ORIGINAL ARTICLE

Biogeography of scorpions in the *Pseudouroctonus minimus* complex (Vaejovidae) from south-western North America: implications of ecological specialization for pre-Quaternary diversification

Robert W. Bryson Jr[1]*, Warren E. Savary[2] and Lorenzo Prendini[3]

[1]*Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Seattle, WA, 98195-1800, USA,* [2]*Department of Entomology, California Academy of Sciences, San Francisco, CA, 94118, USA,* [3]*Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, 10024-5192, USA*

ABSTRACT

Aim The aim of this study was to assess the impact of pre-Quaternary tectonics and orogeny relative to that of Pleistocene climate change on diversification within the *Pseudouroctonus minimus* complex, a group of vaejovid scorpions with stenotopic habitat requirements.

Location South-western North America (United States and Mexico).

Methods Multilocus sequence data (1899 base pairs from two mitochondrial and two nuclear genes) were generated from 65 samples of scorpions in the *minimus* complex. Phylogeographical structure within the *minimus* complex was explored using model-based phylogenetic methods and a general mixed Yule coalescent model to identify independent geographical clusters. A time-calibrated multilocus species tree was reconstructed using a multispecies coalescent approach. Ancestral areas were estimated at divergence events across the tree using a probabilistic Bayesian approach.

Results Extensive geographical structure was evident within two well-supported clades. These clades probably diverged over 25 million years ago (Ma), based on estimated mean divergence dates, followed by 14 divergences in the Miocene (25–5 Ma) and 4 divergences in the Pliocene and Pleistocene (< 5 Ma). The ancestral origin of the *minimus* complex was reconstructed to be across California and the Mexican Highlands. The Chihuahuan Desert was colonized twice from the Mexican Highlands, and one dispersal event occurred from the Mexican Highlands back to California.

Main conclusions Spatial and temporal patterns of evolution in the *minimus* complex support predictions that stenotopy promoted pre-Quaternary diversification. Miocene and Pliocene geomorphology, perhaps in concert with climate change, induced allopatric divergence across the heterogeneous landscape of south-western North America. Stenotopic scorpions such as the *minimus* complex provide a model for exploring correlations between Earth history and biological diversification.

Keywords
Biogeography, diversification, North America, phylogeography, *Pseudouroctonus*, speciation, stenotopy.

*Correspondence: Robert W. Bryson Jr, Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Box 351800, Seattle, WA 98195-1800, USA.
E-mail: brysonjr@uw.edu

INTRODUCTION

The amount of genetic structure in a taxon across a landscape is often correlated with its ecological requirements and dispersal ability (Avise, 2000). Specialized ecological requirements (stenotopy) that exert a consistent and repeated influence upon gene flow among populations may influence diversification across space and time (e.g. Cooper *et al.*, 2011; Derkarabetian *et al.*, 2011; Hamilton *et al.*, 2011; Keith & Hedin, 2012), a prediction of the 'effect hypothesis of

macroevolution' (Vrba, 1980). Many scorpions are characterized by low dispersal ability and stenotopic habitat requirements, increasing their propensity for diversification in association with long-term processes such as geomorphological development and climatic cycles (Prendini, 2001, 2005). Stenotopic scorpions with specialized habitat requirements may disperse only a few metres a year (Polis *et al.*, 1985), and local populations may diverge allopatrically on small spatial scales when there is decreased gene flow across barriers of unsuitable habitat. Consequently, stenotopic scorpion taxa tend to be more species-rich and range-restricted than eurytopic scorpion taxa with generalized ecological requirements (Prendini, 2001).

This study examined the *Pseudouroctonus minimus* complex, a monophyletic group of stenotopic, rupicolous scorpions in the family Vaejovidae Thorell, 1876. The *minimus* complex – here defined as comprising the species *Pseudouroctonus andreas* (Gertsch and Soleglad, 1972), *P. apacheanus* (Gertsch and Soleglad, 1972), *P. chicano* (Gertsch and Soleglad, 1972), *P. minimus* (Kraepelin, 1911), *P. rufulus* (Gertsch and Soleglad, 1972), *P. savvasi* Francke, 2009 and possibly *P. cazieri* (Gertsch and Soleglad, 1972) – is restricted to humid, rocky habitats throughout south-western North America from southern California and northern Baja California across southern Arizona, New Mexico and western Texas, and southwards into the Mexican states of Sonora, Chihuahua and Coahuila (Sissom 2000; Francke & Savary, 2006).

The complex is distributed across three distinct biogeographical regions (Fig. 1). *Pseudouroctonus andreas*, *P. rufulus* and *P. minimus* (including its subspecies *P. m. castaneus* and *P. m. thompsoni*) occur in the Channel Islands and coastal chaparral of southern California and northern Baja California. Several hundred kilometres to the east, *P. apacheanus* and *P. chicano* are distributed across the mesic mixed pine-oak woodlands of the Sierra Madre Occidental and the associated sky-island outliers in southern Arizona and adjacent New Mexico, extending southwards into Chihuahua and Sonora. *Pseudouroctonus apacheanus* is also found in western Texas on several disjunct mountains and along patches of mesic rocky upland habitat within the Chihuahuan Desert. The troglophilous *P. savvasi* inhabits two isolated caves at the eastern edge of the Chihuahuan Desert.

