Testing A Vicariance Model To Explain Haplotype Distribution In The Psammophilic Scorpion Paruroctonus Utahensis

Authors
M. C. Estep, Mary U. Connell, Richard N. Henson, Zack E. Murrell, and Randall L. Small

Abstract
A model of vicariance speciation has been proposed to explain the proliferation of psammophilic (sand loving) species in several endemic genera of scorpions. The current distribution of Paruroctonus utahensis (Williams, 1968) populations on isolated sand dunes seems to fit this model; therefore, this species was examined to test the model for psammophilic population distribution and haplotype variation. A portion of the mitochondrial rDNA gene was sequenced from 44 individuals representing 7 populations of P. utahensis. Three individuals of Paruroctonus gracilior (Hoffmann, 1931) and 1 individual of Paruroctonus boquillas Sissom and Henson, 1998, were used for outgroup comparison. A maximum parsimony heuristic search indicated that the basal populations were in the south and derived populations occurred in the northeastern portion of the range. A nested clade analysis found a statistically significant correlation between haplotypes and their geographical distribution at several clade levels. The biological cause of this association was best explained by allopatric fragmentation. Our data supported the vicariance model of isolation by fragmentation of a larger habitat to explain the variation seen among populations of P. utahensis.
Psammophilic scorpion genera occur in 4 families and all exhibit morphologies that are modified for existence on sand. Many scorpion species in the North American endemic genus *Paruroctonus* Werner, 1934 (Vaejovidae) are psammophiles. These “sand loving” species are allopatrically distributed throughout the range of the genus, across western North America, from southern Canada to Aquascalientes, Mexico.

Scorpion ecologists describe psammophiles as ecomorphotypes with adaptations that facilitate burrowing and locomotion on sand, but impede movement on other substrates (Polis, 1990). Different scorpion lineages that have adapted to living on sand have independently
evolved similar morphological traits (Fet et al., 1998). These adaptations, including elongation of tarsal claws and enlargement of macrochaetae arranged in sand combs, generally increase surface area for locomotion and allow psammophilic scorpions to move freely across loose sandy soils without losing traction or sinking into the substrate. Although the increased surface area allows these scorpion specialists to easily move across the sand, the longer legs and extensive projections hinder movement on less fluid substrates (Polis, 1990). Therefore, these adaptations tend to restrict the movement of psammophilic scorpions between sandy habitats.

Fet et al. (1998) and Prendini (2001) proposed a model of vicariance speciation to explain the proliferation of psammophilic species within endemic genera of scorpions. Their model suggests that a single sandy habitat can be fragmented into multiple smaller isolated sandy habitats and that the ensuing loss of gene flow between the isolated habitats will promote a speciation event. Sandy substrates are known to migrate in the direction of prevailing wind systems during dry climatic periods, yet become stranded due to plant invasion in wet climatic periods, thus suspending further dune migration. According to Avise (1994), if a predictable dispersal mechanism can be reconstructed, then a phylogenetic analysis can be used to test the mechanism for divergence of populations and subsequent speciation.

Paruroctonus utahensis (Williams, 1968) is a psammophilic scorpion of the southwestern United States and small portions of Mexico (Chihuahua). The southernmost distribution of P. utahensis is in the Chihuahuan Desert, with the more northerly populations in desert grasslands of Arizona, New Mexico, Texas, and southern Utah. This scorpion is found in isolated populations on shifting sand dunes throughout its range and on loose sandy soils along the Rio Grande (Stockwell, 1986; R. Henson, pers. observ.). Because P. utahensis possesses typical psammophilic morphological adaptations and does not disperse easily across habitats that are not sand, we used this species to test the validity of the vicariance model.

The fragmentation of sand dunes in the Desert Southwest has occurred on a geologic time scale extending back to glaciated periods of the Pleistocene (Livingstone and Warren, 1996). The episodic events of the Pleistocene period have had an extreme effect on plant and animal communities of the Chihuahuan Desert (Van Devender, 1986). The wet cycles produced alluvium that was deposited along riverbanks and in pluvial lakes, and preexisting dune systems lost some of their ability to move due to the establishment of plant communities. The drier periods allowed sand dunes to grow in size and migrate across surrounding arid environments. Since the end of the Pleistocene, severe aridity can be first detected in the Chihuahuan Desert about 6,000 years ago, with a shift from temperate grasslands to desert grasslands (Betancourt et al., 1990). New dune systems in the more northerly region of desert grasslands were formed during the Alithermal from about 7,500 to 4,500 years ago (Holliday, 1989; Wilkins and Currey, 1999). Dunes in the Guadalupe Mountains are thought to have migrated in a northeasterly direction and dunes in the high plains of western Texas are estimated to have moved in a northwestward direction. Dune movement has generally been in a northerly direction across the Chihuahuan Desert (Holliday, 1989; Wilkins and Currey, 1999).

