2001. V. Fet & P. A. Selden (eds.). *Scorpions 2001. In Memoriam Gary A. Polis*

Mitochondrial DNA reveals a deep, divergent phylogeny in *Centruroides exilicauda* (Wood, 1863) (Scorpiones: Buthidae)

Benjamin Gantenbein[1], Victor Fet[2] and **Mark D. Barker[2]**

[1]Institute of Zoology, Division of Population Biology, University of Berne, Baltzerstrasse 3, CH–3012 Berne, Switzerland

[2]Department of Biological Sciences, Marshall University, Huntington, West Virginia 25755–2510, USA

Summary

A molecular (16S mtDNA sequence) data set recovers a clear divergent phylogeny within the scorpion species *Centruroides exilicauda* (Wood, 1863) (Scorpiones: Buthidae) from Mexico and the USA. Three deep, monophyletic clades are revealed: two distinct peninsular clades—southern Baja (SB) and northern Baja (NB)—and a mainland clade, Sonora/Arizona (SO/AZ). Phylogenetic, taxonomic, and biogeographic implications are discussed.

Introduction

Centruroides Marx, 1890, with 41 species and 24 subspecies (Fet & Lowe, 2000) is one of the most diverse genera of Buthidae (second only to *Tityus* C. L. Koch, 1836), and the only genus of this family found in North America. *Centruroides* is especially diverse in Mexico (Lourenço & Sissom, 2000) and the Caribbean (Armas, 1988), but is also found in Central and South America (Sissom & Lourenço, 1987). Moreover, scorpions of the genus *Centruroides* were introduced by man to the Canary Islands (Goyffon, 1992; Juan *et al.*, 2000). Taxonomy of many *Centruroides* species is confusing, and was traditionally based mainly on morphosculpture and coloration. No modern phylogenetic analysis exists for this genus, and the only existing key (Stahnke & Calos, 1977) is out of date.

Scorpions of the genus *Centruroides* (especially many central Mexican species) are among the most toxic in the world to humans and other mammals (Mazzotti & Bravo-Becherelle, 1963; Simard & Watt, 1990). They are abundant in various natural habitats ranging from tropical forest to temperate deserts. The only medically important scorpion species in the United States (where it is mainly found in Arizona) is *Centruroides exilicauda* (Wood, 1863) (= *C. sculpturatus* Ewing, 1928) (Stahnke, 1971; Curry *et al.*, 1984).

This paper reports new phylogenetic data on *C. exilicauda*, based on the analysis of mitochondrial DNA (16S mtDNA) sequences. These molecular markers have recently become a powerful tool for evaluating the taxonomic status of populations and subspecies/species. The first information on applicability of mtDNA analysis to the species-level taxonomy of *Centruroides* was reported by Fet & Poindexter (1992). Recently, our comparison of 16S rRNA mitochondrial gene sequences allowed clarification of phylogeny at the species level in the genus *Euscorpius* Thorell, 1876 (Euscorpiidae) (Fet *et al.*, 1999; Gantenbein *et al.*, 1999, 2000). In order to further clarify the taxonomic composition of *C. exilicauda* and to confirm its species status assumed by Williams (1980), we initiated a pilot molecular survey applying previously established, scorpion-specific mitochondrial (16S mtDNA) gene markers. Because the particular fragment of the 16S gene was applied to clarify the phylogeny in other scorpion genera as well it is important to test if the number of base pairs that are sequenced from this fragment is high enough for a good phylogenetic signal. Whether the data set is large enough for this particular

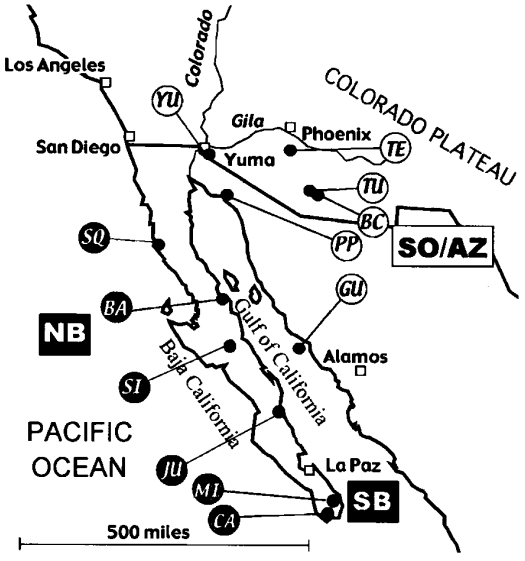

Fig. 1: Sampling sites of *Centruroides exilicauda* from Mexico and the USA. Baja California, Mexico: **SQ** San Quintin; **BA** Bahia de Los Angeles; **SI** San Ignacio; **JU** Juncalito; **CA** Cabo San Lucas; **MI** Miraflores, Sonora, Mexico: **GU** Guaymas Islands (Pajaro); **PP** Puerto Penasco, Arizona, USA: **YU** Yuma; **TE** Tempe; **BC** Black Canyon City, near Tuscon; **TU** Tuscon. **SB** = Southern Baja clade, **NB** = Northern Baja clade, and **SO/AZ** = Sonora desert/Arizona clade, as revealed from phylogenetic analysis.

phylogenetic study is addressed by parametric bootstrapping (Huelsenbeck & Hillis, 1996).

Material and methods

Specimens analysed

Twelve specimens of *Centruroides exilicauda* were analysed from 12 localities from Mexico and USA (Fig. 1): Baja California, Mexico: San Quintin (*Cex*SQ); Bahia de Los Angeles (*Cex*BA); San Ignacio (*Cex*SI); Juncalito (*Cex*JU); Cabo San Lucas (*Cex*CA); Miraflores (*Cex*MI). Sonora, Mexico: Guaymas Islands (Pajaro) (*Cex*GU); Puerto Penasco (*Cex*PP). Arizona, USA: Yuma (*Cex*YU); Tempe (*Cex*TE); Black Canyon City, near Tuscon (*Cex*BC); Tuscon (*Cex*TU).