In this study, species of the *minimus* complex were predicted to retain genetic traces of regional history over relatively small geographical distances because of their stenotopic habitat requirements. Much of the modern landscape of south-western North America evolved prior to the Quaternary (Wilson & Pitts, 2010). The cyclical climate change that characterized the Pleistocene dramatically transformed biotic communities across south-western North America (Webb & Betancourt, 1990), but may have had little effect on the diversification of scorpions in the *minimus* complex closely associated with abiotic rocky habitats. Despite the presence of ephemeral woodland corridors across
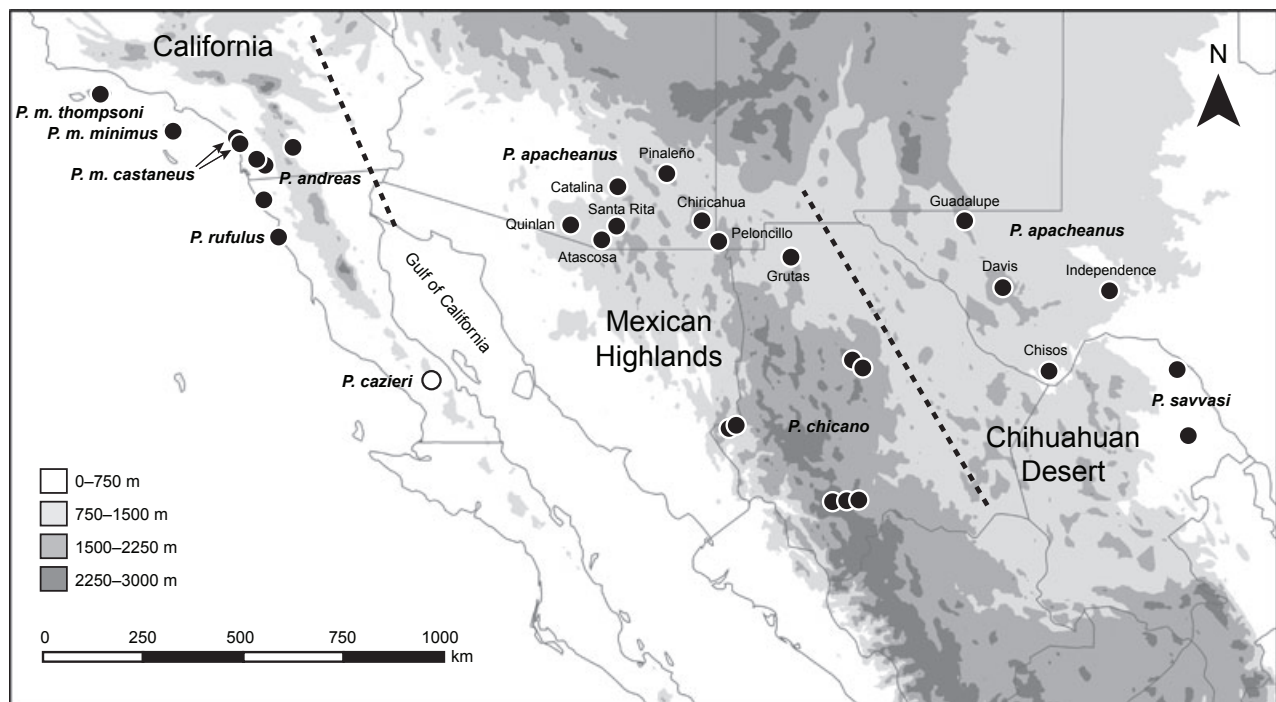
Figure 1 Map of south-western North America plotting collection localities (black circles) of genetic samples of vaejovid scorpions of the *Pseudouroctonus minimus* complex. Putative species and subspecies are indicated, as well as the geographical localities of *P. apacheanus* mentioned in the text. The locality of *P. cazieri*, a possible member of the *minimus* complex missing from the study, is indicated with a white circle. Approximate boundaries between the three main biogeographical regions inhabited by these scorpions are indicated with dashed lines.

much of south-western North America during cooler Pleistocene glacial periods (McCormack *et al.*, 2008; Bryson *et al.*, 2011), the dispersal of scorpions in the *minimus* complex was probably limited to corridors containing rocky habitat. Accordingly, the pre-Quaternary tectonics and orogeny that created and altered rocky terrain across the south-western North American landscape may have had a greater impact on diversification than Pleistocene climate change in these stenotopic scorpions.

These expectations were tested using a molecular phylogenetic approach. First, the phylogeographical structure within the *minimus* complex was explored. Second, a time-calibrated multilocus species tree was reconstructed, and ancestral areas were estimated at divergence events across the tree. The resulting patterns of diversification are discussed in the context of the geomorphological evolution of south-western North America.

MATERIALS AND METHODS

Genetic data

DNA sequence data were generated from 65 samples of scorpions in the *minimus* complex collected throughout the known range (Fig. 1 and Appendix S1 in Supporting Information). The composition of the *minimus* complex was inferred from multilocus and morphological data for samples representing all genera and most species in the family Vaejovidae (http://www.vaejovidae.com/, in preparation). All described species known to belong to the monophyletic *minimus* complex were included (*P. andreas*, *P. apacheanus*, *P. chicano*, *P. minimus*, *P. rufulus* and *P. savvasi*), as well as both subspecies of *P. minimus* (*P. m. castaneus* and *P. m. thompsoni*). *Pseudouroctonus cazieri*, from the Baja California Peninsula, may also belong to the *minimus* complex, but its placement has not been confirmed, as samples could not be obtained for molecular analysis. *Pseudouroctonus reddelli* and *P. williamsi* were included as outgroups (Stockwell, 1989).

Genomic DNA was extracted from leg muscle tissue. Fragments of mitochondrial DNA were sequenced from a protein-coding gene (cytochrome *c* oxidase subunit I, *COI*) and a ribosomal gene (16S rDNA, *16S*) using scorpion-specific primers (Appendix S2). Two nuclear genes, including fragments of the 28S rDNA (*28S*) and the internal transcribed spacer region (*ITS2*) between the 5.8S and 28S rDNAs, were sequenced for a subset of samples ($n = 44$) that were used in species tree analyses (see below). Primer sequences for nuclear genes were obtained from Tully *et al.* (2006) for *28S* and from Ji *et al.* (2003) for *ITS2*. The laboratory protocols used to generate the sequence data are provided in Prendini *et al.* (2003, 2005) and Bryson & Riddle (2012).

Heterozygous sites were identified in nuclear segments when two different nucleotides were observed at the same position in electropherograms of both strands, with the weaker peak reaching at least 50% of the strongest signal.

The gametic phase of the variants was determined computationally using PHASE 2.1.1 (Stephens & Donnelly, 2003). Five separate runs of 400 iterations each were conducted for each nuclear data set, and results with a probability threshold of 0.7 or greater were accepted. All polymorphic sites with a probability < 0.7 were coded in both alleles with the appropriate IUPAC ambiguity code. INDELLIGENT 1.2 (Dmitriev & Rakitov, 2008) was used to resolve insertion/deletion events between homologous nuclear alleles in ITS2. Sequence alignments for individual gene regions were performed with MAFFT 6 (Katoh *et al.*, 2002; Katoh & Toh, 2008) using default settings, the '20PAM/k = 2' scoring matrix for nucleotide sequences, the Q-INS-i algorithm for *16S* and *28S* data, and the E-INS-i algorithm for ITS2 data. The ITS2 gene region was highly variable with a large number of indels among the various sequences, so ambiguously aligned regions were eliminated using GBLOCKS 0.91b (Talavera & Castresana, 2007). A partitioned homogeneity test was conducted in PAUP* 4.0b10 (Swofford, 2002) with 1000 heuristic replicates, to test for conflicting phylogenetic signals between the mitochondrial and nuclear genes used in the species tree analyses.

Phylogeographical estimation

The full mtDNA data set ($n = 67$) was analysed to examine geographical structure and delineate geographically cohesive lineages within the *minimus* complex. Geographical structure was inferred using Bayesian inference and maximum-likelihood phylogenetic methods. Bayesian inference analyses were conducted with MRBAYES 3.2.1 (Ronquist *et al.*, 2012). MRMODELTEST 2.1 (Nylander, 2004) was used to select best-fit models of evolution, based on the Akaike information criterion (Akaike, 1973), for the *COI* and *16S* gene regions. Analyses were conducted with three heated (temperature = 0.05) and one cold Markov chain, sampling every 100 generations for 4 million generations. Adjustment of the heated chain temperature from the default value of 0.2 to 0.05 resulted in higher harmonic mean log-likelihoods and better convergence and mixing. Parameters were unlinked across partitions, and the gamma-shaped rate variation was set to variable. Output parameters were visualized using TRACER 1.5 (Rambaut & Drummond, 2007) to ascertain stationarity and convergence. All samples obtained during the first million (25%) generations were discarded as burn-in. Maximum likelihood analyses were conducted using RAxML 7.2.6 (Stamatakis, 2006) under the GTRGAMMA model, with 1000 nonparametric bootstrap replicates to assess nodal support.

