Cruz,* 5.5 km by road NE Vallegrande (18.467°S, 64.133°W; 1800 m).
68. *Santa Cruz,* 5.5 km by road NNW Vallegrande (18.417°S, 64.133°W; 1800 m).
69. *Tarija,* 1 km S Camatindy (21.000°S, 63.383°W; 650 m).
70. *Tarija,* 3 km WNW Carapari (21.800°S, 63.783°W; 850 m).
71. *Tarija,* Carlazo (21.470°S, 64.530°W; 2300–2400 m).
72. *Tarija,* Chuquiaca (21.783°S, 64.094°W; 890 m).
73. *Tarija,* roadside above Cieneguillas (21.317°S, 65.033°W; 3400 m).
74. *Tarija,* 3 km SE Cuyambuyo (22.267°S, 64.550°W; 900 m).
75. *Tarija,* 5 km NNW Entre Ríos (21.483°S, 64.200°W; 1600 m).
76. *Tarija,* Estancia Bolívar (21.633°S, 62.567°W).
77. *Tarija,* 5 km W Estancia Bolívar (21.633°S, 62.617°W).
78. *Tarija,* 4.5 km E Iscayachi (21.483°S, 64.917°W; 3750 m).
79. *Tarija,* ca. 10 km by road W Narvaez (21.383°S, 64.292°W, 2220 m).
80. *Tarija,* 11.5 km N and 5.5 km E Padcaya (21.783°S, 64.667°W; 1900 m).
81. *Tarija,* Serranía Sama (21.450°S, 64.867°W; 3200 m).
82. *Tarija,* Tapecua (21.433°S, 63.917°W; 1500 m).
83. *Tarija,* 1 km E of Tucumilla (21.450°S, 64.817°W, 2500 m).

BRAZIL

84. *Distrito Federal,* Brasilia, Jardin Botãnico (15.783°S, 47.917°W; 1100 m [Paynter and Traylor, 1991]).
85. *Distrito Federal,* 25 km S Brasilia, IBGE (16.009°S, 47.917°W).
86. *Mato Grosso do Sul,* Fazenda Califórnia (20.683°S, 56.867°W; 650 m).
87. *Tocantins,* Rio da Conceição (11.184°S, 46.844°W).

CHILE

88. *Coquimbo,* Limarí, Parque Nacional Fray Jorge (30.667°S, 71.667°W).
89. *Región Metropolitana de Santiago,* Las Condes, San Carlos de Apoquindo (33.404°S, 70.484°W; 3743 ft).
90. *Región Metropolitana de Santiago,* Santiago, Maipu, Rinconada de Maipu (33.495°S, 70.894°W; 555 m).
91. *Tarapacá,* Iquique, Colchane, Enquelga (19.220°S, 68.745°W; 3960 m).
92. *Valparaíso,* La Campana National Park, Palmas de Ocoa (32.950°S, 71.083°W).
93. *Valparaíso,* Zapalla, Quebrada del Tigre (32.560°S, 71.439°W, 280 m).

PARAGUAY

94. *Concepción,* 7 km NE de Concepción, Escuela Agropecuaria (23.350°S, 57.383°W).
95. *Alto Paraguay,* Fortín Pikyrenda (20.083°S, 61.783°W; 420 m).
96. *Boquerón,* Parque Nacional Teniente Enciso (21.050°S, 61.750°W).

PERU

97. *Ancash,* 1 km N and 12 km E Pariacoto (9.501°S, 77.774°W; 8500 ft).
98. *Arequipa,* 3 mi W Atico (16.233°S, 73.650°W; 100 ft).
99. *Arequipa,* 1 km N Chivay (15.631°S, 71.600°W; 3700 m).
100. *Arequipa,* 3 mi N Mollendo (16.976°S, 72.049°W; 100 ft).
101. *Ayacucho,* 15 mi WNW Puquio (14.617°S, 74.339°W; 12,000 ft).
102. *Lima,* 1 mi W Canta (11.467°S, 76.639°W; 8800 ft).
103. *Lima,* 8 mi NE Yauyos (12.368°S, 75.866°W; 9500 ft).
104. *Tacna,* 65 km W Tacna (17.764°S, 70.747°W; 300 ft).
105. *Tacna,* 4 km N Tarata (17.438°S, 70.033°W; 12,800 ft).

APPENDIX 3: PRIMERS USED TO AMPLIFY CYTB, COX2, AND ND2

Each primer name consists of three parts. First, the locus is indicated. Next, the position of the primer relative to the 5′ end of the locus and the orientation of the primer (forward or reverse) is indicated. “F1” and “R1” designations refer to whole-locus forward and reverse primers. The third component of each name refers to the taxon to which each primer was optimized.