We hypothesized that dune location will help explain the distribution and genetic relatedness of P. utahensis populations. The vicariance model can be tested with molecular techniques to evaluate the degree of gene flow among populations. Restricted gene flow would indicate that populations are isolated on desert sand dunes with little gene exchange between populations. Larger amounts of gene flow would indicate many long distance dispersal events across the range of this species.

METHODS. We collected specimens of P. utahensis from 21 July to 3 August 2000 and 8 to 23 July 2001 from 7 localities that represent the total range of the species in Texas, New Mexico, Arizona, and Utah (Fig. 1). These sites were Monahans Sandhills State Park (M), Texas; Guadalupe Mountains National Park (G), Texas; White Sands National Monument (WS), New Mexico; El Paso (E), Texas; Channing (C), Texas; Bluff (B), Utah; and Willcox (W), Arizona. We also collected specimens of P. boquillas Sissons and Henson, 1998, the sister species of P. utahensis, and P. gracilior (Hoffmann, 1931), the type species for the genus (Fet et al., 2000).

Total DNA was isolated using the DNeasy tissue kit
maximum of Ohio). mosequenase IR700 quenced feature. The amplified GeneAmp fi
CA3 pion recovery (Qiagen, Inc., Valencia, California) and product recovery assessed by gel electrophoresis. A section of mitochondrial 16S rDNA was amplified using scorpion specific primers 5’CGATTGGACTCAGATCA3’ and 5’GTGCCAGGTAGCATAATCA3’ (Gantenbein et al., 1999). Each 25-µl PCR reaction consisted of 10 µl of isolated DNA, the scorpion specific primers, and a Ready-To-Go PCR bead (Amersham Pharmacia Biotech, Piscataway, New Jersey). Forty-fi amplifi cycles were performed in a GeneAmp 9600 Thermostyler (Perkin Elmer, Norwalk, Connecticut) using a 50°C annealing temperature. The amplified region was subsequently sequenced by cycle sequencing using LI-COR Gene Reader 4200 (LI-COR, Inc., Lincoln, Nebraska), IR700 and IR800 labeled primers, and a USB thermosequenase kit (USB Corporation, Cleveland, Ohio). The resulting sequences were edited and aligned for comparison using the AlignIR alignment software (LI-COR, Inc., Lincoln, Nebraska).

Sequence data from 44 individuals representing 7 populations of *P. utahensis*, along with 3 individuals of *P. gracilior* and one specimen of *P. boquillas*, were analyzed using PAUP*4.0* (Swofford, 2000) for a maximum parsimony heuristic search, followed by 100 replicated bootstraps. A 50% majority-rule consensus tree was constructed.

A nested cladistic analysis was performed (Templeton et al., 1995) to determine if sequence variation was spatially distributed. Nested clad analysis is a widely used analytical procedure that tests for significant associations between haplotypes and their geographic distances in the context of a gene genealogy and coalescent theory (Avise, 1998; Templeton, 2001; Knowles and Maddison, 2002; Knowles, 2004; Templeton, 2004; Wakley, 2004). Specifically, nested clad analysis statistically tests for significantly large or small geographic distributions of particular haplotypes or groups of haplotypes (nested clad) by comparing observed distributions to a null distribution obtained from random permutations of the original data. Using the program TCS (Clement et al., 2000), a gene genealogy based on sequence data was constructed and the haplotypes grouped into clades and nested clades based on the number of mutational steps that separate them. The program GeoDis version 2.0 (Posada et al., 2000) then uses these haplotype genealogies and nested clad along with geographic data to calculate the distance statistics used to identify specific geographical associations. Clade distance (Dc), which measures the geographic distribution of the haplotypes, nested clad distance (Dn), which assesses the geographic distribution of a given haplotype relative to the haplotypes nested with it, the interior-tip statistic (I-T), which estimates an average interior distance minus an average tip distance within each nested group, and the permutational P-values are produced by GeoDis. For those clades that showed a statistically significant association between haplotype and geographical distribution based on these statistics, the biological causes of such an association were interpreted using the inference key of Templeton et al. (1995) and Templeton (1998).