A general mixed Yule coalescent (GMYC) model was used to delineate geographically distinct clusters of samples (henceforth 'lineages'), implemented in the R command-line package (R Development Core Team, 2011) SPLITS (Pons *et al.*, 2006). The GMYC model identifies independent evolutionary clusters by detecting a threshold value at the transition from interspecific to intraspecific branching patterns, and provides an objective means of delimiting the 'species'

required for species tree reconstructions (see below). An ultrametric tree for use in SPLITS was generated with BEAST 1.7.4 (Drummond *et al.*, 2012). Separate models of evolution were used for the *COI* and *16S* gene regions. Analyses were run for 40 million generations, with samples retained every 1000 generations, using a Yule tree prior. Results were displayed in TRACER to confirm acceptable mixing and likelihood stationarity, appropriate burn-in, and adequate effective sample sizes above 200 (Drummond *et al.*, 2007) for all estimated parameters. After discarding the first 4 million generations (10%) as burn-in, the parameter values of the samples from the posterior distribution were summarized on the maximum clade credibility tree using TREEANNOTATOR 1.7.4 (Drummond *et al.*, 2012). Analyses using single- and multiple-threshold models were performed with SPLITS. The multiple-threshold model tested whether allowing the species-coalescent transition to vary across the tree significantly improved the fit to the model (Monaghan *et al.*, 2009).

Species tree and divergence date estimation

A time-calibrated species tree was reconstructed for the *minimus* complex from the multilocus data set using *BEAST (Heled & Drummond, 2010; Drummond *et al.*, 2012), a part of the BEAST package. One or two exemplar samples ($n = 44$) were selected from each geographically delimited mtDNA lineage identified by the GMYC analysis, and outgroups were excluded. These GMYC-delineated lineages satisfy the operational requirements of 'species' for the *BEAST analyses (Heled & Drummond, 2010). Best-fit models of evolution were selected using MRMODELTEST, with a Yule speciation prior and relaxed uncorrelated lognormal clocks applied for each gene tree. In analyses of the *COI* and *16S* mtDNA data, which represent a single locus, trees were linked but substitution and clock models were unlinked. The clock was calibrated using scorpion-specific mutation rates of 5×10^{-3} substitutions/site/Myr for *16S* (Gantenbein & Largiadèr, 2003) and of 7×10^{-3} substitutions/site/Myr for *COI* (Gantenbein *et al.*, 2005). Clock rates for the *28S* and ITS2 genes were estimated relative to the mtDNA rates. This 'scorpion clock' was calculated from buthid scorpions, distantly related to the vaejovids of the *minimus* complex, as no other mutation rates have been estimated for scorpions to date. Analyses were run for 8×10^{7} generations, with samples retained every 1000 generations. Results were displayed in TRACER to confirm acceptable mixing and likelihood stationarity, appropriate burn-in, and adequate effective sample sizes. The first 10% of generations were discarded as burn-in, and parameter estimates were summarized on the maximum clade credibility tree using TREEANNOTATOR. This burn-in and visualization procedure was repeated for each of the three gene trees co-estimated by *BEAST.

Ancestral area reconstruction

The ancestral range at each divergence event was reconstructed using Bayesian binary Markov chain Monte Carlo analysis (BBM) as implemented in RASP 2.0b (Yu *et al.*, 2011). This program determines the probability of an ancestral range at a node by averaging over a posterior set of trees, thereby accounting for phylogenetic uncertainty. A total of 40,000 post-burn-in trees were loaded from the *BEAST analyses into RASP. Each sample from the phylogeny was assigned to one of three broad biogeographical regions (Fig. 1; Udvardy, 1975): (1) the coastal chaparral of southern California (including the Channel Islands) and north-western Baja California (hereafter referred to as 'California'); (2) the Sierra Madre Occidental and associated sky islands ('Mexican Highlands'); and (3) upland areas within the Chihuahuan Desert ('Chihuahuan Desert'). The probabilities for nodes in the phylogeny with posterior probability > 0.50 were estimated. The number of areas was set to three, a F81 + G model was used, and analyses were conducted for 1 million generations using 10 chains, sampling every 100 generations. Hypothetical outgroups assigned to the phylogeny by the BBM algorithm prior to the analyses were given a 'wide' distribution across all three biogeographical regions. The first 25% of generations were discarded as burn-in.

RESULTS

Genetic data

The complete mtDNA data set contained 321 parsimony-informative sites (*COI*: 211 in 756 bp; *16S*: 110 in 397 bp). The reduced mtDNA data sets for the species tree analysis contained 200 (*COI*) and 100 (*16S*) parsimony-informative sites, respectively. The nuclear gene loci exhibited much less variation than the mtDNA (parsimony-informative sites: *28S*, 7 in 519 bp; ITS2, 46 in 227 bp). Complete sequence data could not be obtained for two samples: *P. andreas* Mission Trails CA (ITS2) and *P. apacheanus* Vallecito SON (*COI* and ITS2). The *28S* gene contained three heterozygous sites, all confidently resolved. The ITS2 gene contained eight heterozygous sites. Two of those sites in one individual (*P. apacheanus* Pinaleño AZ 1) were not resolved above the 0.7 acceptance threshold. The partitioned homogeneity test revealed no significant conflict between the mitochondrial and nuclear genes in the reduced data set ($P = 0.96$). GTR+I+G models of sequence evolution were selected for both genes in the mtDNA tree data set, and GTR+I+G (*COI*, *16S*), GTR+I (*28S*) and GTR+G (ITS2) models for the species tree data sets. All aligned sequences were deposited in the Dryad repository: doi:10.5061/dryad.q58r0.