Name	Sequence
CYTB-F1-Didelphidae	5′- ATAACCTATGGCATGAAAAACCATTGTTG
CYTB-R1-Didelphidae	5′- GCCTTGTAAGCCAGCAATGAAGG
CYTB-420R-Didelphidae	5′- TGAGGACAAATATCCTTCTGAGGAGC
CYTB-670F-*Thylamys*	5′- GATCCTGTTTCGTGRAGGAA
CYTB-770R-*Thylamys*	5′- TAGGAGACCCTGAYAACTTCAC
CYTB-830R-*Thylamys*	5′- GGCAAATAGAAARTATCACTCTGG
CYTB-1000R-*Thylamys*	5′- ACWGGTTGTCCYCCAATTCA
CYTB-170F-*elegans*	5′- CCTAGCCATACATTACACATCAG
CYTB-260R-*elegans*	5′- CGAAACATTCACGCTAATGGAGC
CYTB-270F-*elegans*	5′- AAACATTCACGCYAATGGAGCYTCAAT

CYTB-340F-*elegans*	5′- CCGAGGACTTTATTATGGATCTTA
CYTB-560F-*elegans*	5′- GCAAAGAATCGGGTAAGTGTAGC
CYTB-620R-*elegans*	5′- TTCCTTCACGAAACAGGATC
CYTB-700F-*elegans*	5′- CACCCCTACTACACTATTAAAGA
CYTB-800R-*elegans*	5′- GCTAACCCACTCAATACCCCTCC
CYTB-900F-*elegans*	5′- CCCAAACAAATTAGGAGGTGT
CYTB-930R-*elegans*	5′- CCACTACTTCATACATCAAACCAAC
CYTB-1030F-*elegans*	5′- CAGCCAATCTTCTAATTTTAACCTG
CYTB-1150R-*elegans*	5′- CTTCAACCAAAATACCCTTAACAT
CYTB-200F-*karimii*	5′- GCTATACACTACACATCGGAC
CYTB-215R-*karimii*	5′- CAGTAGCACATATCTGCCGAGA
CYTB-410F-*karimii*	5′- TAACAGTTATAGCTACCGCATTTG
CYTB-460R-*karimii*	5′- TACAAATCTTCTATCAGCTATTCC
CYTB-630F-*karimii*	5′- CCTGTCTCGTGGAGAAATAGT
CYTB-660R-*karimii*	5′- AACCCAGATTCAGATAAAATCCC
CYTB-810F-*karimii*	5′- ACACCAGCTAACCCACTCAA
CYTB-850R-*karimii*	5′- GCCTATGCAATCCTACGATC
CYTB-990F-*karimii*	5′- TATTCCGACCAATCTCACAAAC
CYTB-1020R-*karimii*	5′- ATCCTAACTTGAATCGGAGGACA
CYTB-501R-*macrurus*	5′- AGTGAATTTGAGGTGGTTTCTCAG
CYTB-530F-*pallidior*	5′- GAATGAATTTGAGGTGGGTTCTCAG
CYTB-940R-*pallidior*	5′- CTTCATACATCAAACCAACGAA
CYTB-320F-*sponsorius*	5′- GCCGAGGGATCTATTACGGA
CYTB-910R-*sponsorius*	5′- AGGAGGAGTTCTAGCACTATTAGC
CYTB-225F-*venustus*	5′- CTGCTTTCTCCTCAGTAGCTCA
CYTB-265R-*venustus*	5′- TATTCATGCTAAYGGRGCTTCT
CYTB-240F-*venustus*	5′- TGCCGAGATGTAAATTTCGGATGA
CYTB-260F-*venustus*	5′- TATTCATGCTAAYGGRGCTTCT
CYTB-285F-*venustus*	5′- GGGRCTTCTATATTTTTTATGTG
CYTB-665R-*velutinus*	5′- TCAGATAAAATCCCCTTCCACCC
CYTB-190F-*velutinus*	5′- CTTAGCCATGCACTACACATC
CYTB-230F-*Lestodelphys*	5′- CGAGACGTAAATTATGGGTGATTAA
COX2-F1-*Thylamys*	5′- ATTTCAAGTCAAYCCCATAACC
COX2-R1-*Thylamys*	5′- CTCCTCAAAACAACATGCCACA
COX2-500F-*Thylamys*	5′- TGAGCAGTTCCATCCTTAGG
ND2-F1-*Thylamys*	5′- CCATACCCCGAAAATGTTGGTT
ND2-R1-*Thylamys*	5′- CGAACGCAAATCGAACGCTT
ND2-790F-*Thylamys*	5′- CYCCACTTACAGGTTTTATACCYAAA