Two congeneric outgroup species were used: *C. vittatus* (Say, 1821) from Arkansas, USA (*Cvi*AR), and *C. bani* Armas & Marcano Fondeur, 1987 from the Dominican Republic (*Cba*DO). Animals were kept alive or in 96% ethanol for DNA analysis.

DNA extraction and sequencing techniques

Comparative analysis of the mitochondrial (mt) 16S ribosomal RNA is widely used in modern molecular evolutionary studies of various arthropods (Simon *et al.*, 1994). The 16S rRNA gene has been recently used for resolving species-level phylogeny of the scorpion genus *Euscorpius* (Gantenbein *et al.*, 1999); this paper should be consulted for the detailed techniques and protocols. Total DNA was extracted from fresh or preserved (in 95% ethanol) muscle tissue using a standard extraction method. An approximately 400 bp fragment of the 16S rRNA gene was amplified by the polymerase chain reaction (PCR) using the universal forward primer LR-J-12887 (Simon *et al.*, 1994) and a scorpion-specific reverse primer (Gantenbein *et al.*, 1999). A total of 12 mtDNA sequences were used for the analysis and aligned by eye. For further analysis, all ambiguities and indels were excluded as suggested by Swofford *et al.* (1996), with final 354 characters remaining.

Phylogenetic analyses

We applied character-matrix-based methods: maximum parsimony (MP) and maximum likelihood (ML) (Felsenstein, 1981) as well as distance-based methods (neighbour-joining (NJ) cluster algorithm) (Saitou & Nei, 1987). The beta version of the computer program PAUP* 4.0b4a (Swofford, 1998) was used for all phylogenetic DNA analyses. We calculated hierarchical likelihood ratio tests in order to find the most appropriate model of DNA substitution using the program MODELTEST 3.0 beta (Posada & Crandall, 1998). This program calculates the likelihood ratio statistic $\delta = 2\log (L_1 - L_0)$, where L_1 is the likelihood under the alternative hypothesis (more complex, parameter rich, model) and L_0 is the likelihood under the null hypothesis (simple model). When the models compared are nested (the null hypothesis is a special case of the alternative hypothesis) and the null hypothesis is correct, the δ statistic is asymptotically distributed as χ^2 with q degrees of freedom, where q is the difference in number of free parameters between the two models. In the next step, a test for the molecular

clock hypothesis (i.e. rate constancy among lineages) was calculated (χ^2 distributed, $df = N - 2$ OTUs). Details on model testing using maximum likelihood ratios are given in Huelsenbeck & Rannala (1997) and in Huelsenbeck & Crandall (1997). The hierarchical likelihood ratio tests suggested the model of Hasegawa *et al.* (1985) with rate heterogeneity (HKY85 + Γ) which is a submodel of the general-time-reversible (GTR) substitution model (Rodríguez *et al.*, 1990; Yang *et al.*, 1994). All parameters (base frequencies, rate matrix) were estimated via maximum likelihood. The rate heterogeneity among sites was assumed to follow a gamma distribution (shape parameter α was ML-estimated) with four categories each represented by its mean (Yang 1994). Phylogenetic analysis is facilitated when rates are equal among lineages. The test for the molecular clock was not rejected at the 0.05 level ($P = 0.0502$). However, because the P value was critically low, we rejected the molecular clock hypothesis and performed tree searching without enforcing a clock.

For ML analyses the tree space was explored using the heuristic search option (100 replicates) with random addition of sequences using the tree bisection-reconnection (TBR) branch-swapping algorithm implemented in PAUP*. For the MP analysis the transitions (ti) were down-weighted relative to transversions (tv) according to the ML-estimated ti/tv ratio (9:1) using the HKY85 (Hasegawa *et al.*, 1985) model, and the tree search was done using the branch-and-bound search option. The consistency index (CI) and the retention index (RI) (Kitching *et al.*, 1998) were calculated as measures for tree stability with PAUP*. Parsimony-uninformative characters were excluded for the calculation of CI, thus we call it CIu (Kitching *et al.*, 1998).

Alternatively, pairwise ML-distances were estimated using the HKY85 + Γ model. Estimating distances via ML has the advantage of constant parameters over all pairwise comparisons and consequently the variance of distances is reduced. These distances were used as a matrix for neighbour-joining (NJ) clustering (Saitou & Nei, 1987). NJ is assumed to be a good heuristic approach for estimating the minimum evolution tree (Page & Holmes, 1998). All trees were rooted using two outgroup species: *Centruroides vittatus* and *C. bani*.

We applied nonparametric bootstrapping by resampling 1000 data sets, and parametric bootstrapping (= Monte Carlo simulation of DNA sequences) by resampling 100 data sets with 177, 354 and 708 characters, respectively. Parametric bootstrapping allows estimating how many characters are needed for resolving a particular phylogenetic tree at a given frequency (Huelsenbeck & Hillis, 1996). In contrast to the nonparametric approach, parametric bootstrapping uses a clearly defined substitution model (HKY85 + Γ in this case), a modeltree (the ML tree revealed by heuristic search) and phylogenetic parameters estimated via ML (Huelsenbeck & Hillis, 1996; Page & Holmes, 1998). Parametric bootstrapping was performed applying the algorithm "evolver" from the PAML 2.0 package (Yang, 1997, 2000).

Furthermore, we used the parsimony reconstructions to test the 16S gene for saturation. The evolutionary change (= pairwise genetic distances) should be at least equal to the minimum estimates of evolutionary change revealed by parsimony branch lengths (= pairwise patristic distances). We compared uncorrected and corrected distances with branch lengths inferred using parsimony as suggested by Philippe *et al.* (1994).

DNA sequence availability

All sequences were deposited in the EMBL Nucleotide Sequence Database with the following accession numbers: *C. exilicauda*: *Cex*SQ = AJ299428; *Cex*BA = AJ389384; *Cex*SI = AJ288634; *Cex*JU = AJ288635; *Cex*CA = AJ288636; *Cex*MI = AJ299429; *Cex*GU = AJ288642; *Cex*PP = AJ288641; *Cex*YU = AJ288640; *Cex*TE = AJ288639; *Cex*BC = AJ288637; *Cex*TU = AJ288638; *C. vittatus*: *Cvi*AR = AJ288643; *C. bani*: *Cba*DO = AJ288644.