Phylogeographical estimation

Extensive geographical structure in the *minimus* complex was evident within two well-supported major clades (Fig. 2). One clade (referred to as the 'NW clade') comprised *P. andreas*, *P. minimus*, and its subspecies. The second clade ('SE clade') comprised *P. apacheanus*, *P. chicano*, *P. rufulus* and *P. savvasi*.
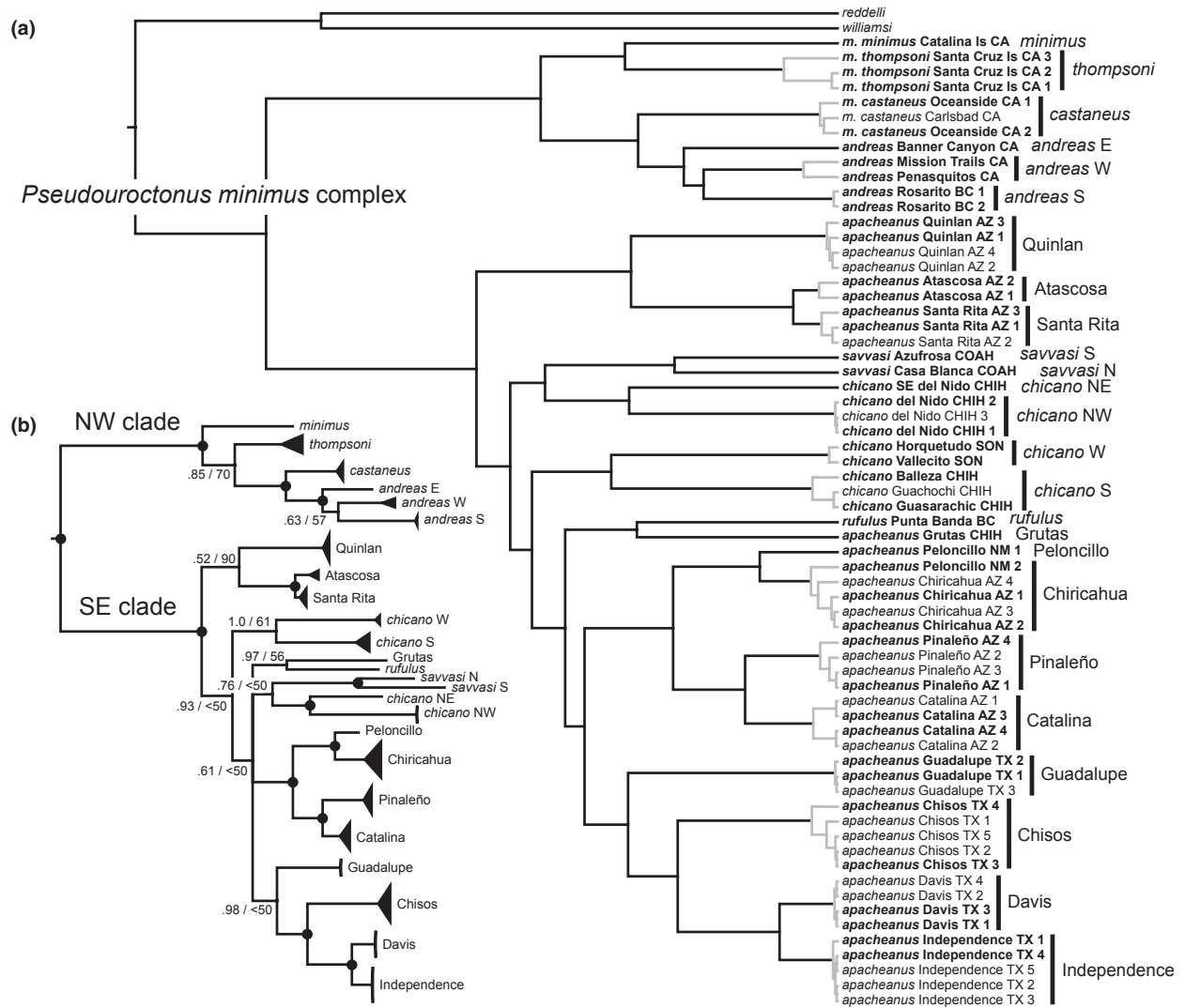
Figure 2 Geographical structure within North American vaejovid scorpions of the *Pseudouroctonus minimus* complex inferred from phylogenetic analyses of 1153 base pairs of mitochondrial DNA. (a) Ultrametric tree generated with BEAST. General mixed Yule coalescent (GMYC) groups delimited from the single-threshold model are designated by grey branches. Samples in bold were used in multilocus species tree reconstructions. No nodal support values are shown for this tree because the focus is on GMYC cluster delimitation. (b) Maternal genealogy based on mixed-model Bayesian inference (tree shown) and maximum likelihood analyses. All major nodes that received ≥ 0.95 Bayesian posterior probability and ≥ 70% bootstrap support are depicted with black dots. Support at all other nodes is denoted by the Bayesian posterior probability followed by maximum likelihood bootstrap values. All lineages identified by the GMYC analysis received ≥ 0.95 posterior probability and ≥ 70% bootstrap support. GMYC groups are collapsed for clarity.

The single-threshold and multiple-threshold GMYC models were not significantly different from each other ($\chi^2 = 1.9109$, d.f. = 9, $P = 0.9927$), so results from the single-threshold model were used to infer geographically delimited lineages within the *minimus* complex (Fig. 2a). Independent evolutionary clusters estimated by the GMYC model were largely concordant with geography. Individuals from isolated mountain ranges and regions in close geographical proximity were monophyletic with the exception of *P. m. thompsoni*. Although the GMYC model split this taxon into two groups, it was treated as one lineage because of its restricted distribution in the northern Channel Islands. *Pseu-*

douroctonus andreas, *P. apacheanus*, *P. chicano* and *P. savvasi* each contained several geographically delimited lineages (including singletons, hereafter referred to as 'lineages' for convenience). *Pseudouroctonus andreas* comprised three lineages, one on the eastern and one on the western slopes of the Peninsular Range in California, and another west of the Peninsular Range in north-western Baja California. The widespread *P. apacheanus* comprised 15 distinct geographically delimited lineages. Eleven of these were grouped within three regional clades. The westernmost clade included three lineages from the Quinlan, Santa Rita and Atascosa mountain ranges (Fig. 1), to the north-east of which was a clade

comprising four lineages from the Catalina, Pinaleño, Chiricahua and Peloncillo Mountains. The Chiricahua lineage included four samples from the Chiricahua Mountains and one from the Peloncillo Mountains. Further east, a third regional clade included four Texas lineages from Independence Creek and the Guadalupe, Davis and Chisos Mountains. The last remaining lineage of *P. apacheanus* comprised one sample from a cave in north-western Chihuahua ('Grutas'). *Pseudouroctonus chicano* formed two lineages in the Sierra del Nido, Chihuahua, one at higher elevations in the west and the other (represented by one sample) at a lower elevation in the east. Two other lineages of *P. chicano* were found, in south-eastern Sonora and further south in the Sierra Madre Occidental, respectively. Both samples of the troglophile *P. savvasi*, collected from isolated caves, were also genetically distinct.