**RESULTS**—A maximum parsimony heuristic search of the aligned sequences from 44 individuals found a total of 112 equally parsimonious trees of 73 steps. Each tree had a consensus index of 0.87 and a retention index of 0.94. A 50% majority-rule consensus tree was constructed from the 112 equally parsimonious trees with bootstrap values indicated at each node in the tree topology (Fig. 2). The *P. utahensis* sequences were subdivided into a basal group of unresolved populations representing individuals from the Guadalupe Mountains National Park and Willcox populations and a clade containing individuals from all 7 localities that was supported with a bootstrap value of 71. Within this derived clade, 3 populations
The 50% majority-rule consensus tree constructed from 112 equally parsimonious trees obtained using a maximum parsimony heuristic search of aligned 16S rDNA sequences from Paruroctonus utahensis. Bootstrap values (≥50%) based on 100 resamples are indicated on the internal branches. Populations of P. utahensis are designated by location. Paruroctonus boquillas and Paruroctonus gracilior are the designated outgroups and abbreviated on the cladogram as BOQ and GRA, respectively.

(White Sands, Channing, Monahans) formed monophyletic subgroups, with the Channing and Monahans populations being sister groups. The White Sands population had a bootstrap value of 65. The Monahans and Channing populations, representing the most northeastern populations, formed the second monophyletic subgroup supported with a bootstrap value of 66. The Monahans and Channing populations were supported with bootstrap values of 79 and 68, respectively.

Twenty-three haplotypes were defined for the P. utahensis populations based on the sequence data (Table 1). Only one haplotype (Put8) occurred in multiple populations. The nested cladistic analysis of these haplotypes resulted in 4 distinct 2-step clades. Eighteen of the 23 haplotypes nested within 2 of these clades, 5 within clade 2.1 and 13 within clade 2.3 (Fig. 3). Clade 2.2 contained both the Channing and Monahan haplotypes as tip clades 1.6 and 1.1, respectively. These clades were connected to the interior Put4 haplotype from the Guadalupe Mountains National Park, which in the analysis is interpreted to be the ancestral haplotype within this nested clade. Three tip clades (1.10, 1.5, and 1.4) were nested within clade 2.3, along with the interior clade 1.7, which represents the ancestral haplotypes for this nested grouping. Nested clade 2.3 included haplotypes from the Guadalupe Mountains National Park, White Sands National Monument, Willcox, El Paso, and Bluff populations. Clade 2.2 included only haplotypes from the El Paso population, while clade 2.4 incorporated only Willcox haplotypes.

The nested clade permutational procedure identified a significant association between haplotypes and geographical location for clades 2.1, 2.4, and the entire cladogram (Table 2). Assessment of the distance statistics, using the inference key of Templeton et al. (1995) and Templeton (1998), indicated that the associations between genetic variation and geographic distance in these clades were best explained by allopatric fragmentation. The significantly small D values for clades 1.1, 1.5, 1.6, 1.7, and 2.2, the significantly small D values for clades 1.1 and 1.5, and the non-overlapping ranges for the individual populations all support this explanation. For clades 1.7, 1.10, and 2.4, which had non-significant Chi-square values in the permutational analysis, the inference key identified panmixia in sexual populations as the reason for their respective distance statistics.

**DISCUSSION** The use of both traditional and nested cladistic analyses of sequence variation in populations of *P. utahensis* revealed haplotype differences in the individual populations, established a statistically significant association between geographical location and this genetic variation, and identified the probable cause for this association as allopatric fragmentation of a once larger habitat. Populations of this scorpion would be expected to accumulate unique mutational differences if the dune systems on which they live have been isolated with little
Table 1—Variable positions in a fragment of 16S mitochondrial DNA from Paruroctonus utahensis defining 23 haplotypes and their distribution (B Bluff, Utah; C Channing, Texas; E El Paso, Texas; G Guadalupe Mountains National Park, Texas; M Monahan Sandhills State Park, Texas, WS White Sands National Monument, New Mexico; and W Willcox, Arizona).