Results

The calculated pairwise distances ranged from 1–10% divergence (uncorrected p) (Table 1). The genetic distances from the ingroup samples of *C. exilicauda* to the outgroup taxa (*Cvi*AR and *Cba*DO) were considerably higher, ranging from 13–16% (uncorrected p). These pairwise genetic distances generally underestimate true genetic divergence. The corrected HKY + Γ distance (as

	*Cex*SQ	*Cex*BA	*Cex*SI	*Cex*JU	*Cex*CA	*Cex*MI	*Cex*BC	*Cex*TU	*Cex*TE	*Cex*YU	*Cex*PP	*Cex*GU	*Cvi*AR	*Cba*DO
***Cex*SQ**	-	0.02	0.03	0.01	0.08	0.08	0.07	0.06	0.07	0.06	0.08	0.07	0.14	0.14
***Cex*BA**	0.03	-	0.04	0.04	0.10	0.10	0.09	0.08	0.09	0.08	0.10	0.09	0.14	0.15
***Cex*SI**	0.03	0.05	-	0.03	0.10	0.10	0.10	0.09	0.10	0.09	0.10	0.10	0.15	0.16
***Cex*JU**	0.02	0.05	0.03	-	0.07	0.07	0.07	0.07	0.08	0.07	0.08	0.08	0.14	0.15
***Cex*CA**	0.14	0.20	0.19	0.13	-	0.01	0.06	0.06	0.05	0.06	0.05	0.06	0.14	0.14
***Cex*MI**	0.14	0.20	0.19	0.13	0.01	-	0.06	0.07	0.06	0.06	0.06	0.06	0.14	0.14
***Cex*BC**	0.14	0.19	0.22	0.15	0.10	0.13	-	0.02	0.02	0.02	0.02	0.03	0.14	0.14
***Cex*TU**	0.11	0.16	0.18	0.14	0.10	0.13	0.02	-	0.02	0.02	0.03	0.03	0.13	0.14
***Cex*TE**	0.14	0.19	0.22	0.15	0.08	0.10	0.02	0.03	-	0.02	0.02	0.01	0.13	0.14
***Cex*YU**	0.12	0.16	0.19	0.13	0.11	0.11	0.02	0.02	0.02	-	0.02	0.02	0.13	0.14
***Cex*PP**	0.16	0.21	0.24	0.17	0.09	0.11	0.03	0.03	0.02	0.03	-	0.03	0.13	0.13
***Cex*GU**	0.14	0.19	0.21	0.15	0.11	0.11	0.03	0.04	0.01	0.02	0.04	-	0.14	0.14
***Cvi*AR**	0.52	0.52	0.56	0.50	0.56	0.52	0.60	0.55	0.55	0.50	0.55	0.56	-	0.15
***Cba*DO**	0.65	0.69	0.72	0.67	0.61	0.61	0.70	0.71	0.67	0.67	0.63	0.68	0.87	-

Table 1: Distance matrix of the sequence divergence (uncorrected p) (upper right) and of HKY85+G (lower left) calculated from pairwise comparisons of 16S mtDNA sequences in *Centruroides*.

suggested by hierarchical likelihood ratio tests) was calculated with the maximum likelihood (ML) estimated parameters (Table 1). The base frequencies were estimated to be $\pi_A = 0.35$, $\pi_C = 0.16, \pi_G = 0.11, \pi_T = 0.38$, and the ti/tv ratio was estimated to 9.00 ($\kappa = 22.58$), respectively. The shape parameter α was ML-estimated to 0.14 (assuming four rate categories). This corresponds to an L-shaped gamma distribution that indicates strong rate heterogeneity among sites. This correction for saturation increased the genetic distances, and therefore, genetic HKY + Γ distances ranged from 1–24% among *C. exilicauda* samples, and went up to 72% between *C. bani* and *C. exilicauda*. The proposed nucleotide substitution model and its estimated parameters are in accordance with the nucleotide substitution models proposed for sequences in other scorpion genera, namely *Hadrurus* (Fet *et al.* 2001), *Mesobuthus* and *Buthus* (Gantenbein, unpublished data). In *Euscorpius* (Gantenbein *et al.*, 2000), the Tamura-Nei (1993) model was selected that differs from the HKY85 model only by the distinction of two different transition ratios.

The NJ tree using HKY + Γ distances as an input matrix revealed three well supported main clusters within *C. exilicauda*, a southern Baja clade (SB), a northern Baja clade (NB), and a Sonora/Arizona clade (SO/AZ), respectively (Fig. 2A). The heuristic search using maximum likelihood (ML) revealed two tree islands with a -ln likelihood of 1146.08, whereas the two tree topologies differ slightly from each other by a dichotomous and a trichotomous branching for the three main clades within *C. exilicauda*. The dichotomous branching order was hit in 83 out of 100 search replicates. The two tree topologies from ML analysis are congruent to the topology revealed by NJ (Fig. 2A).

The Branch-and-Bound tree search using weighted maximum parsimony (MP) resulted in 44 equally parsimonious trees with 150 steps (consensus tree, see Fig. 2B). These trees differed mainly in the branching order of sequences with only one to two bp difference (e.g. *Cex*JU and *Cex*SQ). However, this method confirmed again three divergent clades within *C. exilicauda*. The consistency index (CIu) and the retention index (RI) were relatively high (0.63 and 0.75, respectively) and indicated a high tree stability.

The parametric bootstrap analysis evolving 374 bp along a given model-tree (tree in Fig. 2A) with the parameters estimated from ML analysis revealed high bootstrap values (ranging from 70 to 88% for all nodes, Fig. 2A). Evolving half of the analysed DNA fragment, i.e. 177 bp, using the same tree and parameters resulted in a general lowering of bootstrap values (55–66% for all nodes), and therefore, in a considerable reduction of the phylogenetic signal. Doubling the fragment length (708 bp) did increase bootstrap values (74–96% for all nodes). These findings imply that about 370 bp are enough characters for recovering a strong phylogenetic signal for detecting the three major splits in *Centruroides* given the sequences evolve according to the HKY + Γ model.