Pairwise sequence divergences between sister lineages, calculated with mega 5.05 (Tamura *et al.*, 2011) using a maximum composite likelihood correction, were relatively high. Pairwise divergences between deeply divergent lineages ranged from 8.2% (*P. chicano* from the eastern and western slopes of the Sierra del Nido) to 8.3% (*P. m. minimus* and *P. m. thompsoni*), and those between more closely related lineages ranged from 2.2% (*P. apacheanus* from the Atascosa and Santa Rita mountain ranges) to 3.0% (*P. apacheanus* from the Davis Mountains and Independence Creek).
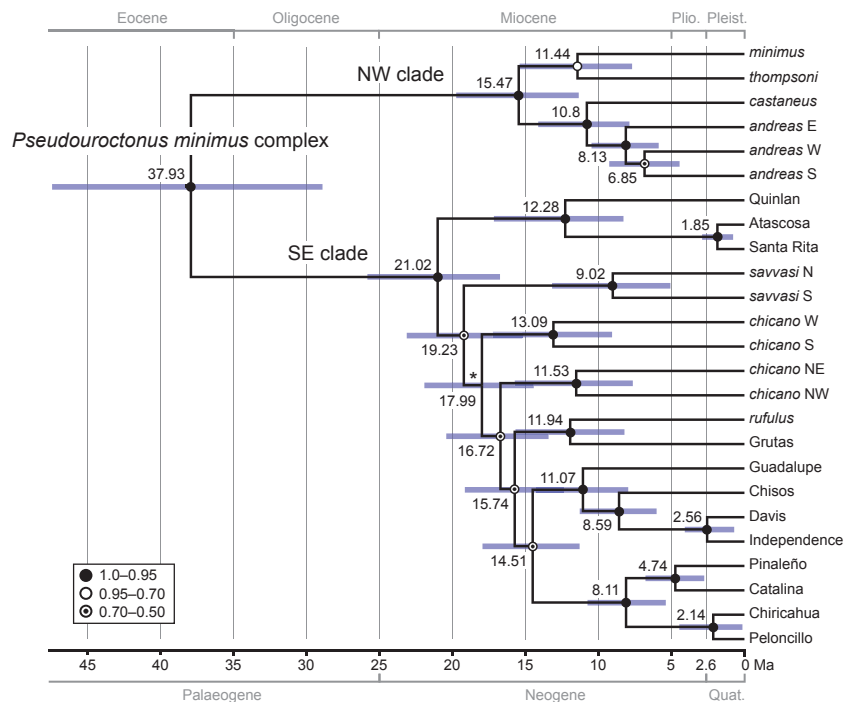
Species tree and divergence times

The topologies of the gene trees were similar and revealed a consistent lack of support at the base (Appendix S3). Mono-

phyly of the NW and SE clades was strongly supported in the species tree reconstruction (Fig. 3). The two clades probably diverged prior to 25 Ma, during the Palaeogene (mean estimated date 37.9 Ma, 95% posterior credibility interval = 47.5–28.9 Ma). A subsequent burst of diversification in the SE clade occurred *c.* 14–21 Ma. During this 7-Myr interval, an estimated six divergence events occurred (based on mean divergence date estimates). A further eight divergences occurred within the SE clade during a second burst of diversification *c.* 13–8 Ma, followed by four final divergences, one near the Miocene/Pliocene boundary *c.* 4 Ma, and three during the Pleistocene. Estimated divergences within the NW clade all occurred during the Miocene from *c.* 15 to 6 Ma.
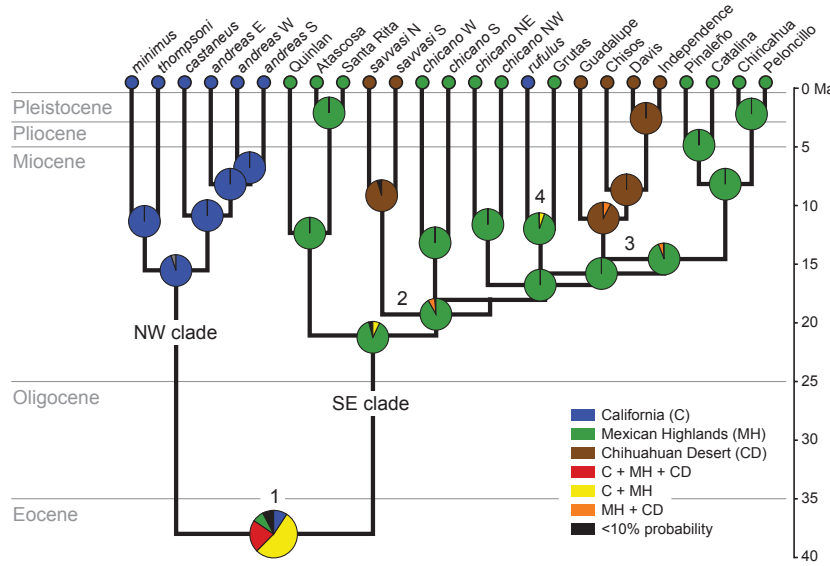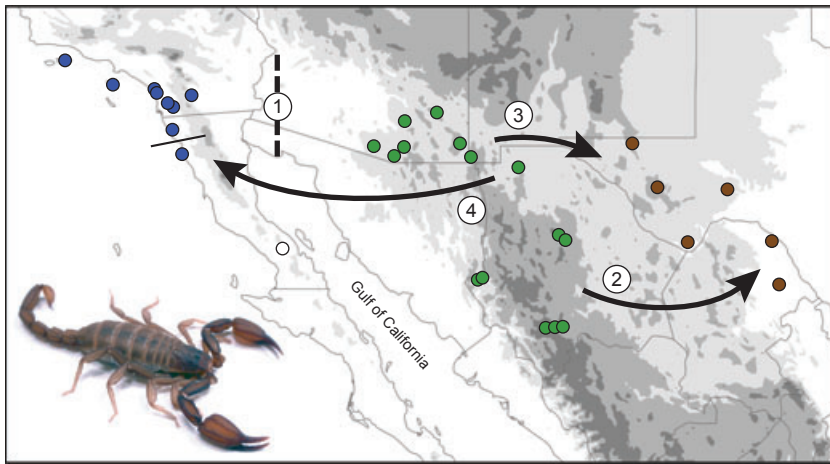
Ancestral area reconstruction

The rasp analysis (Fig. 4) supported a widespread ancestral origin of the *minimus* complex across California and the Mexican Highlands. The marginal probability for this basal node reconstruction (*P* = 53%) was much higher than that for alternative geographical areas (California + Mexican Highlands + Chihuahuan Desert, *P* = 22%; California, *P* = 9%; Mexican Highlands, *P* = 8%; and three other area combinations, *P* < 5% each). One major vicariance event and three major dispersal events appear to have occurred. The basal divergence between the NW clade and the SE clade was probably caused by a vicariance event dating to the Eocene. The Chihuahuan Desert appears to have been colonized twice from the Mexican Highlands by members of the SE clade (Fig. 4), whereas one dispersal event occurred from the Mexican Highlands back to California.

