The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2014, **113**, 1136–1151. With 5 figures

Delimiting species in the genus *Otospermophilus* (Rodentia: Sciuridae), using genetics, ecology, and morphology

MARK A. PHUONG[1]*, MARISA C. W. LIM[1], DANIEL R. WAIT[1], KEVIN C. ROWE[1,2] and CRAIG MORITZ[1,3]

[1]*Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Science Building, Berkeley, CA 94720, USA*
[2]*Sciences Department, Museum Victoria, Melbourne, VIC 3001, Australia*
[3]*Research School of Biology, The Australian National University, Acton, ACT 0200, Australia*

Received 16 April 2014; revised 6 July 2014; accepted for publication 7 July 2014

We apply an integrative taxonomy approach to delimit species of ground squirrels in the genus *Otospermophilus* because the diverse evolutionary histories of organisms shape the existence of taxonomic characters. Previous studies of mitochondrial DNA from this group recovered three divergent lineages within *Otospermophilus beecheyi* separated into northern, central, and southern geographical populations, with *Otospermophilus atricapillus* nested within the southern lineage of *O. beecheyi*. To further evaluate species boundaries within this complex, we collected additional genetic data (one mitochondrial locus, 11 microsatellite markers, and 11 nuclear loci), environmental data (eight bioclimatic variables), and morphological data (23 skull measurements). We used the maximum number of possible taxa (*O. atricapillus*, Northern *O. beecheyi*, Central *O. beecheyi*, and Southern *O. beecheyi*) as our operational taxonomic units (OTUs) and examined patterns of divergence between these OTUs. Phenotypic measures (both environmental and morphological) showed little differentiation among OTUs. By contrast, all genetic datasets supported the evolutionary independence of Northern *O. beecheyi*, although they were less consistent in their support for other OTUs as distinct species. Based on these data, we support the conclusions from a previous study that synonymized *O. atricapillus* with *O. beecheyi*, and we elevate the northern lineage of *O. beecheyi* to a separate species. © 2014 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2014, **113**, 1136–1151.

ADDITIONAL KEYWORDS: diversification – integrative taxonomy – mammals – molecular systematics – species delimitation.

INTRODUCTION

Delimiting species is a challenging task and choosing an appropriate approach has been the subject of much debate (de Queiroz, 1998). A major problem in taxonomy is that species boundaries can differ widely depending on the species concept applied (de Queiroz, 2007). Much emphasis has been placed on developing a unified species concept rooted in the evolutionary origin of species (de Queiroz, 2007). Under the General Lineage Concept, anchored in the logic of the Evolutionary Species Concept (ESC), species are defined as evolutionarily independent lineages and are diagnosed by quantifying secondary characteristics of species (e.g. ecology, morphology, genetics, etc.) that indicate some form of evolutionary independence (Simpson, 1961; Wiley, 1978; de Queiroz, 2007).

The ESC recognizes that the speciation process is heterogeneous across taxa, with a multitude of factors that may be responsible for population divergence (de Queiroz, 1998). Therefore, the appearance of diagnosable characters is contingent upon the evolutionary processes responsible for species formation (de Queiroz, 2007). For example, strong selection gradients can promote phenotypic or ecological character divergence between populations, with accompanying genetic divergences initially seen only in genes of

*Corresponding author. E-mail: markphuong@gmail.com

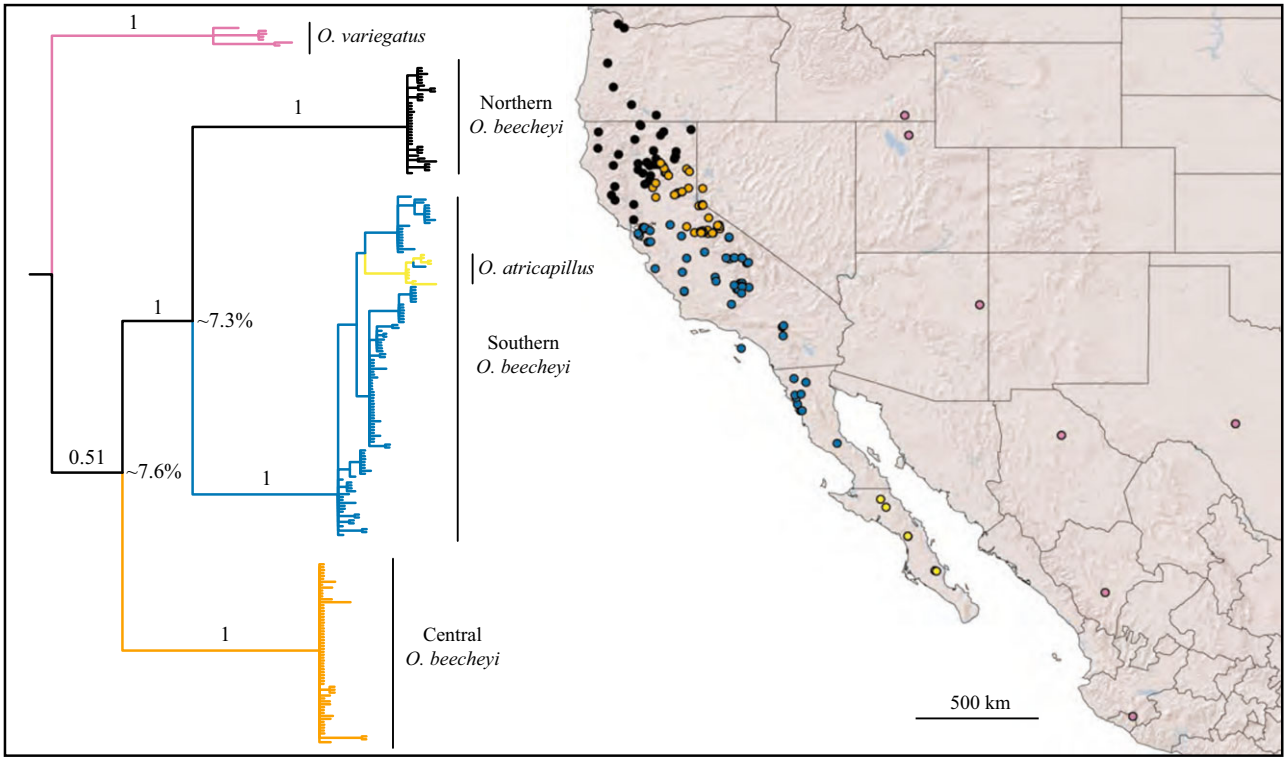

Figure 1. Mitochondrial Bayesian phylogeny of *Otospermophilus* and geographical distribution of samples sequenced for *cytochrome b* (mitochondrial DNA). Labels at nodes represent Bayesian posterior probabilities and average percentage sequence divergence between major clades. Phylogeny rooted with *C. lateralis*, not shown. Map of western United States and Central America created in ARCMAP, version 10.

adaptive significance (Linnen *et al*., 2013). Populations can also become geographically isolated through vicariance or dispersal events, leading to genetic divergence and reproductive isolation without overt changes in morphology (Bickford *et al*., 2007; Singhal & Moritz, 2013). Patterns of non-adaptive genetic divergence can be further complicated by introgression, leading to discordances between mitochondrial DNA (mtDNA) and nuclear markers (Toews & Brelsford, 2012). Hence, given the heterogeneous speciation process, the ESC implies that any set of characters is sufficient, yet no particular character is necessary, to propose and delimit species (Padial *et al*., 2009, 2010). This reasoning has led to explicit calls for integrative taxonomy, or the use of multiple lines of evidence from different datasets to classify and diagnose species (Dayrat, 2005; Padial *et al*., 2010; Schlick-Steiner *et al*., 2010); the term reflects the traditional practice of many taxonomists over multiple decades (e.g. Steppan, 1998; Patton, Da Silva & Malcolm, 2000).

In the present study, we combined different lines of evidence to delineate species boundaries within the genus *Otospermophilus* (Brandt, 1844). This genus currently includes three species of colonial ground squirrels: *Otospermophilus atricapillus* (Bryant, 1889), *Otospermophilus beecheyi* (Richardson, 1829), and *Otospermophilus variegatus* (Erxleben, 1777) (Helgen *et al*., 2009). *Otospermophilus atricapillus* has a narrow distribution (endemic to the Baja California Peninsula) relative to the wide distributions of *O. beecheyi* (far western USA) and *O. variegatus* (southwestern USA and Mexico) (Fig. 1) (Helgen *et al*., 2009). Original species classifications placed these taxa with 38 other species in the genus *Spermophilus* based on external morphological characteristics (e.g. coat colour) and geography (Grinnell & Dixon, 1918). More recent mtDNA and craniodental characters supported the distinctiveness of these three species and elevated them to the genus *Otospermophilus*. (Harrison *et al*., 2003; Herron, Castoe & Parkinson, 2004; Thorington & Hoffmann, 2005; Helgen *et al*., 2009). However, these studies focused on a genus-level revision of *Spermophilus* and the species of *Otospermophilus* were each represented by a few specimens (two or three individuals), thus preventing any assessment of species boundaries. Subsequent phylogeographical analyses of mtDNA from larger sample sizes of *O. atricapillus* and *O. beecheyi* revealed three highly divergent lineages

within *O. beecheyi* and showed *O. atricapillus* to be nested within one lineage of *O. beecheyi*, calling into question the current taxonomy of these species (Álvarez-Castañeda & Cortés-Calva, 2011). Specifically, Álvarez-Castañeda & Cortés-Calva (2011) recommended synonymizing *O. atricapillus* with *O. beecheyi*.