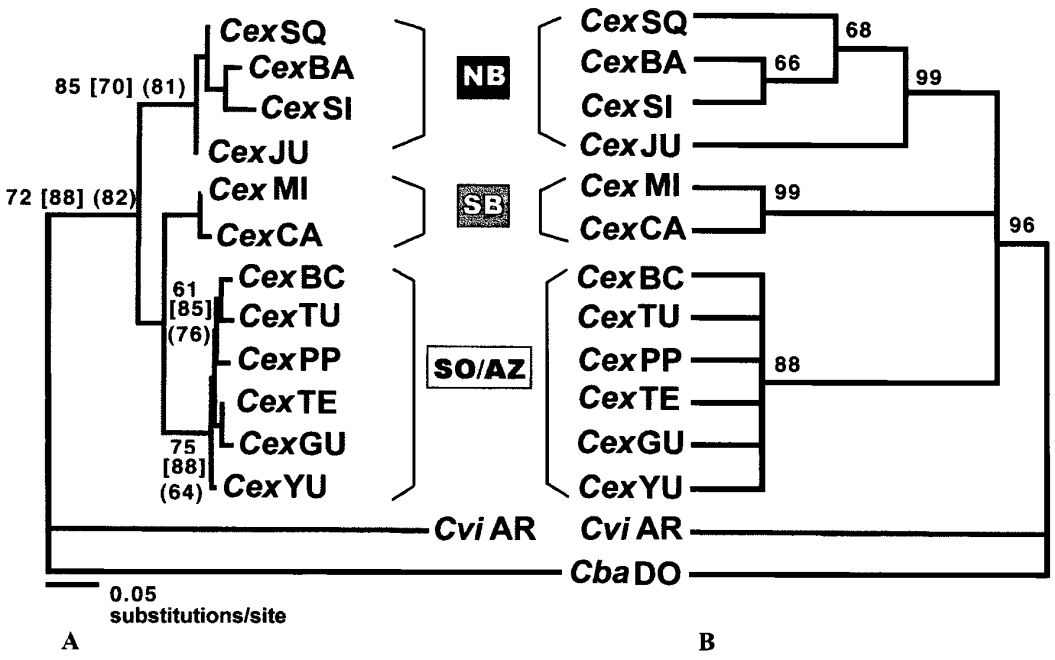

Fig. 2: **A** Neighbour-Joining (NJ) tree inferred from 16S mtDNA sequences of *Centruroides* using the HKY85 substitution model with rate heterogeneity (HKY85 + Γ). Bootstrap values (>50%) are from NJ-bootstrapping, in brackets ( ) from ML-bootstrapping and [ ] from parametric bootstrapping, respectively. **B** Strict consensus tree of 44 equally parsimonious trees of 150 steps revealed by weighed Maximum Parsimony analysis (MP). The consistency index (CIu) was 0.63 and the retention index (RI) was 0.75.

The plot for the saturation as suggested by Phillipe *et al.* (1994) is shown in Figure 3. When the number of inferred steps between taxa estimated on one of the equally parsimonious trees are compared to pairwise distances, there was no evidence for strong saturation among the sequences, using both uncorrected and corrected distances (Fig. 3). Correction for HKY + Γ distances resulted in a general reduction of distances among *C. exilicauda* samples, whereas the distances between the ingroup of *C. exilicauda* and the outgroups were increased (Fig. 3B).

These results show that a deep, divergent phylogeny exists among the studied populations of *C. exilicauda*, and two separate lineages can be distinguished: a "northern peninsular" lineage (NB) and a "mainland" lineage (SO/AZ). The position of Cabo San Lucas populations (*Cex*CA, *Cex*MI), the "southern peninsular" lineage (SB), is less clear.

Discussion

As currently accepted (Fet & Lowe, 2000), the species *Centruroides exilicauda* (Wood, 1863) includes populations from Mexico (Baja California, Sonora) and the United States (all of Arizona, adjacent parts of Utah and Nevada, and occasional records in California). The species is found also on the numerous islands in the Gulf of California (Williams, 1980; Due & Polis, 1986).

Traditionally, however, this taxon was for a long time treated as two separate species: a "peninsular" *C. exilicauda* (Wood, 1863), limited to Baja California only (type locality: Cabo San Lucas; Cockendolpher & Peek, 1991), and a "mainland" *C. sculpturatus* Ewing, 1928 from Mexico and the USA (type locality: Tempe, Arizona). Because Arizona populations attracted attention of toxicologists, the name *C. sculpturatus* was widely used in medical literature (e.g. Curry *et al.*, 1984; Simard & Watt, 1990). Those two species appeared to be allopatric (Williams,