Figure 3 Time-calibrated multilocus species tree for North American vaejovid scorpions of the *Pseudouroctonus minimus* complex. Bars indicate 95% highest posterior densities of divergence dates, with mean estimates in millions of years ago (Ma) given at nodes. Bayesian posterior probability support values for nodes are indicated by coded dots. The single node with a posterior probability below 0.50 is marked with an asterisk. Abbreviations: Pleist., Pleistocene; Plio., Pliocene; Quat., Quaternary.

Figure 4 Dated multilocus phylogeny for North American vaejovid scorpions of the *Pseudouroctonus minimus* complex, showing ancestral area reconstructions. Pie charts indicate the probability of ancestral area for nodes that received ≥ 0.50 posterior probability support. Sample localities are colour-coded to match the three biogeographical regions indicated in Fig. 1. The locality of *P. cazieri*, a possible member of the *minimus* complex missing from the study, is indicated with a white circle. The dashed line and arrows indicate interpretations of four major historical events that affected diversification. The thin solid line indicates the break between *P. andreas* (above the line) and *P. rufulus* (below) at the end of a possible ring distribution formed by the *minimus* complex.

DISCUSSION

Pre-Quaternary diversification

Scorpions of the *minimus* complex appear to be ancient inhabitants of south-western North America. Although the estimated divergence dates are remarkably old and based on a molecular clock rate obtained from a distantly related taxon, the dates of many divergence events within the complex correspond well with geological events and to geologically mediated vicariance. Divergence of the NW and SE clades is estimated to have been prior to 28 Ma, during the Palaeogene (Figs 3 & 4), coincident with dramatic landscape deformation in southern California and the development of the San Andreas fault system at *c.* 30 Ma, following the collision of the Pacific-Farallon and North American plates (Atwater, 1998). The subsequent formation of the Catalina and Channel Islands, *c.* 18–12 Ma (Atwater, 1998; Schoenherr *et al.*, 1999), corresponds well with the basal divergence of the NW clade at *c.* 15 Ma, and to the subsequent isolation of the two insular forms, *P. m. minimus* and *P. m. thompsoni*, at *c.* 11 Ma (Figs 3 & 4). The later diversification of

P. andreas in southern California and northern Baja California may be related to the faulting along the Peninsular Range that began in the late Miocene (Axen & Fletcher, 1998).

Diversification within the widespread SE clade appears to have occurred almost entirely within the Neogene period, in the Miocene and Pliocene. Fifteen of the estimated 18 divergence events (83%) pre-date the Quaternary (Fig. 3). The base of the SE clade is characterized by five weakly supported nodes in the species tree reconstruction, suggesting a rapid radiation. This burst of diversification *c.* 21–14 Ma corresponds with regional geomorphological change and climate shifts during the Miocene, 24–12 Ma (Shafiqullah *et al.*, 1980; Henry & Aranda-Gomez, 2000; Brand & Stump, 2011). Marked extensions of the northern Mexican Highlands occurred during this time (Henry & Aranda-Gomez, 2000; Connell *et al.*, 2005; Brand & Stump, 2011), coincident with the onset of a wetter climate (Retallack, 2001). The ancestral area reconstruction suggests that dispersal and colonization of the Chihuahuan Desert occurred during this period (Fig. 4), perhaps triggered by changing ecosystems associated with the wetter climate (Retallack, 2001). However, at least one of these events may have been caused by allopatric frag-

mentation associated with the mid-Miocene extension of the Rio Grande rift (Fig. 4; Henry & Aranda-Gomez, 2000; Connell *et al.*, 2005).

The divergence of *P. rufulus* on the north-western coast of Baja California from its sister taxon, *P. apacheanus* in northern Chihuahua, at *c.* 12 Ma (mean 11.9 Ma, 95% posterior credibility interval 15.7–8.2 Ma), corresponds well with tectonic plate movements during this time (Henry & Aranda-Gomez, 2000) and to the estimated development of the proto-Gulf of California, *c.* 13–12 Ma (Henry & Aranda-Gomez, 2000; Zhang & Paulssen, 2012). Although the RASP analysis suggested a dispersal event to explain this split, the isolation of *P. rufulus* may be the result of vicariance associated with the separation of the Baja California Peninsula and subsequent rifting along the Pacific plate to the north-west (a passive dispersal event). The dates of this rifting are debated (Umhoefer, 2011), but the two most widely accepted models span 12.5–6 Ma (Oskin *et al.*, 2001; Fletcher *et al.*, 2007), encompassing most of the 95% posterior credibility interval of the estimated date of divergence.

Habitat specialization

Extrinsic habitat features and intrinsic organismal traits synergistically influence the distribution of taxa across landscapes. Species in the *minimus* complex are restricted to humid, rocky habitats throughout south-western North America (Sissom, 2000; Francke & Savary, 2006). The distinct pattern of pre-Quaternary diversification within the complex inferred in this study is probably related to the specialized ecological requirements of these scorpions. With one exception, samples from all mountain ranges in the SE clade form monophyletic groups (Fig. 2a), consistent with the prediction that gene flow between populations on adjacent mountain ranges may have been limited during the Pleistocene. The cooler Pleistocene glacial periods and down-slope expansions of mixed pine-oak woodlands may have provided only one of the habitat requirements that were needed for dispersal. If suitable rocky habitat was absent, dispersal by these stenotopic rupicolous scorpions would have been limited. The sole exception was observed in samples of *P. apacheanus* from the Peloncillo and Chiricahua Mountains (Fig. 2a). One of the two haplotypes from the Peloncillo Mountains grouped with haplotypes from the adjacent Chiricahua Mountains. The closest slopes of these mountain ranges are separated by about 10 km of xeric grassland. At the southern end of the Chiricahua Mountains, however, a low-elevation lava field, stretching 25 km, almost connects to the Peloncillo Mountains. Although perhaps uninhabited by *P. apacheanus* today, this rocky habitat may have provided a stepping stone during cooler Pleistocene glacial periods, permitting gene flow between the two otherwise isolated populations of *P. apacheanus* in these mountains. Future studies with larger sample sizes from each mountain range could better determine whether the inferred pattern of limited gene flow between adjacent mountain ranges is a result of the

relatively few samples per mountain range sequenced in this study.

The two sampled localities of troglophilous *P. savvasi* are separated by 127 km of xeric habitat (Fig. 1). Dating estimates place the divergence between these samples at *c.* 9 Ma (Fig. 3), suggesting a long period of isolation, probably caused by the first significant expansions of semi-arid and arid habitats across south-western North America that started during the late Miocene (Axelrod, 1950, 1979; Graham, 1999). Although *P. savvasi* possesses no evident troglomorphies (Francke, 2009), its reliance on a cool humid cave environment may have prevented subsequent expansions across the more suitable mesic habitats that extended across much of the Chihuahuan Desert during Pleistocene glacial times (Morafka, 1977). The relatively ancient split between *P. chicano* populations in the Sierra del Nido might reflect divergences between a highland form (del Nido W) and an epigean troglophile on the lower-elevation slopes (del Nido E). The Sierra del Nido represents an eastern outlier of the Sierra Madre Occidental flanked by Chihuahuan Desert scrub. Cavernicole forms are found at low elevations to the north-west (*P. apacheanus* Grutas) and east (*P. savvasi*) of these mountains, suggesting that they are or were more widespread and have gone extinct or undetected.