Here, we examined additional characters that may have been important in the evolutionary history of *O. atricapillus* and *O. beecheyi*. Species boundary inferences based solely on mtDNA can be unreliable as a result of selection on physiological processes, asymmetric introgression, and sex-linked dispersal (Ballard & Whitlock, 2004). Furthermore, inferring evolutionary relationships from any single locus may lead to inaccurate species designations as a result of stochastic variance in gene histories (Knowles & Carstens, 2007). Therefore, we expanded geographical sampling of mtDNA diversity and evaluated additional genetic data from 11 microsatellite markers and 11 nuclear loci (exons and introns) to investigate the patterns of divergence at the population and species level. In addition, the regions that *O. atricapillus* and *O. beecheyi* inhabit (specifically, California and Baja California) are environmentally heterogeneous, which could promote diversification through adaptive divergence along climatic niche space (Riemann & Ezcurra, 2005; Davis *et al*., 2008). As such, we analyzed eight bioclimatic variables to determine whether environmental heterogeneity could have contributed to lineage divergence. Finally, craniodental morphology is commonly used to delimit species boundaries in these and other rodents and is correlated with dietary preferences (Howell, 1938; Helgen *et al*., 2009; Samuels, 2009). Thus, we quantified 23 linear skull measurements to test whether morphological divergence reflects lineage divergence.

Our species delimitation hypotheses were based upon the maximum number of taxa possible derived from current taxonomy and initial mtDNA results (i.e. four; *O. atricapillus*, Northern *O. beecheyi*, Central *O. beecheyi*, and Southern *O. beecheyi*). These operational taxonomic units (OTUs) were subsequently tested with each line of evidence collected. Final species designations were assessed qualitatively by examining all datasets collectively and by considering evolutionary processes that may have generated observed patterns of concordance and discordance of species boundaries between datasets.

MATERIAL AND METHODS

Genetic sampling

We obtained tissue samples from eight *O. atricapillus* individuals and 207 *O. beecheyi* individuals from various institutions (see Supporting information, Table S1). We included four samples of *O. variegatus* for outgroup comparison in sequence data analyses (see Supporting information, Table S1). We chose samples to include individuals at geographically disparate localities to encompass each species' known range. To obtain genomic DNA, we used salt extraction (Aljanabi & Martinez, 1997). In addition, we extracted DNA from 35 individuals of *O. beecheyi* from skin fragments taken from museum specimens using a protocol described by Mullen & Hoekstra (2008) with modifications described in Rowe *et al*. (2011).

mtDNA data and analysis

In our analyses, we included the *cytochrome b* (*cytb*) sequences available on GenBank and reported in Álvarez-Castañeda & Cortés-Calva (2011) for 10 *O. atricapillus* individuals, 75 *O. beecheyi* individuals (11 Northern *O. beecheyi*, 28 Central *O. beecheyi*, 36 Southern *O. beecheyi*), and three *O. variegatus* individuals (see Supporting information, Table S1). We sequenced an additional 170 *O. beecheyi* (45 Northern *O. beecheyi*, 45 Central *O. beecheyi*, 80 Southern *O. beecheyi*) and four *O. variegatus* to increase the geographical sampling of each lineage and to narrow geographical gaps between lineages (see Supporting information, Table S1). We also included one sample of *Callospermophilus lateralis* (GenBank #AF157887) to root the mtDNA phylogeny. For DNA extracted from modern tissue samples, we amplified approximately 800 bp of *cytb* using the primers MVZ05/MVZ16 (see Supporting information, Table S2). When individuals did not successfully amplify the 800-bp region, we used alternative primers to target shorter regions (see Supporting information, Table S2). For DNA extracted from museum skins, we created novel primers to target shorter regions of *cytb* because DNA from skin specimens are fragmented (see Supporting information, Table S2). Polymerase chain reaction (PCR) conditions and loci information are provided in the Supporting information (Appendix S1). We generated sequence data using an ABI 3730 automated DNA sequencer and created alignments using Geneious Pro (Biomatters). Only sequences that were at least 600 bp were used in downstream analyses.

For phylogenetic inference, we inferred the best-fitting substitution model using MRMODELTEST, version 2.3 (Nylander, 2008) and estimated a phylogeny in MrBayes, version 3.1.2 (Huelsenbeck & Ronquist, 2001). The chain was sampled every 10 000 generations over 10 million generations with a four million generation burn-in, which was sufficient for convergence.

We characterized genetic variation for each OTU and *O. variegatus* in ARLEQUIN, version 3.1

(Excoffier, Laval & Schneider, 2005). To assess interspecific genetic variation, we calculated sequence divergence between OTUs (*Dxy*). To assess within OTU genetic variation and deviations from neutrality, we calculated nucleotide diversity, Tajima's *D*, and Fu's F_S. For these analyses, we used the substitution model with closest affinity to the model inferred by MRMODELTEST if it was not available in ARLEQUIN.

Microsatellite data and analysis

We genotyped 11 polymorphic microsatellite loci for 205 specimens of *O. beecheyi* (40 Northern *O. beecheyi*, 61 Central *O. beecheyi*, 104 Southern *O. beecheyi*) and eight *O. atricapillus* individuals from across their range (see Supporting information, Table S1). We used these data to test whether clustering of individuals by their multilocus nuclear genotypes are congruent with the four OTUs. In addition, our mtDNA analyses identified two areas where mtDNA lineages within *O. beecheyi* co-occur and we used microsatellite variation to test for evidence of introgression at these contact zones. The 11 loci previously characterized for ground squirrels and analyzed for the present study were: IGS-BP1, IGS-1, IGS-6 (May *et al*., 1997), MS45 (Hanslik & Kruckenhauser, 2000), 2g2 (Kyle *et al*., 2004), MA018 (Da Silva *et al*., 2003), SS-Bibl18, SS-Bibl4 (Goossens *et al*., 1998), GS17, GS22, and GS25 (Stevens, Coffin & Strobeck, 1997) (see Supporting information, Table S3). Additional information on loci and PCR conditions is provided in the Supporting information (Appendix S1). We used an ABI 3730 DNA Analyzer (Applied Biosystems, Inc.) to size PCR products and GENEMAPPER, version 4.0 (Applied Biosystems, Inc.) to manually score alleles.

We used MSA (Microsatellite Analyzer), version 4.05 (Dieringer & Schlötterer, 2003) to check for errors (unexpected mutation steps, large gaps in data, or unexpected short or long allele sizes) and FREENA to check for the presence of null alleles (Chapuis & Estoup, 2007). When checking for errors and null alleles, we grouped samples by mtDNA assignment. We tested for Hardy–Weinberg equilibrium (HWE) for each OTU in Arlequin v3.1. For the HWE analysis, we grouped individuals based on the assignments given by either the mtDNA phylogeny or the two population assignment methods described below (Excoffier *et al*., 2005). We adjusted the data using a Bonferroni correction to avoid inflation of Type I errors given multiple tests. To summarize genetic divergence between each OTU, we calculated F_{ST} in ARLEQUIN, version 3.1 (Excoffier *et al*., 2005).

With the microsatellite data, we identified clusters of individuals using STRUCTURE, version 2.3 (Pritchard, Stephens & Donnelly, 2000) and by performing a discriminant analysis of principal components (DAPC) (Jombart, Devillard & Balloux, 2010). STRUCTURE uses a Bayesian approach to assign individuals to genetic clusters (*K*). For *K* ranging from 1–5, we ran STRUCTURE 10 times for one million generations, with a burn-in of 100 000 generations. We used STRUCTURE HARVESTER (Earl & vonHoldt, 2012) to summarize across runs, CLUMPP, version 1.1.2 (Jakobsson & Rosenberg, 2007) to cluster results from STRUCTURE run repetitions, and DISTRUCT, version 1.1 (Rosenberg, 2004) to visualize population structure. DAPC performs a PC analysis (PCA) that is followed by a discriminant ana1ysis of the retained PCs (Jombart *et al*., 2010). DAPC was implemented with the *adegenet* package in R (Jombart, 2008, R Development Core Team, 2012) and we retained 60 PC axes for this analysis.

Nuclear sequence data and analysis

To account for stochastic variation among loci when inferring species boundaries from genealogies, we sequenced 11 nuclear loci for a subset of individuals from each OTU and from *O. variegatus* for outgroup comparison (four *O. atricapillus*, five from each *O. beecheyi* OTU, and three *O. variegatus*). Four of the loci were previously published: the exons GHR (Adkins *et al*., 2001) and RAG1 (Steppan, Storz & Hoffmann, 2004) (Table S2), and the introns MDH2 (Debry & Seshadri, 2001) and SPTBN1 (Matthee *et al*., 2001). Seven introns are new to the present study: CCR2, ETIF, HRG, HSP90, IGFB, SEC16, and TP132 (see Supporting information, Table S2). To generate primers for the new loci, we aligned the transcriptomes from the Belding's ground squirrel, *Urocitellus beldingi*, and the alpine chipmunk, *Tamias alpinus*, and chose genes with high sequence divergence (> 7%) between the two species as candidates for marker development (*T. alpinus*: Bi *et al*., 2012; *U. beldingi*: K. Bi, unpublished data). We designed exon-primed intron-crossing primers from this subset of genes. PCR conditions and information on loci are provided in the Supporting information (Appendix S1). We generated sequence data from the above loci using an ABI 3730 automated DNA sequencer and created alignments using Geneious Pro (Biomatters). We manually adjusted alignments around indels by eye and coded indels as 'N'. We inferred haplotypes from each nuclear locus using PHASE, version 2.1.1 and kept heterozygous positions with > 90% probability for downstream analyses (Stephens & Donnelly, 2003; Flot, 2010).