APPENDIX 4: PRIMERS USED TO AMPLIFY ANONYMOUS LOCI AND INTRONS

Name	Sequence
ThyAnon61-F	5’-TGGAAGAGTTGGGGTTCAAAGTGGGT
ThyAnon61-R	5’-TGCTTTCCCATCCATATGCCTTTTGCC
ThyAnon72-F	5’-CCCAGCTAGTGAAGAGGACTGTCACC
ThyAnon72-R	5’-TGTGGGCTGCTGCTCTTATTGGTAGT
ThyAnon78-F	5’-GGGTGTGTGTTGTAATGTGTTGGACA
ThyAnon78-R	5’-TGCGTGAGTCTGTATGTGTCTTTATGCG

ThyAnon85-F	5'-ATGAGAAAGAGATGTCTTGCAAATCCTAACA
ThyAnon85-R	5'-AGCCTTGATTTCTTCATCAGAAAACTGTCTATT
ThyAnon86-F	5'-TCTATGCACTGGGGAATATAAAGGAAGGT
ThyAnon86-R	5'-TTGCTACTGCTAGTGATGCTAATACAATGTT
ThyAnon89-F	5'-TGTGCAATTTTTGTGAATGAAATGTAATGCT
ThyAnon89-R	5'-GGCAGGGACTGGTTTCTTTTTGCCTC
ThyAnon90-F	5'-TGACAGGTGGGTCATTGAGCTATTGTT
ThyAnon90-R	5'-ATGGGCAGGACTAACAAGTTGGGGAA
ThyAnon94-F	5'-TCTGGGAGATGCAGTCCAAGGTTCAAA
ThyAnon94-R	5'-GCCACCAACAATCACTTATGTGGCAGA
ThyAnon98-F	5'-GCAGAATACATGGCACGTACCTGGG
ThyAnon98-R	5'-CCTCAGACACTGTGACCCTAGGCAA
ThyAnon101-F	5'-GGACCAAAGGAATGCAAGTTCAGACCC
ThyAnon101-R	5'-TGGGAGCGGGAGTAAGGGAGTAAGTAA
ThyAnon115-F	5'-AGTGGTATGGTTGGATCAAAGGGCA
ThyAnon115-R	5'-CAGTCACTGTCGTCCACTGTCATAG
ThyAnon121-F	5'-TTCACAGTGATGGTGTCAAGAGAATATCAGG
ThyAnon121-R	5'-AGTTAAAATATGGCACCAGGTAAGGATCTCCA
ThyAnon122-F	5'-TTGAGAAAAGGCAGGGAGGGACCAAAT
ThyAnon122-R	5'-CCTAGGTTCCTCGTGGTAGACACCAAC
ThyAnon128-F	5'-CTTACACCAGGCACCAACTCTGAGACA
ThyAnon128-R	5'-CTCTAAACTGCCATCCCAGGGTCACTC
OGT-F	5'-AAATCATTTCATCGACCTTTCTCAG
OGT-R	5'-ATTCCCTGTAATGGAAAAGCAGC
KREMEN-F1	5'-TGTTGCTGTTGCTACCCCCTAACCT
KREMEN-R1	5'-GGTAGTCCGCCCCATTCACCTGG
P4HB-F1	5'-GCAGTTAAGGTTCACAGTTTCCCT
P4HB-R1	5'-ATCTTGTCCACCACTCTCCAG
PPIC-F1	5'-CCCAAGACTGTGGAGAATTTC
PPIC-R1	5'-CAACAGTAAAGTCTCCACCTTGAA
SLC38-F1	5'-TGGTTTCAGTGGTGCTTATT
SCL38-R1	5'-CAATCAGAAAGAACACCATACACA

APPENDIX 5: SUBSET OF INDIVIDUALS SEQUENCED FOR ANONYMOUS LOCI AND OGT

List of individuals sequenced and collecting localities (details in Appendix 2). Haplogroup assignments are based on the results from analysis of the mitochondrial gene cytochrome b (**Figure 2**). Asterisks indicate *T. venustus* individuals with a possible history of introgression with *T. sponsorius*.

Tissue No.	**Voucher No.**	**Species**	**Haplogroup**	**Country: Province/State**
NBH 76-97	FMNH 162495	*pallidior*	A	Bolivia: Tarija
NK 14721	MSB 57003	*pallidior*	A	Bolivia: Chuquisaca
NK 23533	MSB 87099	*pallidior*	A	Bolivia: Tarija
NK 96072	MSB 133108	*pallidior*	A	Chile: Tarapacá
AC 47	MLP 24.X.01.3	*pallidior*	B	Argentina: Córdoba
Arg 2690/OCGR 2153	OMNH 29957	*pallidior*	B	Argentina: Jujuy
Arg 395/OCGR 343	OMNH 23485	*pallidior*	B	Argentina: San Juan
Arg 43/OCGR 43	OMNH 23482	*pallidior*	B	Argentina: Mendoza
Arg 6392/OCGR 3957	OMNH 32559	*pallidior*	B	Argentina: Salta
Arg 6578/OCGR 7196	OMNH 34903	*pallidior*	B	Argentina: Catamarca
Arg 6683/OCGR 7279	OMNH 34908	*pallidior*	B	Argentina: Salta