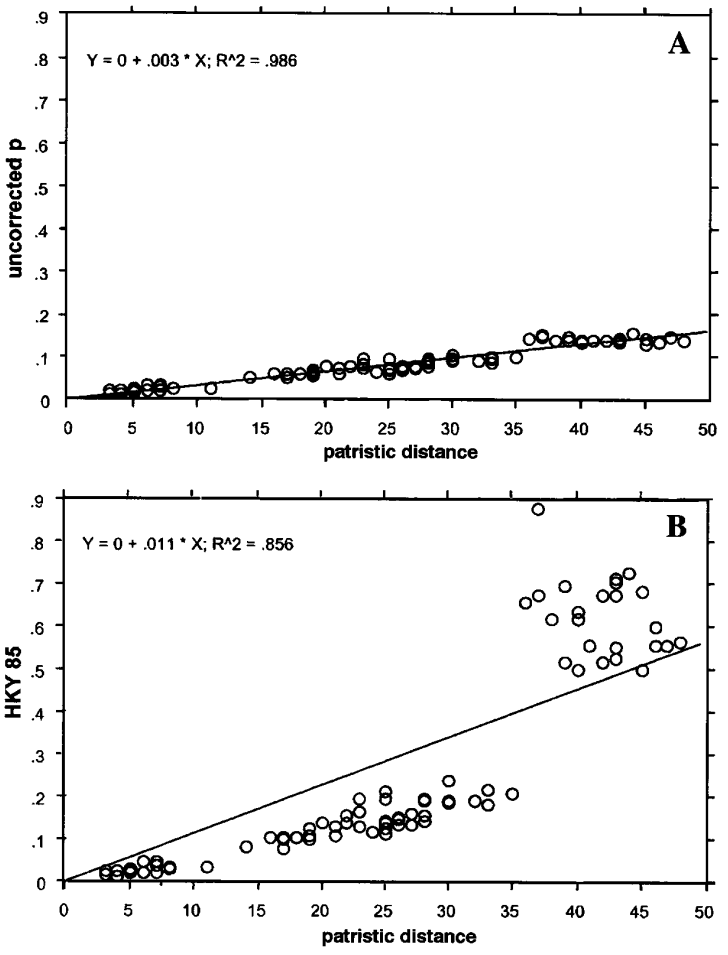

Fig. 3: Correlation between the number of substitutions between taxa from one arbitrarily chosen tree out of 44 equally parsimonious trees (= patristic distances) (X axis) and the number of observed differences from pairwise genetic distances (Y axis) for 16S rDNA data. Y axis in **A** uncorrected p and in **B** HKY + Γ distances. The saturation level is important in spite of the fact that the number of analysed sequences is low. The plotted distances should be most ideally distributed on the estimated regression line (saturation test after Phillipe *et al.*, 1994).

1980; J. Bigelow, pers. comm.). The indirect and anecdotal evidence exists that, while mainland populations were highly toxic, the Baja California populations do not exhibit high venom potency (Mazzotti & Bravo-Becherelle, 1963; G. A. Polis, pers. comm.), although no study attempted comparison of toxicity between peninsular and mainland populations.

C. exilicauda and *C. sculpturatus* were listed as separate taxa in the latest key to the genus (Stahnke & Calos, 1977). Later, Williams (1980) synonymized Ewing's mainland species *C. sculpturatus* with *C. exilicauda*. However, no detailed study of these two species has been published by this author; he simply stated (Williams, 1980: 7) that "comparison of *C. exilicauda* from Baja California with *C. sculpturatus* from Tempe, Arizona, reveal insignificant differences in morphology. Minor differences do appear when widely separated populations are compared; however, these seem to reflect only local racial adaptations." Wide variability of morphology in the Baja California populations of this species was addressed in an unpublished study by Walker (1973). A very high variability was also described from the Arizona populations (Stahnke, 1971). Synonymy of Williams (1980) has been accepted by the later authors (Fet & Lowe, 2000), although there was no attempt to study the situation in detail.

Our mtDNA data indicate that there might be in fact more than one species currently united under *Centruroides exilicauda*. Similar situations have been discovered elsewhere in scorpions, e.g. in the Alpine species *Euscorpius germanus* (C. L. Koch, 1837) (Euscorpiidae) (Gantenbein *et al.*,

position				11111	1111111111
position	11112	2333333344	6667777888	9999933355	5566666666
position	67891	9012456923	2890147134	0346901935	6701256789
Indel	=====	==========	==========	=========D	=======D==
Informativ	FFFFF	=F=FFF==FF	==FFF=F=FF	=FFF=FF=FF	====FFF=F=
ti	NN=N=	N=NNNN=NNN	NNNNNNNNNN	NNNNNNNNN=	NNNNNNN==N
tv	==VVV	=V===VV=V=	==========	=====VVV==	=VV=====VV
CexSQ	CCATC	GGAAGGAATT	GATACGGTAG	AGTTTGCAA*	TAGTTCC*TT
CexBA	T----	-----A----	----------	G----A--G-	---------C
CexSI	T----	----------	-------C--	--------G-	-G--------
CexJU	-----	----------	----------	------T---	------T---
CexCA	--CC-	---G----C-	----T-A-GA	-A---TT--A	----CTT---
CexMI	---C-	---G----C-	----T-A-GA	-A---TT---	----CTT---
CexBC	--C--	------C--C	---GT-A---	-AC--CT--A	----CTT-A-
CexTU	--C--	---------C	-G-GT-A---	-ACC-C----	----CTT-A-
CexTE	--CC-	---G-----C	---GT-A---	-ACC-CT---	----CTT-A-
CexYU	-----	---------C	----T-A---	-ACC-CT--A	----CTT-A-
CexPP	--C--	--------CC	---GT-A---	-ACC-TT---	----CTT-A-
CexGU	---C-	--GG-----C	---GT-A---	-ACC-CT--A	----CTT-A-
CviAR	TT--A	-T--AC-G--	A-CGTAA--A	-AC-CATT-A	--A--TT---
CbaDO	TT-AA	AT--AA--AC	--C-T-A--A	-A---TAG-A	CTCC-TTA-A

position	1111111111	1222222222	2222222223	3333333333	3333333333	333
position	7777788899	9011122223	3477778890	0001111111	2223333344	466
position	1235778934	6678935693	5623457970	3671235789	2890245923	506
Indel	======DD==	==========	==========	==========	==========	===
Informativ	=FFFF=F=FF	===FF=F===	F==F====F=	F=====F=FF	===FFFFF==	F=F
ti	=NNNN===NN	NNNNNNNNN=	NNNNNNNNNN	NNNNN=NNNN	NN=NNNNNNN	N=N
tv	VVV==V====	======V==V	===V======	VV=VVVV===	==V=======	=V=
CexSQ	TACCAT*TTC	ATTTTTATAA	TCAAACTTAA	GCAGAAAGTT	GAACCTCTTC	CAC
CexBA	-G--------	----------	----------	A---------	----------	---
CexSI	----------	---CC-----	--G----CG-	----------	----------	---
CexJU	----------	---C--G---	--------G-	----------	----------	---
CexCA	--TTG-A-C-	----CCG---	----------	------G-CC	-----CT---	T--
CexMI	--TTGAA-C-	----C-G---	C---------	------G-CC	-----CT---	T--
CexBC	--TT---*CT	----------	----------	T-------CC	----TCT---	--T
CexTU	--TT----C-	----------	----------	T-------C-	----TCT---	---
CexTE	--TT----CT	----------	----------	T-------C	----TCT---	T--
CexYU	-TTT----CT	----------	----------	T-------CC	----TCT---	---
CexPP	--TT----CT	----------	---G------	T-------CC	----TCTC--	T--
CexGU	--TT----CT	----------	----------	T--------C	----TCT--T	---
CviAR	ATAT------	GCC-C----T	CT-G--C-GG	ATGTT-T---	---TTCT-C-	T--
CbaDO	--AT----CT	----C-TCG-	---TGT----	-A-AGT-A--	AGCTTC-C--	TCT