To visualize nuclear (n)DNA relationships across individuals, we generated a multilocus network. We

used POFAD (Phylogeny of Organisms From Allelic Data), which explicitly incorporates allelic variation among loci to generate a network of similarity among individuals (Joly & Bruneau, 2006). To create the network, for each locus, we inferred a substitution model for each locus using MRMODELTEST, version 2.3 (Nylander, 2008), estimated distance matrices using PAUP (Phylogenetic Analysis Using Parsimony) (Swofford, 2002), and calculated individual-based distance matrices using POFAD. We visualized the results using SPLITSTREE (Huson & Bryant, 2006). In addition, we generated a gene tree for each locus with RAXML (Randomized Axelerated Maximum Likelihood), version 7.0.4 (Stamatakis, 2006). We used the simplest substitution model in RAXML (GTR + Γ) if substitution models selected by MRMODELTEST were simpler than those implemented in RAXML.

We characterized genetic variation for the nuclear sequence data from each OTU and *O. variegatus* using ARLEQUIN, version 3.1 (Excoffier *et al*., 2005). To assess between OTU genetic variation, we calculated sequence divergence (*Dxy*) for each nuclear locus. To assess within OTU genetic variation and deviations from neutrality, we calculated nucleotide diversity, Tajima's *D*, and Fu's F_S for each locus separately. We conducted all genetic variation analyses using the substitution model inferred from MRMODELTEST. If the model was not available in ARLEQUIN, we chose the substitution model with the closest affinity to the one inferred by MRMODELTEST.

Bayesian species delimitation

Traditional species delimitation incorporating molecular data typically used criteria such as fixed differences or reciprocal monophyly to identify species (Zachos *et al*., 2012). However, reciprocal monophyly is not expected for all loci, especially if speciation events are recent (Rosenberg, 2007). Recently developed methods now apply coalescent theory to delimit species that account for stochastic variation among loci (Fujita *et al*., 2012). We used one such approach with BPP (Bayesian Phylogenetics & Phylogeography), version 2.0 (Yang & Rannala, 2010) to test for evolutionary independence of lineages (i.e. the four OTUs) based on results from the 11 sequenced nuclear loci. BPP estimates speciation probabilities at the nodes between lineages under a conservative model of no gene-flow after speciation and this method accounts for gene-tree discordance by explicitly modeling the coalescent process (Leaché & Fujita, 2010). BPP assumes that the loci are (1) phased and (2) free from intralocus recombination. To identify nonrecombining blocks within loci, we filtered phased nuclear data through the program IMgc (Woerner, Cox & Hammer, 2007). BPP also requires a user-specified guide topology and results are sensitive to mis-specified guide trees (Yang & Rannala, 2010). Therefore, we provided multiple plausible topologies from current taxonomy, the mtDNA phylogeny, and microsatellite clusterings (see Supporting information, Fig. S1). BPP also requires specified priors for the effective population size (θ) and divergence time (τ_0). Previous studies showed a significant impact of prior specification on the results (Leaché & Fujita, 2010). Thus, we analyzed each guide tree under divergence scenarios that were chosen to represent combinations of extreme values for both priors (large ancestral population size and deep divergence – $\theta \sim G(1,10)$ and $\tau_0 \sim G(1,10)$; small ancestral population size and shallow divergence – $\theta \sim G(2,2000)$ and $\tau_0 \sim G(2,2000)$; large ancestral population size and shallow divergence – $\theta \sim G(1,10)$ and $\tau_0 \sim G(2,2000)$). We tested both available species delimitation algorithms (species algorithm 0 and 1) for all possible scenarios. Each run consisted of 500 000 generations, sampling every five generations with a burn-in of 50 000 generations, which resulted in high effective sample sizes (ESS) values (i.e. > 200).

Species tree reconstruction

To resolve phylogenetic relationships among independent lineages inferred from BPP, we estimated a species tree using a coalescent, Bayesian method implemented in *BEAST (Bayesian Evolutionary Analysis Sampling Trees), version 1.6.1 (Heled & Drummond, 2010). *BEAST allows for the incorporation of intraspecific polymorphism and incomplete lineage sorting in phylogeny estimation and requires the *a priori* designation of species (i.e. it assumes no migration among lineages) (Heled & Drummond, 2010). Therefore, independently evolving lineages identified by the BPP analysis were used as the primary taxa in generating a topology. We used phased nuclear alleles without evidence of intralocus recombination. We used the substitution models for each locus inferred from MRMODELTEST and used an uncorrelated, relaxed molecular clock model. We set the mean.ucld parameter for each locus to follow a gamma distribution ($k = 1$, $\theta = 1$). We left all other priors on the default setting and let the Markov chain Monte Carlo analysis run for 100 million generations, sampling every 4000th iteration. We discarded 40% of the trees and evaluated convergence by examining ESS values in TRACER, version 1.4.1 (Rambaut & Drummond, 2007).

Environmental analysis

Occurrence records for each OTU were queried using MANIS (Mammal Networked Information System;

http://www.manisnet.org; accessed on 2 April 2012). We partitioned locality points for *O. beecheyi* into the Northern, Central, and Southern OTUs by incorporating information on the geographical distribution of the mtDNA lineages. We georeferenced genetic samples included in the analyses using GEOLOCATE, version 3.22, from locality information if geographical coordinates were not recorded (Rios & Bart, 2010). Because of the low number of unique localities for *O. atricapillus*, we georeferenced additional records from MANIS using locality information with GEOLOCATE and added them to the analyses. We discarded geographical coordinates if: (1) coordinate uncertainty was greater than 10 km; (2) no coordinate uncertainty was reported; or (3) the collection date was prior to 1950 (climate layers are averaged from 1950–2000; Hijmans *et al*., 2005). To avoid overparameterization of downstream analyses, we chose eight out of 19 BIOCLIM variables from the WorldClim database (see Supporting information, Table S4) (Hijmans *et al*., 2005) that were not correlated with each other (Pearson correlation coefficient $|r| < 0.7$). To ensure that only unique localities were included in the analyses, we removed all but one point within an occupied cell within the bioclimatic layers using a custom R script (R Development Core Team, 2012). To consider the total environmental space these OTUs could encounter, we generated 500 random points across the range of *O. atricapillus* and *O. beecheyi*. To visualize potential environmental niche divergence within the context of the environment available to the OTUs, we conducted a PCA.

Furthermore, we quantified and statistically tested environmental niche divergence between OTUs using an ordination technique developed by Broennimann *et al*. (2011). Ordination techniques enable the direct comparison of environmental space between species of interest (Broennimann *et al*., 2011). The method constructs an environmental grid from the region of interest using the first two axes from a PCA of environmental variables. Then, it incorporates (1) the relative species occurrence density in each cell within the environmental grid estimated by a kernel density function and (2) the relative frequency of different environmental conditions to create a standardized metric of environmental occupancy (z_{ij}) (Broennimann *et al*., 2011). The environmental occupancy metric is used to calculate the niche overlap metric, Schoener's *D*, where values of *D* range from 0 (niches are completely discordant) to 1 (niches are identical) (Warren, Glor & Turelli, 2008). This approach accounts for nonrandom distribution of records across a species' range and incorporates information about the environment available to the species of interest (Broennimann *et al*., 2011). To statistically test the significance of niche overlap values, we performed the niche similarity test. This test generates simulated niche overlap values by comparing the environmental niche of one taxon to randomly shifted occurrence points across the available environmental space in another taxon (Broennimann *et al*., 2011). Here, we modify the test by treating it as a one-tailed test with a null hypothesis of niche similarity. If the observed value is less than 5% of the simulated values, then niche similarity is rejected and the taxa are more different than expected. We used the ordination technique 'PCA-env', which creates the environmental grid based on total environment available to the taxa under comparison and performs more accurately than other strategies (Broennimann *et al*., 2011). To characterize the environmental space available for each OTU, we extracted 500 random points using polygons created in ARCMAP, version 9.3 (Environmental Systems Research Institute) that circumscribed each OTU's distribution. We performed the test in R (R Development Core Team, 2012) on geographically adjacent taxa with 1000 replicates.

Morphological analysis

We collected craniodental measurements from 369 specimens spanning all OTUs (10 *O. atricapillus*, 83 Northern *O. beecheyi*, 36 Central *O. beecheyi*, 240 Southern *O. beecheyi*; see Supporting information, Table S1); we partitioned samples into OTUs by using information on the geographical extent of each OTU. We used adults only, defined here as individuals with a fully erupted M3 and at least some wear to molar cusps (Helgen *et al*., 2009). We measured twenty-three highly repeatable and reliable craniodental variables from Helgen *et al*. (2009) and Patton, Huckaby & Álvarez-Castañeda (2008) for each individual (see Supporting information, Table S5). We collected measurements with handheld digital calipers to the nearest 0.01 mm and, to ensure consistency, all measurements were taken by a single individual (DRW).