Fragments of an ancient ring distribution?

Although few areas of the world present the geographical conditions necessary for ring speciation (Mayr, 1963), recent research suggests that the unique geomorphology of the Baja California Peninsula may have created ring distributions in several taxa (Mulcahy & Macey, 2009). Based on the inferred close relationship between two samples from widely disjunct localities (*P. apacheanus* from Grutas el Sabinal in northern Chihuahua, and *P. rufulus* from north-western Baja California), scorpions of the *minimus* complex appear to fit this pattern of distribution, forming a ring around the Gulf of California (Fig. 4), the terminus of which is situated in northern Baja California where the distributions of the two species come into proximity (Fig. 4). This ring distribution may have been intact until the mid-Miocene separation of the Baja California Peninsula from the Mexican mainland severed habitat connectivity. Future studies should test this hypothesis with additional samples of *P. apacheanus* from Sonora and of *P. rufulus* from Baja California, and with the inclusion of *P. cazieri*, a possible member of the *minimus* complex, also from Baja California, which could not be obtained for the present study.

CONCLUSIONS

Many aspects of biogeography seek to explain the distributions of species in terms of historical factors and contemporary ecology. Stenotopic scorpions such as the *minimus* complex offer a model for exploring correlations between Earth history and biological diversification. Spatial and tem-

poral patterns of evolution in the *minimus* complex suggest that stenotopy promoted pre-Quaternary diversification in these rupicolous scorpions. The heterogeneous landscape of south-western North America provided the backdrop for diversification. Miocene and Pliocene geomorphology, perhaps in concert with climate change, appears to have induced allopatric divergence on a small spatial scale.

ACKNOWLEDGEMENTS

REFERENCES

Akaike, H. (1973) Information theory as an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* (ed. by B.N. Petrov and F. Csaki), pp. 276–281. Akadémiai Kiadó, Budapest.

Atwater, T. (1998) Plate tectonic history of southern California with emphasis on the western Transverse Ranges and Santa Rosa Island. *Contributions to the geology of the northern Channel Islands, southern California* (ed. by P.W. Weigand), pp. 1–8. American Association of Petroleum Geologists, Pacific Section, Bakersfield, CA.

Avise, J. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Axelrod, D.I. (1950) Evolution of desert vegetation in western North America. *Carnegie Institute of Washington Publications*, **590**, 215–306.

Axelrod, D.I. (1979) Age and origin of Sonoran Desert vegetation. *Occasional Papers of the California Academy of Sciences*, **132**, 1–74.

Axen, G.J. & Fletcher, J.M. (1998) Late Miocene–Pleistocene extensional faulting, northern Gulf of California, Mexico and Salton Trough, California. *International Geology Review*, **40**, 217–244.

Brand, P. & Stump, E. (2011) *Tertiary extension and fault block rotation in the transition zone, Cedar Mountains Area, Arizona v.1.1*. Arizona Geological Survey, Tucson, AZ.

Bryson, R.W. & Riddle, B.R. (2012) Tracing the origins of widespread highland species: a case of Neogene diversification across the Mexican sierras in an endemic lizard. *Biological Journal of the Linnean Society*, **105**, 382–394.

Bryson, R.W., Murphy, R.W., Graham, M.R., Lathrop, A. & Lazcano-Villareal, D. (2011) Ephemeral Pleistocene woodlands connect the dots for highland rattlesnakes of the *Crotalus intermedius* group. *Journal of Biogeography*, **38**, 2299–2310.

Connell, S.D., Hawley, J.W. & Love, D.W. (2005) Late Cenozoic drainage development in the southeastern basin and range of New Mexico, southeasternmost Arizona, and western Texas. *New Mexico's ice ages* (ed. by S.G. Lucas, G.S. Morgan and K.E. Zeigler), pp. 125–150. New Mexico Museum of Natural History and Science, Albuquerque, NM.

Cooper, S.J.B., Harvey, M.S., Saint, K.M. & Main, B.Y. (2011) Deep phylogeographic structuring of populations of the trapdoor spider *Moggridgea tingle* (Migidae) from southwestern Australia: evidence for long-term refugia within refugia. *Molecular Ecology*, **20**, 3219–3236.

Derkarabetian, S., Ledford, J. & Hedin, M. (2011) Genetic diversification without obvious genitalic morphological divergence in harvestmen (Opiliones, Laniatores, *Sclerobunus robustus*) from montane sky islands of western North America. *Molecular Phylogenetics and Evolution*, **61**, 844–853.

Dmitriev, D.A. & Rakitov, R.A. (2008) Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Computational Biology*, **4**, e1000113.

Drummond, A.J., Ho, S.Y.W., Rawlence, N. & Rambaut, A. (2007) *A rough guide to BEAST 1.4*. Available at: http://beast.bio.ed.ac.uk/Main_Page/ (accessed 1 July 2012).

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Fletcher, J.M., Grove, M., Kimbrough, D., Lovera, O. & Gehrels, G.E. (2007) Ridge–trench interactions and the Neogene tectonic evolution of the Magdalena Shelf and southern Gulf of California; insights from detrital zircon U-Pb ages from the Magdalena Fan and adjacent areas. *Geology Society of America Bulletin*, **119**, 1313–1336.

Francke, O.F. (2009) Description of a new species of troglophile *Pseudouroctonus* (Scorpiones: Vaejovidae) from Coahuila, Mexico [Descripción de una nueva especie de *Pseudouroctonus* troglófilo (Scorpiones: Vaejovidae) de Coahuila, México]. *Texas Memorial Museum, Speleological Monographs*, **7**, 11–18.

Francke, O.F. & Savary, W.E. (2006) A new troglobitic *Pseudouroctonus* Stahnke (Scorpiones: Vaejovidae) from northern Mexico. *Zootaxa*, **1302**, 21–30.

Gantenbein, B. & Largiadèr, C.R. (2003) The phylogeographic importance of the Strait of Gibraltar as a gene flow barrier in terrestrial arthropods: a case study with the scorpion *Buthus occitanus* as model organism. *Molecular Phylogenetics and Evolution*, **28**, 119–130.

Gantenbein, B., Fet, V., Gantenbein-Ritter, I.A. & Balloux, F. (2005) Evidence for recombination in scorpion mitochondrial DNA (Scorpiones: Buthidae). *Proceedings of the Royal Society B: Biological Sciences*, **272**, 697–704.

Graham, A. (1999) *Late Cretaceous and Cenozoic history of North American vegetation*. Oxford University Press, New York.