Table 2: Polymorphic sites (indels, parsimony informative sites, transitions (ti), and transversions (tv)) in the analysed 16S rRNA gene sequences. Abbreviations for the haplotypes are explained in **Material and methods**. The sequence alignment was 373 bp long.

2000). In *C. exilicauda*, at least two lineages distinguished by mtDNA analysis are confined to Baja California peninsula and to mainland, respectively. The haplotype of the population from the type locality of *C. sculpturatus* Ewing, 1928 (Tempe, Arizona) (*Cex*TE) groups together with other mainland haplotypes, although ML and MP techniques do not agree on its exact placement. The haplotype of the population from the type locality of *C. exilicauda* (Wood, 1863) (Cabo San Lucas, Baja California) (*Cex*CA) in our analysis is separated from the mainland lineage. If the monophyly of these lineages is further confirmed, and a taxonomic level of species is eventually assigned to them, it might be necessary to consider a separate taxon for a still another,

northern peninsular lineage (*Cex*SQ, *Cex*BA, *Cex*SI, and *Cex*JU) which forms a monophyletic group in both ML and MP analyses.

Centruroides is one of a few extant scorpion genera of which Cenozoic fossils exist, from both the Miocene amber of the Dominican Republic (Schawaller, 1979), with the minimal age 15–20 Ma (Itturalde-Vinent & MacPhee, 1996), and the Miocene/Oligocene amber of Chiapas, Mexico (Santiago-Blay & Poinar, 1993). Thus, the age of the extant lineages can be very old, and their evolution could be considered against the geological events, which occurred since the Miocene. The Gulf of California formed only 12–5 Ma ago, and the Baja California Peninsula assumed its modern position in the Pliocene (3–5 Ma ago) (Gastil *et al.*, 1983). The general aspects of the evolution of North American desert fauna in general (Morafka *et al.*, 1992), and Baja California in particular (Murphy, 1983; Grismer, 1994) indicate southern origin for many taxa which later spread northward with progressing aridization.

The information about the assumed time of divergence (3–5 Ma ago) between the Baja Peninsula and the mainland allows to correlate genetic divergences against time. Assuming that the populations of *C. exilicauda* on the Baja peninsula and on the mainland are authochtonous, and that these populations were separated by the formation of the Baja Peninsula (3–5 Ma ago) we can estimate the evolutionary rate, which is the sequence divergence per time unit (Li, 1997). Phylogenetic reconstruction suggests an almost simultaneous splitting of *Centruroides* in three distinct clades (SB, NB, and SO/AZ). Dividing the observed genetic divergences between peninsular (SB and NB) and mainland (AZ/SO) haplotypes, which range from 6–10% ($\bar{x} = 7 \pm 1.5\%$) sequence divergence (uncorrected p), (Table 1), by an assumed estimated time of separation of 4 Ma reveals a molecular rate of 1.75% uncorrected divergence per Ma. Similarly, corrected HKY + Γ distances ranged from 8–24% ($\bar{x} = 15 \pm 4\%$) and the divergence rate is estimated to 3.75% HKY + Γ distance per Ma based on the identical separation time. If we consider an even earlier or later isolation time of mainland and peninsular lineages of 3 or 5 Ma the rates vary from 1.4–2.3% uncorrected sequence divergence and from 3–5% HKY + Γ distance per Ma. These estimates of molecular evolutionary rates in mitochondria lie in the range of the clock calibration in primates by Brown *et al.* (1979) based on mtDNA RFLP data. This study predicts an evolutionary rate of mitochondrial DNA of about 2% uncorrected divergence per Ma and was cited many times for organisms where no fossil data or dated vicariant events are available. More important for invertebrates, however, is the estimated divergence rate for the mitochondrial 16S gene in Hawaiian *Drosophila* species that was again estimated to about 2% uncorrected divergence per Ma (DeSalle *et al.*, 1987). Another clock calibration used pairwise genetic divergences of sister taxa of snapping shrimps of the genus *Alpheus* across the Isthmus of Panama and revealed a rate of 2.4% uncorrected sequence divergence (Knowlton *et al.*, 1993). These rates imply that the divergence rate in the scorpion mitochondrial 16S rRNA gene is comparable to other rates estimated in other invertebrates and even in vertebrates. Whether mitochondria in scorpions evolve generally with a comparable rate found in other invertebrates should be tested in other scorpion species with well known divergence dates. However, these clock calibrations are dependent on the analysed gene region and on the precision of the dating of the vicariant event(s) (Kumar & Hedges, 1998). Most calibrations are based on somewhat tenuous biogeographical dates and assume *ad hoc* a constant evolutionary rate, i.e. a molecular clock, and therefore, should be considered with caution.

The shape and clade support of the recovered mtDNA phylogeny unequivocally indicates that the mainland populations (Arizona + Sonora) are separated from the peninsular ones (Baja California). There was no resolution among the mainland populations that could indicate that these haplotypes are sufficiently divergent to provide the necessary phylogenetic signal in the 16S rDNA gene. On the other hand, the clade of the northern peninsular populations is well supported, and their phylogeny has a comblike branching shape indicating a clear geographic south-to-north gradient in haplotype changes, resembling a so-called "peninsular effect" in faunal changes observed in Baja California (Due & Polis, 1986). Thus, our phylogeny suggests that both peninsular and mainland populations originate from the southern ancestors. The Cape region of Baja California, according to the geological data (Gastil *et al.*, 1983), underwent a history of separation from the mainland as an archipelago of Miocene islands, and only later joined the territory of the modern Baja California Peninsula. Isolation of the Cape area at different times resulted in endemic forms of reptiles (Murphy, 1983; Grismer, 1994) and rodents (Riddle *et al.*, 2000); it also could be responsible

for a separation of southern (*Cex*CA, *Cex*MI) lineage of *C. exilicauda* which is observed in our phylogeny.