We performed a PCA to assess the degree of overlap between the OTUs and we used the PC values to test for morphological differences in two ways. First, we performed a multivariate analysis of covariance (MANCOVA) on the first five PC axes (eigenvalues > 1) at the same time as controlling for latitude and sex, two variables that are known to influence body size in *Otospermophilus* (Blois, Feranec & Hadly, 2008). Second, we explicitly tested whether geographical shifts in morphology were concordant with genetic shifts based on OTU geographical boundaries. If morphological differentiation was consistent with molecular differentiation, we expect to observe abrupt shifts in skull morphology along

geographical transects at genetically defined OTU boundaries. Partitions were made within each OTU using an ecoregions map to provide uniformity of habitat and vegetation within groupings (United States Environmental Protection Agency). To execute this analysis, we first performed a multivariate analysis of variance (MANOVA) on the first five PC axes to account for variation as a result of sex differences. Using residuals for the first two PC axes from the MANOVA, we tested for intraspecific and interspecific differences between groupings using an ANOVA with Tukey's honestly significant difference method.

To examine whether species from our final taxonomic decisions could be discriminated by skull characters, we performed a discriminant function analysis (DFA) on the 23 skull measurements and on the residuals of the 23 measurements after correcting for latitude using a MANOVA. We separated the analyses by sex.

All sequence data are available on GenBank (GenBank accession numbers KM504528–KM504936; see Supporting information, Appendix S2). Microsatellite data, final sequence alignments, environmental data, and morphological measurements are available on Dryad (Phuong *et al*., 2014).

RESULTS

Mitochondrial phylogeny and analysis

Bayesian phylogenetic inference of mtDNA supported four lineages with high confidence (Bayesian posterior probabilities = 1), consistent with the topology in Álvarez-Castañeda & Cortés-Calva (2011) (Fig. 1). Based on the mtDNA, we were unable to resolve whether Central *O. beecheyi* or *O. variegatus* was the earliest diverging lineage within *Otospermophilus* (Bayesian posterior probability = 0.51) (Fig. 1). The mtDNA analyses supported three lineages within *O. beecheyi* and the Southern lineage is paraphyletic based on the placement of *O. atricapillus* within it. Our more extensive geographical sampling facilitated greater resolution of the distributional extent of each *O. beecheyi* lineage relative to Álvarez-Castañeda & Cortés-Calva (2011); otherwise, our results are consistent with their previous findings (Fig. 1):

(1) The Northern lineage ranges from southern Washington to the north of the Sacramento–San Joaquin River Delta
(2) The Central lineage is mostly found in the Sierra Nevada and foothills on both slopes
(3) The Southern lineage extends south from the Sacramento-San Joaquin River Delta, through central and southern California to northern Baja California
(4) We identified a contact zone between the Northern and Central lineages at Lake Almanor in Plumas County and a contact zone between the Central and Southern lineages at Mono Lake in Mono County (Fig. 2)

Net mitochondrial sequence divergence between the three major lineages of *O. beecheyi* ranged from 0.0719 to 0.0799 substitutions per site, whereas *O. atricapillus* was only slightly divergent from Southern *O. beecheyi* individuals (0.0136 substitutions per site) (Table 1). Nucleotide diversity (ND) was highest in Southern *O. beecheyi* (ND = 0.005) and lowest in Northern *O. beecheyi* (ND = 0.0018; see also Supporting information, Table S6). Tajima's D for Northern *O. beecheyi* and Fu's F_S for Central and Southern *O. beecheyi* significantly differed from neutral expectations; all other neutrality tests were not significant (see Supporting information, Table S6).

Microsatellite analysis

MSA found no errors with the genotype dataset, and null allele frequency was low for all loci analyzed (< 0.2; see Supporting information, Table S7). The analysis of genotype data showed that genetic diversity is high within OTUs (Northern: H_O = 0.118–0.825, Central: H_O = 0.361–0.852, Southern: H_O = 0.047–0.795; see Supporting information, Table S8). We detected a strong signal of heterozygote deficiency in Southern *O. beecheyi* ($P < 0.001$), where several loci deviated from HWE, irrespective of how we defined OTUs (see Supporting information, Table S8). F_{ST} results indicated that Northern *O. beecheyi* is distinct from both the Central *O. beecheyi* (F_{ST} = 0.11) and Southern *O. beecheyi* (F_{ST} = 0.10) lineages, whereas the latter OTUs were more similar to each other (F_{ST} = 0.03; see Supporting information, Table S9). *Otospermophilus atricapillus* was distinct from all *O. beecheyi* taxa, although this may be due to isolation by distance ($F_{ST} > 0.100$; see Supporting information, Table S9).

The STRUCTURE output supported two genetic populations based on the Evanno, Regnaut & Goudet (2005) criterion: Northern *O. beecheyi* and one population including Central *O. beecheyi*, Southern *O. beecheyi*, and *O. atricapillus* (Fig. 2; see also Supporting information, Figs S2, S3). For $K = 3$, population assignment was largely congruent with the mtDNA partitions by splitting *O. beecheyi* into the three mtDNA lineages and including *O. atricapillus* within Southern *O. beecheyi* (Fig. 2). However, we observed several mismatches between mtDNA assignment and microsatellite assignment for Central and Southern *O. beecheyi* individuals from geographically disparate localities (Fig. 2). At the contact zone

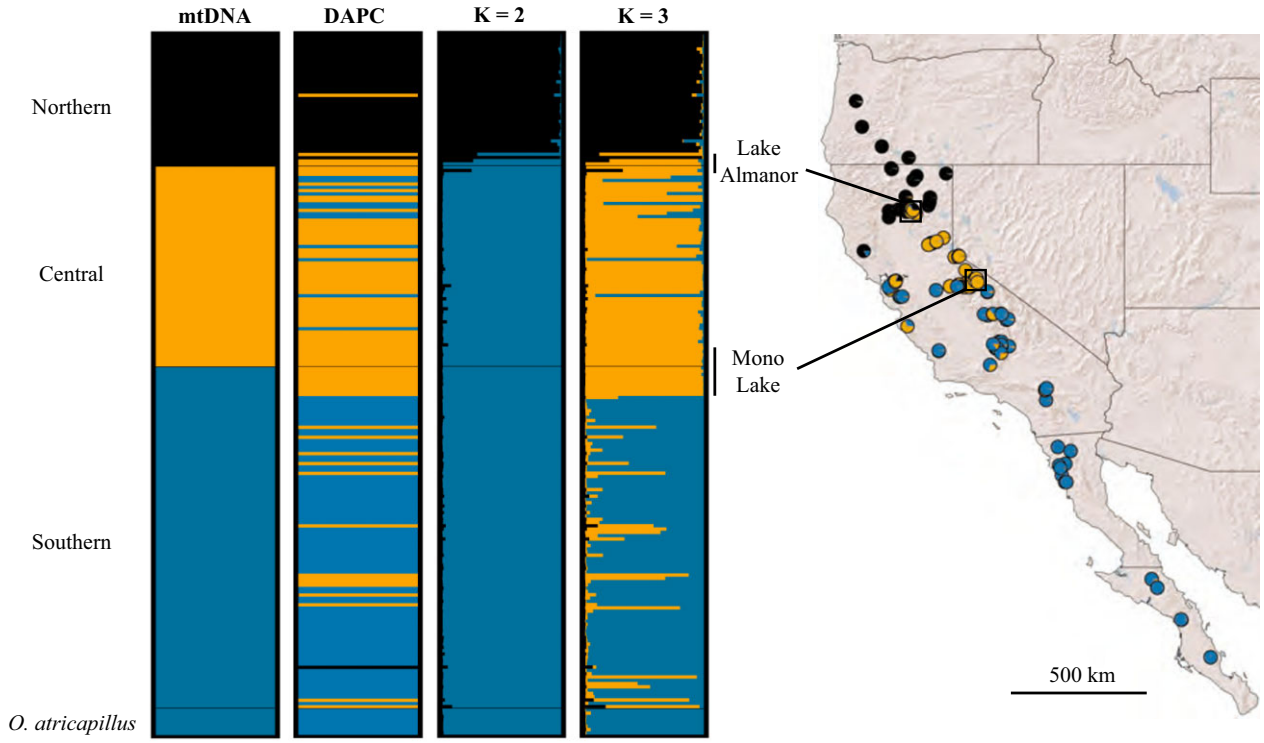

Figure 2. Comparison of mitochondrial DNA (mtDNA) and microsatellite population assignments (discriminant analysis of principal components; DAPC, STRUCTURE $K = 2$, STRUCTURE $K = 3$), with individuals (horizontal bars) arranged from North to South. $K = 2$ is the most likely cluster of individuals according to microsatellite data (Evanno *et al.*, 2005). Map showing geographical distribution of individuals genotyped at microsatellite loci with population assignments at STRUCTURE $K = 3$. Each pie chart represents an individual. Map of western USA and Central America created in ARCMAP, version 10.