Hamilton, C.A., Formanowicz, D.R. & Bond, J.E. (2011) Species delimitation and phylogeography of *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): cryptic diversity in North American tarantulas. *PLoS ONE*, **6**, e26207.

Heled, J. & Drummond, A. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Henry, C.D. & Aranda-Gomez, J. (2000) Plate interactions control middle–late Miocene, proto-Gulf and Basin and Range extension in the southern Basin and Range. *Tectonophysics*, **318**, 1–26.

Ji, Y.-J., Zhang, D.-X. & He, L.-J. (2003) Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Molecular Ecology Notes*, **3**, 581–585.

Katoh, K. & Toh, H. (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics*, **9**, 212.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Keith, R.M. & Hedin, M. (2012) Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae, Hypochilidae, *Hypochilus*), with implications for species delimitation. *Journal of Arachnology*, **40**, 167–181.

Mayr, E. (1963) *Animal species and evolution*. Belknap Press, Cambridge, MA.

McCormack, J.E., Bowen, B.S. & Smith, T.B. (2008) Integrating paleoecology and genetics of bird populations in two sky island archipelagos. *BMC Biology*, **6**, 28.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J.G., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G. & Vogler, A.P. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

Morafka, D.J. (1977) A biogeographical analysis of the Chihuahuan Desert through its herpetofauna. *Biogeographica*, **9**, 1–313.

Mulcahy, D.G. & Macey, J.R. (2009) Vicariance and dispersal form a ring distribution in nightsnakes around the Gulf of California. *Molecular Phylogenetics and Evolution*, **53**, 537–546.

Nylander, J.A.A. (2004) *MRMODELTEST v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala.

Oskin, M.E., Stock, J.M. & Martín-Barajas, A. (2001) Rapid localization of Pacific–North America plate motion in the Gulf of California. *Geology*, **29**, 459–462.

Polis, G.A., McReynolds, C.N. & Ford, R.G. (1985) Home range geometry of the desert scorpion *Paruroctonus mesaensis*. *Oecologia*, **67**, 273–277.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Prendini, L. (2001) Substratum specialization and speciation in southern African scorpions: the Effect Hypothesis revisited. *Scorpions 2001: in memoriam Gary A. Polis* (ed. by V. Fet and P.A. Selden), pp. 113–138. British Arachnological Society, Burnham Beeches, UK.

Prendini, L. (2005) Scorpion diversity and distribution in southern Africa: pattern and process. *African biodiversity: molecules, organisms, ecosystems* (ed. by B.A. Huber, B.J. Sinclair and K.-H. Lampe), pp. 25–68. Springer, New York.

Prendini, L., Crowe, T.M. & Wheeler, W.C. (2003) Systematics and biogeography of the family Scorpionidae Latreille, with a discussion of phylogenetic methods. *Invertebrate Systematics*, **17**, 185–259.

Prendini, L., Weygoldt, P. & Wheeler, W.C. (2005) Systematics of the *Damon variegatus* group of African whip spiders (Chelicerata: Amblypygi): evidence from behaviour, morphology and DNA. *Organisms, Diversity and Evolution*, **5**, 203–236.

R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Rambaut, A. & Drummond, A.J. (2007) *Tracer v1.5*. Available at: http://beast.bio.ed.ac.uk/Tracer/.

Retallack, G.J. (2001) Cenozoic expansion of grasslands and climatic cooling. *Journal of Geology*, **109**, 407–426.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Schoenherr, A.A., Feldmeth, C.R. & Emerson, M.J. (1999) *Natural history of the islands of California*. University of California Press, Berkeley, CA.

Shafiqullah, M., Damon, P.E., Lynch, D.T., Reynolds, S.J., Rehrig, W.A. & Raymond, R.H (1980) K-Ar geochronology and geologic history of southwestern Arizona and adjacent area. *Studies in western Arizona: Arizona Geological Society Digest*, Vol. XII (ed. by J. Jenney and C. Stone), pp. 201–260. Tucson, AZ.

Sissom, W.D. (2000) Family Vaejovidae. *Catalog of the scorpions of the world (1758–1998)* (by V. Fet, W.D. Sissom, G. Lowe and M.E. Braunwalder), pp. 503–553. New York Entomological Society, New York.

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.

Stockwell, S.A. (1989) *Revision of the phylogeny and higher classification of scorpions (Chelicerata)*. PhD Thesis. University of California, Berkeley, CA.

Swofford, D.L. (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10*. Sinauer Associates, Sunderland, MA.

Talavera, G. & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**, 564–577.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tully, T., D'Haese, C., Richard, M. & Ferrière, R. (2006) Two major evolutionary lineages revealed by molecular phylogeny in the parthenogenetic collembolan *Folsomia candida*. *Pedobiologia*, **50**, 95–104.

Udvardy, M.D.F. (1975) *A classification of the biogeographical provinces of the world*. International Union for Conservation of Nature and Natural Resources, Gland, Switzerland.

Umhoefer, P.J. (2011) Why did the Southern Gulf of California rupture so rapidly?—Oblique divergence across hot, weak lithosphere along a tectonically active margin. *GSA Today*, **21**, 4–10.

Vrba, E.S. (1980) Evolution, species and fossils: how does life evolve? *South African Journal of Science*, **76**, 61–84.

Webb, R.H. & Betancourt, J.L. (1990) The spatial and temporal distribution of radiocarbon ages from packrat middens. *Packrat middens: the last 40,000 years of biotic change* (ed. by J.L. Betancourt, T.R.V. Devender and P.S. Martin), pp. 85–103. University of Arizona Press, Tucson, AZ.

Wilson, J.S. & Pitts, J.P. (2010) Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: a resource for biologists. *Progress in Physical Geography*, **34**, 419–441.

Yu, Y., Harris, A. & He, X. (2011) RASP (reconstruct ancestral state in phylogenies). Available at: http://mnh.scu.edu.cn/soft/blog/RASP/.

Zhang, X. & Paulssen, H. (2012) Geodynamics of the Gulf of California from surface wave tomography. *Physics of the Earth and Planetary Interiors*, **192–193**, 59–67.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Collection locality data for genetic samples used in the phylogeographical study of North American vaejovid scorpions of the *Pseudouroctonus minimus* complex.

Appendix S2 Scorpion-specific mtDNA primers used to amplify DNA sequence data in the phylogeographical study of North American vaejovid scorpions of the *Pseudouroctonus minimus* complex.

Appendix S3 Each of the three gene trees (right) embedded within the shared species tree (left) for North American vaejovid scorpions of the *Pseudouroctonus minimus* complex, co-estimated with *BEAST.

BIOSKETCHES

Robert W. Bryson Jr is interested in the origin and distribution of genetic diversity in taxa distributed in south-western North America and the Neotropics.

Warren Savary and **Lorenzo Prendini** are interested in arachnid systematics, evolution and biogeography and are part of an international team of researchers studying the endemic North American scorpion family Vaejovidae.

Editor: Aristeidis Parmakelis