Our results suggest that more than one monophyletic lineage (species?) exist within currently accepted *Centruroides exilicauda* (Wood, 1863). Further, detailed investigations should be carried out to test this hypothesis: one has to study many more populations from the entire range of *C. exilicauda*, and search for other data sets (nuclear genes, allozymes, morphology, toxin structure/activity) to be able to establish the true taxonomic and genetic structure of the populations currently grouped under this species.

Acknowledgements

First and foremost, we are grateful to Gary A. Polis: his friendship, advice, and tutelage allowed V.F. and his family (as well as scores of other lucky people) to be a part of an exciting Baja research team for many, never to be forgotten, years. V.F.'s travel to Baja in 1993–1994 was supported by a NSF Research Opportunity Award as a supplement to DEB-9207855 grant to G. A. Polis. We thank everybody who donated specimens for this study: Joe Bigelow, Dietmar Huber, Gary Polis, Michael Soleglad, Barry Tomberlin, and Tsunemi Yamashita. We also thank Brian J. Poindexter whose help and enthusiasm were instrumental for the first understanding of *Centruroides* mtDNA.

References

ARMAS, L. F. De 1988: *Sinopsis de los escorpiones antillanos*. La Habana: Editorial Científico-Técnica.

BROWN, W. M., GEORGE, M. Jr & WILSON, A. C. 1979: Rapid evolution of animal mitochondrial DNA. *Proc. nat. Acad. Sci. U.S.A.* **44:** 1967–1971.

COCKENDOLPHER, J. C. & PEEK, M. L. 1991: Horatio C Wood (1841–1920) and his contributions to arachnology. *Rev. Arachnol.* **9:** 89–101.

CURRY, S., VANCE, M., RYAN, P., KUNKEL, D. & NORTHEY, W. 1983(1984): Envenomation by the scorpion *Centruroides sculpturatus*. *J. Toxicol. clin. Toxicol.* **21:** 417–449.

DESALLE, R., FREEDMAN, T., PRAGER, E. M. & WILSON, A. C. 1987: Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. molec. Evol.* **26:** 157–164.

DUE, A. D. & POLIS, G. A. 1986: Trends in scorpion diversity along the Baja California peninsula. *Am. Nat.* **128:** 460–468.

FELSENSTEIN, J. 1981: Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. molec. Evol.* **17:** 368–376.

FET, V., BARKER, M. & GANTENBEIN, B. 1999: Species-level variation of the mitochondrial 16S rRNA gene sequence: use in molecular systematics and biogeography. *Proc. W. Va Acad. Sci.* **71:**15.

FET, V. & LOWE, G. 2000: Family Buthidae C. L. Koch, 1837. *In* V. Fet, W. D. Sissom, G. Lowe & M. E. Braunwalder. *Catalog of the scorpions of the world (1758–1998)*. New York, NY: New York Entomological Society: 54–286.

FET, V. & POINDEXTER, B. J. 1992: Genetic variation of mitochondrial DNA in *Centruroides exilicauda* (Scorpiones: Buthidae) as revealed by the polymerase chain reaction. *Am. Zool.* **32:** 136A.

FET, V., SOLEGLAD, M. E. & BARKER, M. D. 2001: Phylogeny of the "hirsutus" group of the genus *Hadrurus* Thorell, 1876 based on morphology and mitochondrial DNA (Scorpiones: Iuridae). *In* V. Fet & P. A. Selden (eds.). *Scorpions 2001: In Memoriam Gary A. Polis*. Burnham Beeches, Bucks.: British Arachnological Society: 139–160.

GANTENBEIN, B., FET, V., BARKER, M. & SCHOLL, A. 2000: Nuclear and mitochondrial markers reveal the existence of two parapatric scorpion species in the Alps: *Euscorpius germanus* (C. L. Koch, 1837) and *E. alpha* Caporiacco, 1950, stat. nov. (Euscorpiidae). *Revue suisse Zool.* **107:** 843–869.

GANTENBEIN, B., FET, V., LARGIADÈR, C. R. & SCHOLL, A. 1999: First DNA phylogeny of *Euscorpius* Thorell, 1876 (Scorpiones: Euscorpiidae) and its bearing on taxonomy and biogeography of this genus. *Biogeographica* **75:** 49–65.

GASTIL, G., MINCH, J. & PHILLIPS, R. P. 1983: The geology and ages of islands. *In* T. J. Case & M. L. Cody (eds.). *Island biogeography in the Sea of Cortez*. San Diego, CA: University of California Press: 13–25.

GRISMER, L. L. 1994: The origin and evolution of the Peninsular herpetofauna of Baja California, Mexico. *Herpetol. nat. Hist.* **2:** 51–106.

GOYFFON, M. 1992: Le rôle de l'homme dans l'expansion territorial de quelques espèces de scorpions. *Bull. Soc. zool. Fr.* **177:** 15–19.

HASEGAWA, M., KISHINO, K. & YANO, T. 1985: Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *J. molec. Evol.* **22:** 160–174.

HUELSENBECK, J. P. & CRANDALL, K. A. 1997: Phylogeny estimation and hypothesis testing using maximum likelihood. *A. Rev. Ecol. Syst.* **28:** 437–466.

HUELSENBECK, J. P. & HILLIS, D. M. 1996: Parametric bootstrapping in molecular phylogenetics: Applications and performance. *In* J. D. Ferraris & S. R. Palumbi (eds.). *Molecular zoology: advances, strategies, and protocols*. New York, NY: Wiley-Liss, Inc.: 19–45.

HUELSENBECK, J. P. & RANNALA, B. 1997: Phylogenetic methods come of age: testing hypothesis in an evolutionary context. *Science, N. Y.* **276:** 227–232.