Table 1. Sequence divergence (*Dxy*) of concatenated nuclear loci (above diagonal) and mitochondrial data (below diagonal)

	Northern	Central	Southern	*Otospermophilus atricapillus*	*Otospermophilus variegatus*
Northern		0.0049	0.0053	0.0049	0.0055
Central	0.0799		0.0023	0.0030	0.0039
Southern	0.0724	0.0708		0.0025	0.0037
Otospermophilus atricapillus	0.0756	0.0749	0.0136		0.0034
Otospermophilus variegatus	0.0943	0.0719	0.0709	0.0742	

around Mono Lake, microsatellite data assigned all individuals to a Central *O. beecheyi* population irrespective of their mtDNA haplotypes (Fig. 2; see also Supporting information, Fig S4). Although we were unable to summarize results across runs for $K = 4$ and 5, these higher K values continued to subdivide Central and Southern *O. beecheyi* into smaller populations, while consistently identifying Northern *O. beecheyi* as a single distinct population (see Supporting information, Fig. S4). STRUCTURE never assigned *O. atricapillus* as a single genetic population at any K value (Fig. 2; see also Supporting information, Fig. S4).

Microsatellite data suggested that there was little introgression between Northern *O. beecheyi* and the other OTUs. At all K values, STRUCTURE assigned all but three individuals with a Northern mitochondrial haplotype to a Northern microsatellite population (Fig. 2; see also Supporting information, Fig. S4). The three exceptions were recovered only at the contact zone around Lake Almanor (Fig. 2; see also Supporting information, Fig. S4).

The DAPC analysis found support for between four and six clusters with similar likelihoods based on the Bayesian Information Criterion (see Supporting information, Fig. S5). All cluster values supported

Northern *O. beecheyi* as distinct (see Supporting information, Fig. S6). These clusters did not identify any additional geographically meaningful populations within the other OTUs as a result of high overlap between clusters (see Supporting information, Fig. S6). Given the lack of information in higher numbers of clusters, we focus on population assignments based on three clusters. With three clusters, the DAPC ordination of microsatellite profiles shows clear separation of Northern *O. beecheyi* but considerable mixing between Central and Southern *O. beecheyi* (Fig. 2; see also Supporting information, Fig. S6). *Otospermophilus atricapillus* individuals are located within the group dominated by Southern *O. beecheyi* (Fig. 2; see also Supporting information, Fig S6). Individuals from both contact zones clustered with Central *O. beecheyi* (Fig. 2; see also Supporting information, Fig S6). These results generally agree with the STRUCTURE analyses (Fig. 2; see also Supporting information, Fig. S6).

Nuclear phylogeny and analysis

The POFAD nuclear network supported Northern *O. beecheyi* and *O. variegatus* as distinct lineages, whereas Central *O. beecheyi*, and Southern *O. beecheyi* are not clearly separated (Fig. 3A). *Otospermophilus atricapillus* forms a somewhat distinct cluster but nests close to Central and Southern *O. beecheyi*. Individual gene trees revealed widespread topological discordance and incomplete lineage sorting across OTUs (see Supporting information, Fig. S7). Of the 7.39 kbp obtained, overall sequence divergence ranged from 0.0023–0.0055 substitutions per site (Table 1). Nucleotide diversity ranged from 0.0003–0.0025 (see Supporting information, Table S6). Deviations from neutrality were not detected (see Supporting information, Table S6).

For the BPP analysis, the topology of the guide tree influenced the number of distinct evolutionary lineages the method inferred. The majority of guide trees strongly supported five independent lineages within *Otospermophilus* (speciation probability = 1; Fig. 3B; see also Supporting information, Fig. S1). Northern *O. beecheyi* is consistently confirmed as a separate lineage, as was *O. atricapillus*, even when placed sister to Southern *O. beecheyi* as implied by the microsatellite results. By contrast, when the Central and Southern *O. beecheyi* OTUs were designated as sister taxa, analyses did not support them as significant, independent lineages (Fig. 3B; see also Supporting information, Fig. S1). Discrepancies in results between guide trees can be explained by the placement of divergent lineages as sister taxa, which can inflate divergences that BPP interprets as speciation events (Leaché & Fujita, 2010). Different species delimitation algorithms did not affect the outcome under any guide tree. θ and τ_0 priors affected the node between the Central and Southern *O. beecheyi* OTU: under shallow divergence and small ancestral population sizes, the algorithm interpreted the node as a speciation event (Fig. 3B; see also Supporting information, Fig. S1). Previous analyses have suggested that the algorithm is sensitive to smaller ancestral population sizes (Leaché & Fujita, 2010).

Using lineages identified by BPP as taxon sets for *BEAST, the resultant species tree placed the collapsed Central and Southern *O. beecheyi* OTUs sister to *O. atricapillus* with high support (posterior probability = 1; Fig. 3C). However, the analyses did not confidently resolve the basal node within *Otospermophilus*, creating a polytomy among *O. variegatus*, Northern *O. beecheyi*, and the clade comprising Central *O. beecheyi*, Southern *O. beecheyi*, and *O. atricapillus*.

Ecological phenotype

The first and second PCs of the climatic variables explained 75% of the variance in the data (see Supporting information, Table S10). The variables contributing to most of the variation in PC1 are Mean Annual Temperature (BIO1), Temperature Seasonality (BIO4), Precipitation of Driest Month (BIO14), and Precipitation Seasonality (BIO15). Variation in PC2 is explained most by Mean Temperature of Wettest Quarter (BIO8). The PCA plot showed no clear breaks in environmental space between OTUs, with the OTUs lying along a continuum in environmental space (Fig. 4A). Niche overlap values among OTUs of *O. beecheyi* ranged from 0.424 to 0.542, suggesting moderate overlap in environmental space. By contrast, the comparison of *O. atricapillus* to Southern *O. beecheyi* showed very little niche overlap ($D = 0.081$; see Supporting information, Table S11). However, niche overlap values were not significantly different given the environmental space available to each OTU, such that there is no evidence for significant niche differences ($P > 0.05$; see Supporting information, Table S11).

Craniodental phenotype

The first two PCs explained 51% of the variance in the data. Condylobasal length contributed most to the first axis, whereas palate width measurements contributed to the second axis (see Supporting information, Table S12). PCA plot showed demonstrable overlap between all putative taxa (Fig. 4B), yet the results of the MANCOVA showed significant differences among OTUs ($P < 0.0005$; see Supporting information, Table S13). However, explicit spatial analyses showed