<table>
<thead>
<tr>
<th>Population</th>
<th>Haplotype</th>
<th>Nucleotide position:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 1 8 9 9 1 1 4 5 5 5 5 5 6 7 7 8 9 9 0 0 4 4 8 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 2 8 6 7 8 2 3 6 1 4 6 9 1 0 5 0 1 4 6 8 3 7 6 4</td>
</tr>
</tbody>
</table>

Table 2—Nested contingency analysis of geographic structure of the Paruroctonus clades based on 1,000 resamples. Clades lacking genetic variation, geographical variation, or both are not included because no test is possible.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Permutational Chi-square statistic</th>
<th>Probability of random Chi-square being &gt; observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>8.0000</td>
<td>0.1080</td>
</tr>
<tr>
<td>1.10</td>
<td>0.6857</td>
<td>1.0000</td>
</tr>
<tr>
<td>2.1</td>
<td>23.0000</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>2.3</td>
<td>69.5312</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>2.4</td>
<td>1.3333</td>
<td>1.0000</td>
</tr>
<tr>
<td>Entire cladogram</td>
<td>86.7020</td>
<td>&lt;0.0001*</td>
</tr>
</tbody>
</table>

* Probability values <0.05 indicate significant geographical structure.

The possibility of dispersion from one locale to another. The isolated nature of these aeolian sand dunes is evident when their locations are viewed on a map of the Southwest (http://geochange.er.usgs.gov/sw/impacts/geology/sand/swsand.html; Muhs, 2004). In the cladistic analysis, the grouping of all individuals from a single population into a monophyletic clade, as seen with the Monahans, Channing, and White Sands populations, suggested that isolation contributed to the unique genetic structure and further suggested a closer historical relationship between the Monahans and Channing populations based on their position as sister clades. This was strongly supported in the nested clade analysis by the placement of Monahans and Channing populations as tip clades within clade 2.1 and by the distance statistics for the both these tip clades and the entire 2.1 clade.
While genetic structure in the remaining populations was evident, it was more difficult to interpret using a traditional cladistic approach because no strong monophyletic groups emerged. The nested clad analysis grouped representative haplotypes from all the populations, except Channing and Monahans, within the larger 2.3 clade with nested clades 1.7 and 1.10 containing haplotypes from more than one population. This suggested that either these populations had been isolated for a shorter period or there had been rare, long-distance dispersal events.

The nested clad analysis allowed us to assess the cause of the haplotype variation that we saw in the different populations of *P. utahensis*. The ability to evaluate the relative roles of historical processes, such as range fragmentation or expansion, versus current biological processes, like gene flow, in fashioning the spatial distribution of haplotypes was one of the strengths of this type of analysis. This contributed to the widespread use of nested clad analysis for population studies of organisms across the biological spectrum, including flowering plants (Ge et al., 2002), land snails (Pfenninger and Posada, 2002), freshwater shrimp (Hurwood and Hughes, 2001), insects (Gomez-Zurita et al., 2000; Althoff et al., 2002; Contreras-Diaz et al., 2003), freshwater mussels (Turner et al., 2000), crayfish (Fetzner and Crandall, 2003), brown trout (Bernatchez, 2001), amphibians (Mead et al., 2001; Hoffman and Blouin, 2004), birds (Sgariglia and...
Burns, 2003), rabbits (Branco et al., 2002), and humans (Hammer et al., 1998) to mention just a few.

The patterns of genetic variation we found between P. utahensis populations were concordant with the psammophilic nature of these scorpions and the geographic isolation of the dunes on which they live. The biological cause of the geographical associations of haplotypes was found to be allopatric fragmentation, supporting our initial hypothesis.

We thank V. Davila (Big Bend National Park, permit BIBE-2000-043 and BIBE-2001-SCI-0053), F. Armstrong (Guadalupe Mountains National Park, permits GMO-00-007 and GMO-2001-SCI-0066), B. Conrad (White Sands National Monument, White Sands permit), and D. Riskind (Texas Parks and Wildlife, permit 49-00) for their help obtaining collecting permits, R. H. Goetz for translating the abstract into Spanish, and S. Windelspecht (Ricochet Productions) for her assistance with the graphics. Funding provided by Appalachian State University.

LITERATURE CITED


Livingstone, I., and A. Warren. 1996. Aeolian geo-


Associate Editor was Steven Goldsmith.