ITTURALDE-VINENT, M. A. & MACPHEE, R. D. E. 1996: Age and paleogeographical origin of Dominican amber. *Science, N. Y.* **273:** 1850–1852.

JUAN, C., EMERSON, B. C., OROMI, P. & HEWITT, G. M. 2000: Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends Ecol. Evol.* **15:** 104–109.

KITCHING, I. J., FOREY, P. L., HUMPHRIES, C. J. & WILLIAMS, D. M. 1998: *Cladistics*. Second edition. Oxford, UK: Oxford University Press.

KNOWLTON, N., WEIGT, L. A., SOLORZANO, L. A., MILLS, D. K. & BERMINGHAM, E. 1993: Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science, N.Y.* **260:** 1629–1631.

KUMAR, S. & HEDGES, B. 1998: A molecular timescale for vertebrate evolution. *Nature, Lond.* **392:** 917–920.

LI, W.-H. 1997: *Molecular evolution*. Sunderland, MA: Sinauer Associates.

LOURENÇO, W. R. & SISSOM, W. D. 2000: Scorpiones. *In* J. Llorente Bousquets, E. González Soriano & N. Papavero (eds.). *Biodiversidad, taxonomía y biogeographía de artrópodos de México*. México City: Universidad Nacional Autónoma de México **2:** 115–135.

MAZZOTTI, L. & BRAVO-BECHERELLE, M. A. 1963: Scorpionism in the Mexican Republic. *In* H. L. Keegan & W. V. MacFarlane (eds.). *Venomous and poisonous animals and noxious plants of the Pacific region*. New York, NY: Pergamon Press: 119–131.

MORAFKA, D. J., ADEST, G. A., REYES, L. M., AGUIRRE L., G. & LIEBERMAN, S. S. 1992: Differentiation of North American deserts: a phylogenetic evaluation of a vicariance model. *In* S. P. Darwin & A. L. Welden (eds.). *Biogeography of Mesoamerica. Tulane Studies in Zoology and Botany, Supplementary Publication 1*. New Orleans, LA: Tulane University: 195–206.

MURPHY, R. W. 1983: The reptiles: origin and evolution. *In* T. J. Case & M. L. Cody (eds.). *Island biogeography in the Sea of Cortez*. San Diego, CA: University of California Press: 130–158.

PAGE, R. D. M. & HOLMES, E. C. 1998: *Molecular evolution. A phylogenetic approach*. London: Blackwell Science.

PHILIPPE, H., SÖRHANNUS, U., BAROIN, A., PERASSO, R., GASSE, F. & ADOUTTE, A. 1994: Comparison of molecular and palaeontological data in diatoms suggests a major gap in the fossil record. *J. evol. Biol.* **7:** 247–264.

POSADA, D. & CRANDALL, K. A. 1998: Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

RIDDLE, B. R., HAFNER, D. J. & ALEXANDER, L. F. 2000: Phylogeography and systematics of the *Peromyscus eremicus* species group and the historical biogeography of North American warm regional deserts. *Molec. Phylog. Evol.* **17:** 145–160.

RODRÍGUEZ, F., OLIVER, J. L., MARÍN, A. & MEDINA, J. R. 1990: The general stochastic model of nucleotide substitution. *J. theor. Biol.* **142:** 485-501.

SAITOU, N. & NEI, M. 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **4:** 406–425.

SANTIAGO-BLAY, J. A. & POINAR, G. O. JR 1993: First scorpion (Buthidae: *Centruroides*) from Mexican amber (Lower Miocene to Upper Oligocene). *J. Arachnol.* **21:** 147–151.

SCHAWALLER, W. 1979: Erst Nachweis eines Skorpions im dominikanischen Bernstein. *Stuttg. Beitr. Naturk.,Ser. B* **45:** 1–15.

SIMARD, M. & WATT, D. D. 1990: Venoms and toxins. *In* G. A. Polis (ed.). *The biology of scorpions*. Stanford, CA: Stanford University Press: 414–444.

SIMON, C., FRATI, F., BECKENBACH, A., CRESPI, B., LIU, H., & FLOOK, P. 1994: Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. ent. Soc. Am.* **87:** 651–701.

SISSOM, W. D. & LOURENÇO, W. R. 1987: The genus *Centruroides* in South America (Scorpiones, Buthidae). *J. Arachnol.* **15:** 11–28.

STAHNKE, H. 1971: Some observations of the genus *Centruroides* Marx (Buthidae, Scorpionida) and *C. sculpturatus* Ewing. *Ent. News* **82:** 281–307.

STAHNKE, H. & CALOS, M. 1977: A key to the species of the genus *Centruroides* Marx (Buthidae, Scorpionida). *Ent. News* **88:** 111–120.

SWOFFORD, D. L. 1998: *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4*. Sunderland, MA: Sinauer Associates.

SWOFFORD, D. L., OLSEN, G. J., WADDELL, P. J. & HILLIS, D. M. 1996: Phylogenetic inference. Chapter 11. *In* D. M. Hillis, C. Moritz & B. K. Mable (eds.). *Molecular systematics*. Second edition. Sunderland, MA: Sinauer Associates: 407–425.

TAMURA, K. & NEI, M. 1993: Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molec. Biol. Evol.* **10:** 512–526.

WALKER, J. M. 1973: *A study of variability in the scorpion* Centruroides exilicauda *(Wood) (Scorpionida: Buthidae)*. Unpublished M.S. thesis, San Francisco State University, San Francisco, CA.

WILLIAMS, S. C. 1980: Scorpions of Baja California, Mexico, and adjacent islands. *Occ. Pap. Calif. Acad. Sci.* **135:** 1–127.

YANG, Z. 1994: Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. molec. Evol.* **39:** 306–314.

YANG, Z. 1997: PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13:** 555–556.

YANG, Z. 2000: *Phylogenetic Analysis by Maximum Likelihood (PAML), Version 2.0*. London: University College.

YANG, Z., GOLDMAN, N. & FRIDAY, A. 1994: Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molec. Biol. Evol.* **11:** 316–324.