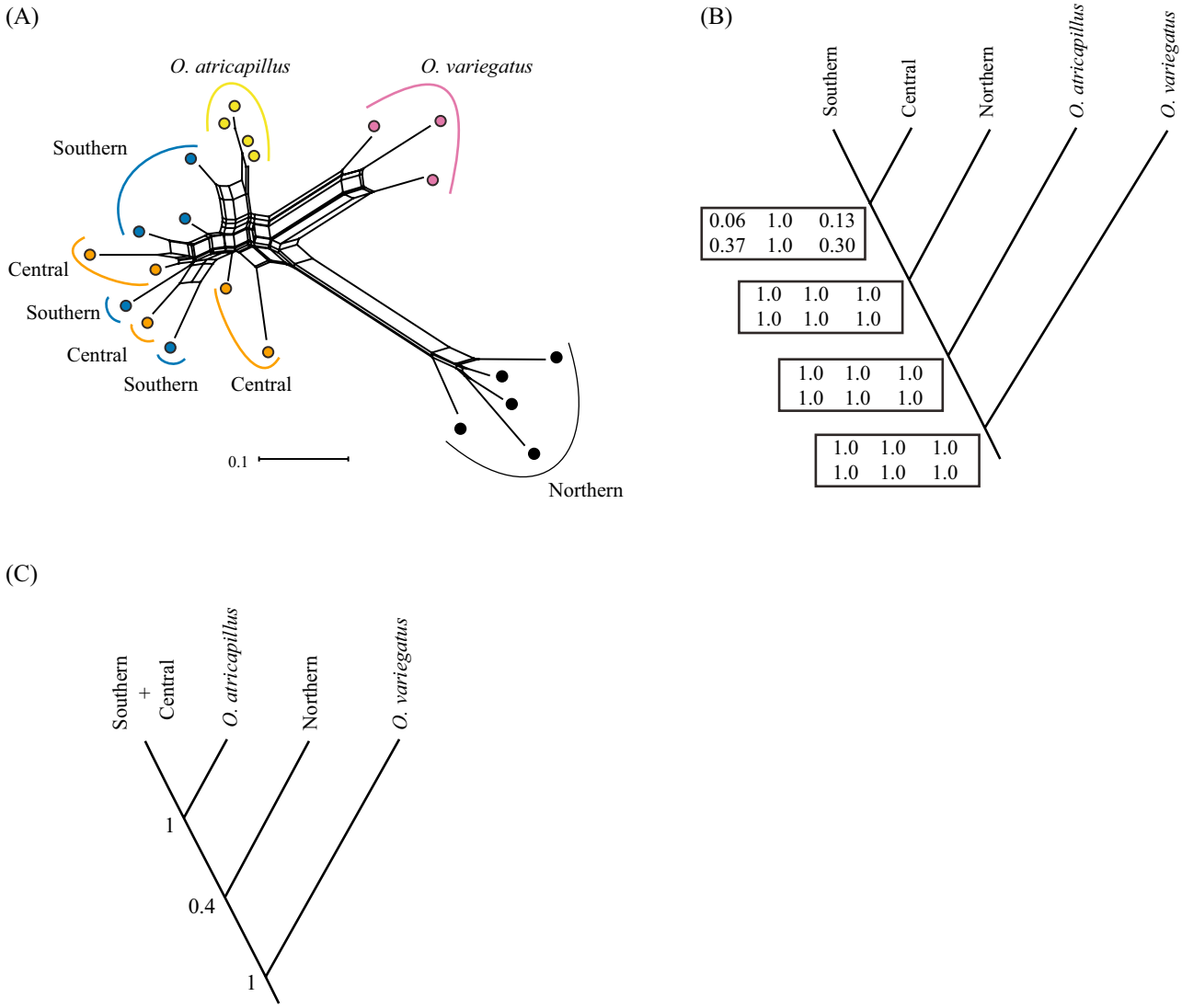

Figure 3. A, nuclear network based on 11 loci. Scale is a standardized, non-unit based measurement reported by POFAD. B, exemplar guide tree tested using BPP. Speciation probabilities are given at each node. Rows represent species algorithm (top, species algorithm 0; bottom, species algorithm 1) and columns represent different prior combinations (left, prior means = 0.1; middle, prior means = 0.001; right, prior mean $\theta = 0.1$, prior mean $\tau_0 = 0.001$). C, species tree inferred from *BEAST. Values at nodes represent posterior probabilities.

no demonstrable pattern of morphological breaks concordant with the known mtDNA genetic breaks within each *O. beecheyi* OTU or between them and *O. atricapillus* ($P > 0.05$; see Supporting information, Fig. S8). This suggests that there was no strong signature of morphological differentiation congruent with OTU definition.

The DFA correctly assigned 84.3% Northern *O. beecheyi* individuals and 94.8% of the individuals in the grouping of *O. atricapillus*, Central *O. beecheyi*, and Southern *O. beecheyi* (see Supporting information, Table S14). When latitude is taken into account, Northern *O. beecheyi* was misclassified 53% of the time, whereas the latter grouping was misclassified 10.8% of the time (see Supporting information, Table S14).

DISCUSSION

Here, we build on recent observations on the mtDNA phylogeography and colour pattern analyses that led to the suggestion that the Baja California rock ground squirrel, *O. atricapillus*, should be considered a part of the California ground squirrel, *O. beecheyi*, rather than a separate species (Álvarez-Castañeda & Cortés-Calva, 2011). In the spirit of integrative tax-

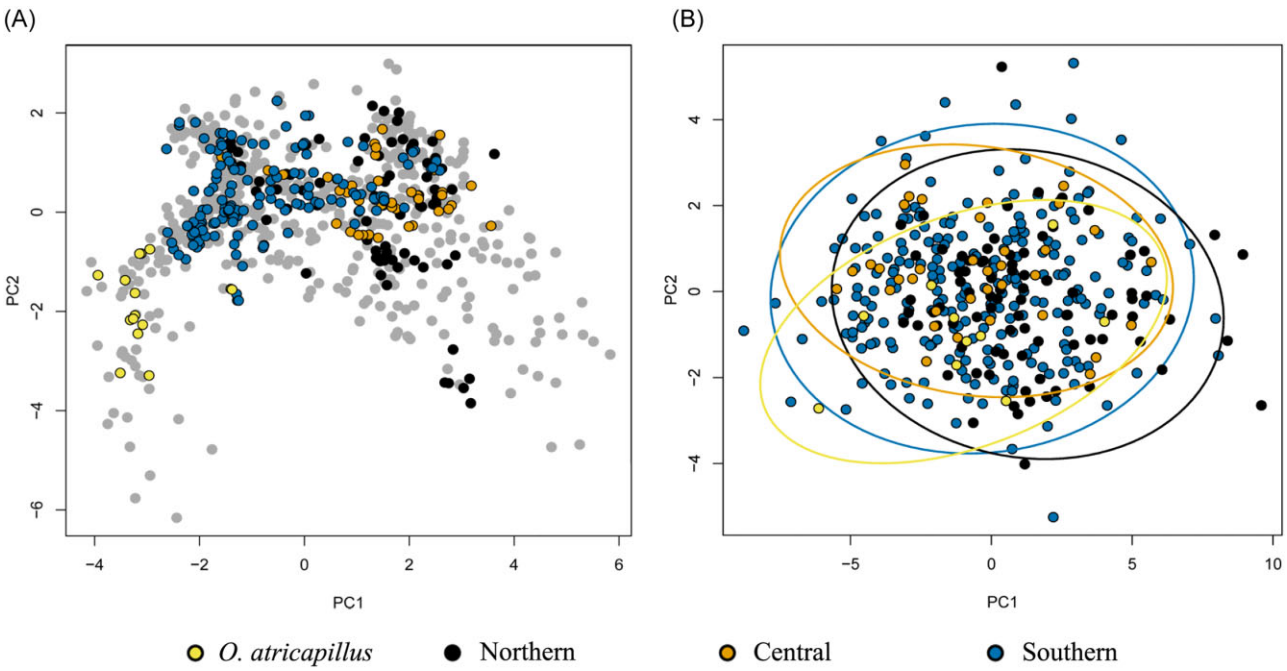

Figure 4. A, scatterplot of values for the first two principal components calculated from environmental data. Each point represents a unique sampling locality and is coded by the operational taxonomic unit (OTU) at that locality. Grey points represent environments available to the putative taxa. B, scatterplot of values for the first two principal components calculated from morphological data. Each point represents a measured individual and is coded by the OTU. Ellipses circumscribe 95% of the individuals within each putative taxon.

onomy (a long established practice), we combined evidence on morphological, bioclimatic, nuclear genes (11 microsatellites and 11 sequenced loci), and an expanded mtDNA dataset to test this proposition and to evaluate whether highly divergent mtDNA clades within *O. beecheyi* warrant species status. Working within the ESC and using these multiple lines of evidence, we aimed to identify evolutionarily independent lineages but recognize that discordance across datasets can arise through complex evolutionary processes.

Species delimitation

Our results showed disagreement between datasets on where species boundaries should be placed (Fig. 5). Taking into account all of the analyses generated in the present study, we propose to split *O. beecheyi* into a northern (Northern *O. beecheyi*) and southern (Central and Southern *O. beecheyi*) species, and in agreement with Álvarez-Castañeda & Cortés-Calva (2011), also synonymize *O. atricapillus* with the southern species of *O. beecheyi*. We explain our rationale below.

We did not detect a strong signal of bioclimatic or morphological differentiation between OTUs. Analysis of bioclimatic space occupied relative to the background environment revealed no significant divergence and suggested little opportunity for bioclimatic factors to drive divergence. Multivariate analyses of craniodental morphology did show significant differences between OTUs; however, the strong degree of overlap in the PCA and lack of obvious morphological divergence at geographical boundaries of OTUs suggest little biological relevance. Other morphological features such as pelage coloration and bacula morphology have been inconclusive in delineating species in *Otospermophilus* as well (Burt, 1960; Álvarez-Castañeda & Cortés-Calva, 2011). Without strong signatures of ecological and morphological divergence to aid in discerning species boundaries, we turned to results from our genetic data.

All genetic analyses were consistent with evolutionary independence and long-term isolation of Northern *O. beecheyi* from all other OTUs. Inferences from mtDNA, microsatellite analyses (STRUCTURE and DAPC), and nuclear sequence analyses (POFAD and BPP) demonstrated strong differentiation of Northern *O. beecheyi* from other OTUs. In addition, microsatellite analyses showed limited introgression at Lake Almanor (where Northern *O. beecheyi* is syntopic with Central *O. beecheyi*) but not outside the contact zone. Furthermore, nuclear analyses (Fig. 3A) showed that Northern *O. beecheyi* had greater genetic

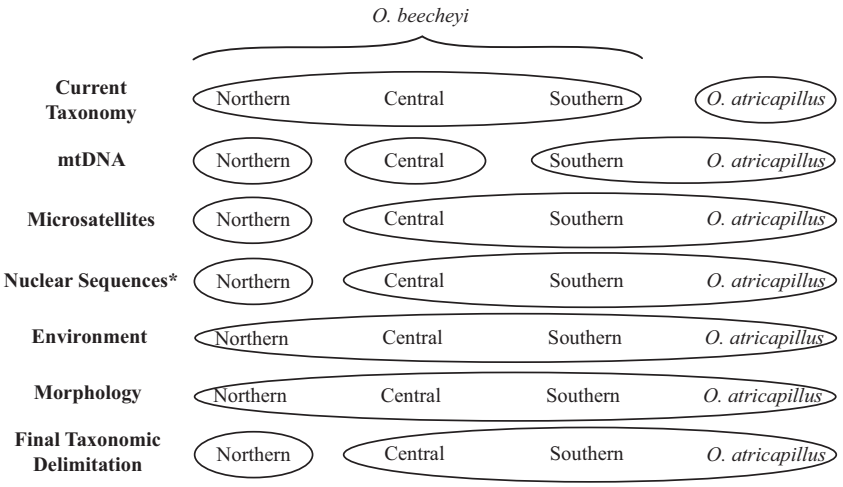

Figure 5. Summary of species boundary inferences from each dataset and final taxonomic grouping. *Based on POFAD analysis.

divergence from all other OTUs than *O. variegatus*, a currently recognized species in the genus. Taken together, these results supported a scenario of morphologically cryptic divergence and provide sufficient evidence to recognize Northern *O. beecheyi* as a species under the ESC.