Arg 687/OCGR 624	OMNH 23489	*pallidior*	B	Argentina: San Luis
LTU 77	CNP 1919	*pallidior*	B	Argentina: Buenos Aires
UP 397	CNP 1921	*pallidior*	B	Argentina: Neuquén
BDP 3309	FMNH 162507	*sponsorius*	A	Bolivia: Tarija
BDP 3345	FMNH 162505	*sponsorius*	A	Bolivia: Tarija
NK 23647	MSB 140295	*sponsorius*	A	Bolivia: Tarija
NK 23719	MSB 67009	*sponsorius*	A	Bolivia: Tarija
NK 23762	MSB 67014	*sponsorius*	A	Bolivia: Tarija
NK 23763	MSB 67015	*sponsorius*	A	Bolivia: Tarija
NK 23874	[uncataloged]	*sponsorius*	A	Bolivia: Tarija
NK 23899	[uncataloged]	*sponsorius*	A	Bolivia: Tarija
NK 23903	MSB 67010	*sponsorius*	A	Bolivia: Tarija
NK 23904	MSB 67012	*sponsorius*	A	Bolivia: Tarija
Arg 1449/OCGR 1346	OMNH 29967	*sponsorius*	B	Argentina: Tucumán
Arg 1466/OCGR 1363	OMNH 29965	*sponsorius*	B	Argentina: Catamarca
Arg 2233/OCGR 1860	IADIZA 4009	*sponsorius*	B	Argentina: Tucumán
Arg 2460/OCGR 2047	OMNH 32566	*sponsorius*	B	Argentina: Tucumán
Arg 2659/OCGR 2144	OMNH 29970	*sponsorius*	B	Argentina: Jujuy
Arg 4609/OCGR 3600	OMNH 29974	*sponsorius*	B	Argentina: Jujuy
Arg 6179/OCGR 3929	OMNH 32553	*sponsorius*	B	Argentina: Tucumán
Arg 6499/OCGR 3979	OMNH 32545	*sponsorius*	B	Argentina: Salta
Arg 6540/OCGR 3986	OMNH 32548	*sponsorius*	B	Argentina: Salta
NBH 20-20	AMNH 103761	*venustus*	A	Bolivia: Cochabamba
NK 21546	AMNH 263555	*venustus*	A	Bolivia: Chuquisaca
NK 21556	AMNH 263556	*venustus*	A	Bolivia: Chuquisaca
NK 25027	MSB 87109	*venustus*	A	Bolivia: Cochabamba
NK 30425	AMNH 275427	*venustus*	A	Bolivia: Cochabamba
NK 30437	MSB 87100	*venustus*	A	Bolivia: Cochabamba
NK 30479	AMNH 275429	*venustus*	A	Bolivia: Cochabamba
NK 12114	AMNH 260030	*venustus*	B	Bolivia: Santa Cruz
NK 22811	[uncataloged]	*venustus*	B	Bolivia: Santa Cruz
NK 22813	MSB 87107	*venustus*	B	Bolivia: Santa Cruz
NK 22815	MSB 87106	*venustus*	B	Bolivia: Santa Cruz
NK 22844	AMNH 275428	*venustus*	B	Bolivia: Cochabamba
NK 22845	MSB 67001	*venustus*	B	Bolivia: Cochabamba
NK 22946	MSB 67003	*venustus*	B	Bolivia: Santa Cruz
NK 22949	[uncataloged]	*venustus*	B	Bolivia: Santa Cruz
NK 22952	AMNH 275433	*venustus*	B	Bolivia: Santa Cruz
NK 22986	MSB 67004	*venustus*	B	Bolivia: Santa Cruz
Arg 1108/OCGR 1007*	OMNH 29966	*venustus*	C	Argentina: Tucumán
Arg 2496/OCGR 2071	OMNH 29952	*venustus*	C	Argentina: Jujuy
NK 12575	AMNH 261245	*venustus*	C	Bolivia: Chuquisaca
NK 12637	AMNH 261254	*venustus*	C	Bolivia: Chuquisaca
NK 21237	MSB 63260	*venustus*	C	Bolivia: Santa Cruz
NK 21368	MSB 63262	*venustus*	C	Bolivia: Chuquisaca
NK 21516	MSB 63269	*venustus*	C	Bolivia: Chuquisaca
NK 23023	MSB 67005	*venustus*	C	Bolivia: Santa Cruz
NK 23392	MSB 67008	*venustus*	C	Bolivia: Tarija
NK 23992*	MSB 67392	*venustus*	C	Bolivia: Tarija