By contrast, although Central *O. beecheyi* and Southern *O. beecheyi* are highly distinct in mtDNA, both classes of nuclear data [multilocus sequencing (POFAD and BPP) and microsatellites (STRUCTURE at $K = 2$)] collapsed them into one distinct lineage. Microsatellites analyzed from STRUCTURE at $K = 3$ and the DAPC analysis showed considerable mixing with respect to mtDNA haplotype that was geographically widespread rather than confined to a narrow contact zone. One possible scenario that could result in such strong discordance between mtDNA and nDNA is extensive historical introgression (Singhal & Moritz, 2012). Higher STRUCTURE K values and DAPC cluster values only increased the degree of discordant assignment among mitochondrial OTUs and erased any signature of Central and Southern populations. Based on the lack of nDNA divergence, we treat these two OTUs as one species.

For *O. atricapillus*, the genetic analyses were somewhat inconsistent. Neither mtDNA, nor microsatellite analyses supported this OTU as a distinct entity and suggested a close relationship to Southern *O. beecheyi*. For the multilocus nuclear sequences, the POFAD network showed only slight separation from Central and Southern *O. beecheyi*, yet BPP analyses consistently identified *O. atricapillus* as an independently evolving lineage. BPP may have identified *O. atricapillus* because it formed a monophyletic cluster slightly discrete from the diversity across Central and Southern *O. beecheyi* (Fig. 3B). Although robust to some model violations (Zhang *et al*., 2011), BPP can split more finely than alternative genetic delimitation methods and needs to be considered in light of other evidence (Carstens *et al*., 2013). Álvarez-Castañeda & Cortés-Calva (2011) examined colour variation in pelage in *O. atricapillus*, and concluded that the range of variation in this (i.e. the primary diagnostic character) was within the range of *O. beecheyi*. Therefore, we agree with Álvarez-Castañeda & Cortés-Calva (2011) and choose to synonymize *O. atricapillus* with Central and Southern *O. beecheyi*.

Given the available information analyzed in the present study and accounting for the multitude of ways that species can form, we consider the proposed split of *O. beecheyi* into two species is justified as a case of morphologically cryptic divergence. When latitudinal differences in skull characteristics are taken into account, the DFA indicated that it is difficult to morphologically discriminate between the two taxa. Although discriminatory power is increased when excluding latitude as a covariate, 15.7% of Northern *O. beecheyi* individuals were still misclassified (see Supporting information, Table S14). These findings are consistent with previous reports showing that besides differences in average size, the skulls of these species are indistinguishable (Hall, 1981). The presence of morphologically cryptic species is common throughout the tree of life, whereby speciation can occur without overt changes to morphology (Bickford *et al*., 2007; Singhal & Moritz, 2013). Here, our datasets exemplified this phenomenon, where multiple lines of genetic evidence support the recognition of Northern *O. beecheyi* as a separate species. We hypothesize that Northern *O. beecheyi* diverged from the rest of the species complex in allopatry and have recently

come into contact; this proposal is consistent with phylogeographical studies that found similarly located genetic breaks in other taxa, suggestive of a common biogeographical history (Barrowclough, Gutierrez & Groth, 1999; Feldman & Spicer, 2006). We are not currently able to explicitly test the divergence histories hypothesized here and above as a result of low sequence diversity at the nuclear loci sequenced. Tests of these scenarios and parameterization of divergence models require more loci, which is the subject of ongoing studies using high-throughput sequencing technologies (Bi *et al*., 2012).

Taxonomic implications

We elevate the Northern *O. beecheyi* OTU to species status as *Otospermophilus douglasii* (Richardson, 1829), the earliest available name with a type locality within the known geographical limits of this taxon. The Central *O. beecheyi* and Southern *O. beecheyi* retain the name, *Otospermophilus beecheyi* (Richardson, 1829), and *O. atricapillus* (Bryant, 1889) is demoted to a subspecies within *O. beecheyi*. The diagnostic features of these two taxa largely rest on the geographical location from which the organism is captured and characters from molecular data, with mtDNA sequences as the simplest molecular marker to discern between these two species as a result of high sequence divergence. Further studies are needed to define non-molecular diagnostic characters among these species.

Otospermophilus beecheyi (Richardson, 1829)

Distribution

Sierra Nevada with its Northern limit at Lake Almanor. South from the Sacramento-San Joaquin River Delta, into the Central Valley, with its southernmost extent in Baja California. Includes an allopatric population in Baja California Sur. Type locality: 'neighbourhood of San Francisco and Monterey, in California'.

Otospermophilus douglasii (Richardson, 1829)

Distribution

Southern Washington; western Oregon; northern California to the north of the Sacramento–San Joaquin River Delta. Type locality: 'bank Columbia River, Oregon.'

Further taxonomic information can be found in the supplementary species accounts (Supporting information, Appendix S1). The splitting of the common and widespread *O. beecheyi* into two species may have implications for pest management and human health. Numerous studies investigating the ecology and life history of *O. beecheyi* in relation to agriculture and disease biology focus on patterns in discrete populations that are extrapolated to the entire species (Hubbart, Jachowski & Eads, 2011). Many traits including behaviour and life history were not investigated in the present study and may reveal important differences between *O. beecheyi* and *O. douglasii* with implications for disease biology and management.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Moritz laboratory and the MVZ community for thoughtful advice and discussions, in particular, K. Bi, T. L. Morelli, J. L. Patton, M. Tonione, S. Singhal, and A. B. Smith. We thank the Burke Museum of Natural History and Culture, the Museum of Southwestern Biology, the Museum of Texas Tech University, the Grinnell Resurvey Project Team, the Deck family, ST Álvarez-Castañeda, T. L. Morelli, and J. L. Patton for the majority of samples used in this study. We thank K. Bi for assistance in designing nDNA loci from his transcriptome sequences, as well as J. L. Patton, S. Singhal, A. B. Smith, and three anonymous reviewers for valuable comments on earlier drafts of this article. This work was supported by UC Berkeley Summer Undergraduate Research Fellowships to MAP and ML, UC Berkeley Biology Fellows Program Awards to DRW and MAP, the Museum of Vertebrate Zoology Undergraduate Biodiversity Award, and the NSF DEB 0640859 to CM.

REFERENCES

Adkins RM, Gelke EL, Rowe D, Honeycutt RL. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Molecular Biology and Evolution* **18:** 777–791.

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25:** 4692–4693.

Álvarez-Castañeda ST, Cortés-Calva P. 2011. Genetic evaluation of the Baja California rock squirrel *Otospermophilus atricapillus* (Rodentia: Sciuridae). *Zootaxa* **3138:** 35–51.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* **13:** 729–744.

Barrowclough GF, Gutierrez RJ, Groth JG. 1999. Phylogeography of spotted owl (*Strix occidentalis*) populations based on mitochondrial DNA sequences: gene flow, genetic structure, and a novel biogeographic pattern. *Evolution* **53:** 919–931.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* **13:** 403.

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

Blois JL, Feranec RS, Hadly EA. 2008. Environmental influences on spatial and temporal patterns of body-size variation in California ground squirrels (*Spermophilus beecheyi*). *Journal of Biogeography* **35:** 602–613.

Brandt J. 1844. Observations sur les différentes espèces de sousliks de Russie, suivies de remarques sur l'arrangement et la distribution géographique du genre *Spermophilus*, ansé que sur la classification de la familie des ecureuils (Sciurina) en général. *Bulletin Scientifique l'Académie Impériale des Sciences de Saint-Pétersburg* **1844[1843]:** 357–382.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin MJ, Randin C, Zimmermann NE, Graham CH, Guisan A. 2011. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21:** 481–497.

Bryant WE. 1889. Provisional description of supposed new mammals from California and Lower California. *Proceedings of the California Academy of Sciences of the United States of America* **2:** 25–27.

Burt WH. 1960. *Bacula of North American mammals*. Ann Arbor, MI: University of Michigan.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology* **22:** 4369–4383.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:** 621–631.

Da Silva A, Luikart G, Allaine D, Gautier P, Taberlet P, Pompanon F. 2003. Isolation and characterization of microsatellites in European alpine marmots (*Marmota marmota*). *Molecular Ecology Notes* **3:** 189–190.

Davis EB, Koo MS, Conroy C, Patton JL, Moritz C. 2008. The California Hotspots Project: identifying regions of rapid diversification of mammals. *Molecular Ecology* **17:** 120–138.

Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85:** 407–415.

De Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard DJ, Berlocher SH, eds. *Endless forms: species and speciation*. Oxford: Oxford University Press, 57–75.

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

Debry RW, Seshadri S. 2001. Nuclear intron sequences for phylogenetics of closely related mammals: an example using the phylogeny of *Mus*. *Journal of Mammology* **82:** 280–288.

Dieringer D, Schlötterer C. 2003. Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3:** 2002–2004.

Earl DA, VonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* **4:** 359–361.

Erxleben JCP. 1777. *Systema regni animalis per classes, ordines, genera, species, varietates cum synonymia et historia animalium. Classis I. Mammalia*. Lipsiae: Impensis Weygandianis.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1:** 47–50.

Feldman CR, Spicer GS. 2006. Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. *Molecular Ecology* **15:** 2201–2222.

Flot JF. 2010. SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources* **10:** 162–166.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* **27:** 480–488.

Goossens B, Graziani L, Waits L, Farand E, Magnolon S, Coulon J, Bel M, Taberlet P, Allainé D. 1998. Extra-pair paternity in the monogamous alpine marmot revealed by nuclear DNA microsatellite analysis. *Behavioral Ecology and Sociobiology* **43:** 281–288.

Grinnell J, Dixon J. 1918. *Natural history of the ground squirrels of California*. Sacremento, CA: California State Printing Office.

Hall E. 1981. *The mammals of North America*. New York, NY: Wiley.

Hanslik S, Kruckenhauser L. 2000. Microsatellite loci for two European sciurid species (*Marmota marmota*, *Spermophilus citellus*). *Molecular Ecology* **9:** 2163–2165.

Harrison RG, Bogdanowicz SM, Hoffmann RS, Yensen E, Sherman PW. 2003. Phylogeny and evolutionary history of the ground squirrels (Rodentia: Marmotinae). *Journal of Mammalian Evolution* **10:** 249–276.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27:** 570–580.

Helgen KM, Cole FR, Helgen LE, Wilson DE. 2009. Generic revision in the holarctic ground squirrel genus *Spermophilus*. *Journal of Mammalogy* **90:** 270–305.

Herron MD, Castoe TA, Parkinson CL. 2004. Sciurid phylogeny and the paraphyly of Holarctic ground squirrels (*Spermophilus*). *Molecular Phylogenetics and Evolution* **31:** 1015–1030.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Howell AH. 1938. *Revision of the North American ground squirrels with a classification of the North American Sciuridae*. Washington, DC: North American Fauna.

Hubbart JA, Jachowski DS, Eads DA. 2011. Seasonal and among-site variation in the occurrence and abundance of fleas on California ground squirrels (*Otospermophilus beecheyi*). *Journal of Vector Ecology* **36:** 117–123.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Joly S, Bruneau A. 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology* **55:** 623–636.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11:** 94.

Knowles LL, Carstens BC. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* **56:** 887–895.

Kyle CJ, Karels TJ, Clark B, Strobeck C, Hik DS, Davis CS. 2004. Isolation and characterization of microsatellite markers in hoary marmots (*Marmota caligata*). *Molecular Ecology Notes* **4:** 749–751.

Leaché AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London Series B, Biological Sciences* **277:** 3071–3077.

Linnen CR, Poh YP, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339:** 1312–1316.

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

May B, Gavin T, Sherman PW, Korves TM. 1997. Characterization of microsatellite loci in the northern Idaho ground squirrel *Spermophilus brunneus brunneus*. *Molecular Ecology* **6:** 399–400.

Mullen LM, Hoekstra HE. 2008. Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution* **62:** 1555–1570.

Nylander JAA. 2008. MrModelTest v2.3. Program distributed by author. Uppsala: Uppsala University, Evolutionary Biology Centre.

Padial JM, Aurélien M, De la Riva I, Vences M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* **7:** 1–14.

Padial JM, Castroviejo-Fisher S, Köhler J, Vilà C, Chaparro JC, De la Riva I. 2009. Deciphering the products of evolution at the species level: the need for an integrative taxonomy. *Zoologica Scripta* **38:** 431–447.

Patton JL, Da Silva MNF, Malcolm J. 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 JL, Huckaby DG, Álvarez-Castañeda ST. 2008. *The evolutionary history and a systematic revision of woodrats of the Neotoma lepida group*. Berkeley, CA: University of California Press.

Phuong MA, Lim MCW, Wait DR, Rowe KC, Moritz C. 2014. Data from: delimiting species in the genus, *Otospermophilus* (Rodentia: Sciuridae), using genetics, ecology, and morphology. *Dryad Digital Repository*. doi:10.5061/dryad.s4h2j.

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

R Development Core Team. 2012. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org.

Rambaut A, Drummond AJ. 2007. *Tracer*, Version 1.4. Computer program and documentation distributed by the author. Available at: http://beast.bio.ed.ac.uk/Tracer

Richardson J. 1829. *Fauna Boreali-Americana; or the zoology of the Northern Parts of British America*. London: J. Murray.

Riemann H, Ezcurra E. 2005. Plant endemism and natural protected areas in the peninsula of Baja California, Mexico. *Biological Conservation* **122:** 141–150.

Rios NE, Bart HL. 2010. GEOLocate*, Version 3.22.* Computer program. Bell Chasse, LA: Tulane University Museum of Natural History.

Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137–138.

Rosenberg NA. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution* **61:** 317–323.

Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi K, Moritz CC. 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* **11:** 1082–1092.

Samuels JX. 2009. Cranial morphology and dietary habits of rodents. *Zoological Journal of the Linnean Society* **156:** 864–888.

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology* **55:** 421–438.

Simpson GG. 1961. *Principles of animal taxonomy*. New York, NY: Columbia University Press.

Singhal S, Moritz C. 2012. Testing hypotheses for genealogical discordance in a rainforest lizard. *Molecular Ecology* **21:** 5059–5072.

Singhal S, Moritz C. 2013. Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society of London Series B, Biological Sciences* **280:** 1–8.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73:** 1162–1169.

Steppan SJ. 1998. Phylogenetic relationships and species

limits within *Phyllotis* (Rodentia: Sigmodontinae): concordance between mtDNA sequence and morphology. *Journal of Mammology* **79:** 573–593.

Steppan SJ, Storz BL, Hoffmann RS. 2004. Nuclear DNA phylogeny of the squirrels (Mammalia?: Rodentia) and the evolution of arboreality from c-myc and RAG1. *Molecular Phylogenetics and Evolution* **30:** 703–719.

Stevens S, Coffin J, Strobeck C. 1997. Microsatellite loci in Columbian ground squirrels *Spermophilus columbianus*. *Molecular Ecology* **6:** 493–495.

Swofford DL. 2002. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*, Version 4. Sunderland, MA: Sinauer Associates.

Thorington JR, Hoffmann RS. 2005. Family Sciuridae. In: Wilson DE, Reeder DM, eds. *Mammal species of the world: a taxonomic and geographic reference*, 3rd edn. Baltimore, MD: John Hopkins University Press, 754–818.

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

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62:** 2868–2883.

Wiley EO. 1978. The evolutionary species concept reconsidered. *Systematic Biology* **27:** 17–26.

Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. *Bioinformatics* **23:** 1851–1853.

Yang Z, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 9264–9269.

Zachos F, Apollonio M, Bärmann EV, Festa-Bianchet M, Göhlich U, Habel JC, Haring E, Kruckenhauser L, Lovari S, McDevitt AD, Pertoldi C, Rössner GE, Sánchez-Villagra MR, Scandura M, Suchentrunk F. 2012. Species inflation and taxonomic artefacts – a critical comment on recent trends in mammalian classification. *Mammalian Biology* **78:** 1–6.

Zhang C, Zhang DX, Zhu T, Yang Z. 2011. Evaluation of a bayesian coalescent method of species delimitation. *Systematic Biology* **60:** 747–761.

SHARED DATA

Data deposited in the Dryad digital repository (Phuong *et al*., 2014).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Guide topologies tested with BPP (Bayesian Phylogenetics & Phylogeography).
Figure S2. *K* versus delta *K* plot from STRUCTURE analyses.
Figure S3. *K* versus mean likelihood plot from STRUCTURE analyses.
Figure S4. STRUCTURE plots for $K = 4$ and $K = 5$.
Figure S5. Bayesian information criterion value versus number of clusters from discriminant analysis of principal components analysis.
Figure S6. Discriminant analysis of principal components Bayesian information criterion = 3, 4, 5, and 6 plots.
Figure S7. Maximum likelihood gene trees of nuclear sequence data.
Figure S8. Map of explicit spatial analyses of morphological variation.
Table S1. Samples used in the present study.
Table S2. Sequence marker information.
Table S3. Microsatellite marker information.
Table S4. WorldClim variables used in the present study.
Table S5. Skull characters measured in the present study.
Table S6. Summary statistics of genetic diversity and tests of neutrality from sequence data.
Table S7. Null allele frequency.
Table S8. Allelic richness and Hardy–Weinberg equilibrium results from microsatellite data.
Table S9. F_{ST} values from microsatellite data.
Table S10. Principal components analysis eigenvalues and loadings for WorldClim variables used in the present study.
Table S11. Niche similarity test comparison results and niche overlap values.
Table S12. Principal components analysis eigenvalues and loadings for the 23 morphological variables used in the present study.
Table S13. Multivariate analysis of covariance results for morphological data.
Table S14. Discriminant function analysis classifications.
Table S15. Substitution models inferred for mitochondrial and nuclear loci.
Appendix S1. Detailed methods for molecular techniques.
Species accounts
Appendix S2. GenBank accession numbers.