Phylogeny of the "hirsutus" group of the genus *Hadrurus* Thorell, 1876 based on morphology and mitochondrial DNA (Scorpiones: Iuridae)

Victor Fet[1], **Michael E. Soleglad[2]** and **Mark D. Barker[1]**

[1]Department of Biological Sciences, Marshall University, Huntington, West Virginia 25755, USA

[2]P.O. Box 250, Borrego Springs, California 92004, USA

Summary

Phylogenetic analysis using both morphology and DNA sequences is presented for the "hirsutus" group of the scorpion genus *Hadrurus*. Parsimony-based cladistic analysis is conducted using important morphological diagnostic characters. Several analytical techniques are applied against the 16S mtDNA sequences, including genetic distance estimation, cluster analysis, maximum likelihood, maximum parsimony and direct optimization. The resulting cladogram for this group is ((*H. pinteri*, (*H. concolor, H. hirsutus*)), ((*H. spadix, H. obscurus*), *H. arizonensis*)), demonstrating monophyly of the "hirsutus" and "arizonensis" subgroups. Based on molecular evidence, the subspecies *H. arizonensis pallidus* Williams, 1970, is synonymized with the nominotypical form.

Introduction

Williams (1970) and Soleglad (1976) presented taxonomies of the scorpion genus *Hadrurus* Thorell, 1876, in which each proposed system was based on a different set of morphological characters. Although both authors discussed phylogenetic groupings, Williams using a combination of subtle character differentiations and biogeographical affinities, and Soleglad using the presence or absence of accessory trichobothria on the chelae, neither work presented rationale for these hypotheses using formal cladistic techniques. We recently obtained ample material for DNA extraction of all six species comprising the "hirsutus" group of *Hadrurus*, and therefore took the opportunity to conduct a full cladistic analysis of this group, encompassing both morphology and molecular sequence analysis. For morphology all important structural diagnostic characters as discussed and identified by Williams and Soleglad were coded for cladistic analysis. Many phylogenetic analyses were initiated against the 16S rRNA mitochondrial DNA sequences: genetic distance measurements, cluster analysis, maximum parsimony, maximum likelihood, and "direct optimization", an analysis approach that allows the detection of evolutionary events without the use of multiple sequence alignments (Wheeler, 1996).

The primary goal of this study is to determine the phylogenetic support for the two subgroups defined by Soleglad (1976) within the "hirsutus" group: the "hirsutus" and "arizonensis" subgroups. A secondary goal is to evaluate the monophyly of species complexes within the two subgroups. Within the context of this study, four species complexes are considered using groups and subgroups as defined by Soleglad (1976):

"arizonensis" subgroup

1. "arizonensis" complex: *H. a. arizonensis* Ewing and *H. a. pallidus* Williams, 1970 (*H. a. austrinus* Williams, 1970, not considered)
2. "spadix" complex: *H. spadix* Stahnke, 1940, and *H. obscurus* Williams, 1970

"hirsutus" subgroup:

3. "hirsutus" complex: *H. hirsutus* (Wood, 1863) and *H. concolor* Stahnke, 1969 (Note: this species was described and always listed as *H. concolorous*; however, the correct Latin adjective is "concolor", while "concolorous" should be
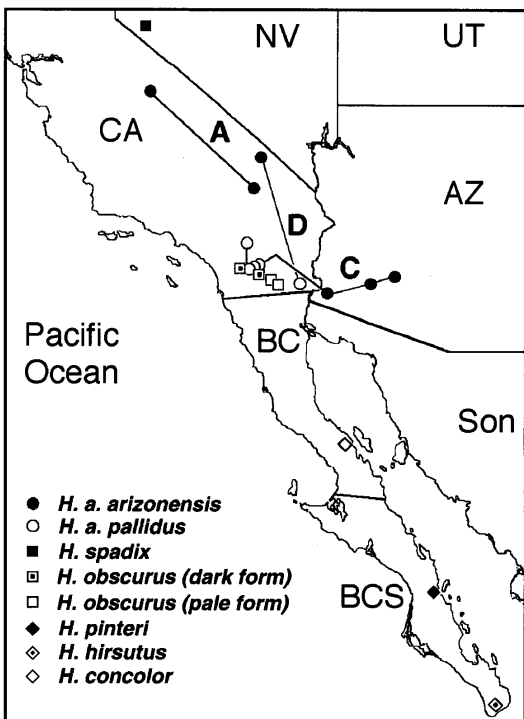
Fig. 1: South-western USA and northern Mexico showing distribution of specimens used in DNA analysis. See Table 1 for definitions of *H. arizonensis* genetic groups **A** = Hari_A, **C** = Hari_C, and **D** = Hari_D.

regarded as an incorrect original spelling according to the International Code of Zoological Nomenclature; see Sissom & Fet, 2000)

4. "pinteri" complex: *H. pinteri* Stahnke, 1969

Recent taxonomic history

Eight species comprise *Hadrurus* Thorell, 1876, presented here in chronological order of their description: *H. hirsutus* (Wood, 1863); *H. aztecus* Pocock, 1902; *H. arizonensis* Ewing, 1928; *H. spadix* Stahnke, 1940; *H. concolor* Stahnke, 1969; *H. pinteri* Stahnke, 1969; *H. obscurus* Williams, 1970; and *H. gertschi* Soleglad, 1976. Two subspecies were described in *H. arizonensis*, in addition to the nominotypic form: *H. arizonensis pallidus* Williams, 1970 and *H. a. austrinus* Williams, 1970. A reasonably complete set of information on the taxonomic history of this interesting genus can be obtained

from the original descriptions, and also from Williams (1970, 1980), Stahnke (1971), and Hjelle (1972).

The first comprehensive study of *Hadrurus* was conducted by Williams (1970). This excellent work involved large geographic samples of each species, and was important for many reasons; of particular interest was establishment of the true identity of *H. hirsutus*. Williams presented a taxonomy based primarily on the coloration and its patterns, specifically of the carapace and mesosoma, emphasizing morphometrics and setation, with less emphasis on other structures. His species key (non-phylogenetic in nature) was comprised of eight couplets, five entirely based on coloration, two partly relying on coloration, and only one dealing solely with morphology. At this time seven species were described, including one new species and two new subspecies.

Soleglad (1976) studied this genus entirely on trichobothrial patterns of the pedipalp chelae. His species groups and subgroups were delineated solely from these patterns. Of equal importance was the discovery of accessory trichobothria on both the external and internal aspects of the chela of several species, a very rare condition in the order as a whole (first reported by Gertsch & Soleglad, 1972, figs. 108–112). In Soleglad's (1976) study eight species were recognized, including one new species; and two groups, each divided into two subgroups, were defined using trichobothrial patterns. Although formal cladistic techniques were not performed or discussed, these phenetic groups and subgroups were considered and presented as "natural" groupings within the genus. The groups defined by Soleglad (1976) were in general consistent with the overall view of Williams's (1970) phylogeny of *Hadrurus*.

Material and methods

Material for DNA Analysis

All specimens were identified using diagnostic characters as defined by Williams (1970) and Soleglad (1976). We used a total of 19 specimens belonging to all six species of the "hirsutus" group (see Fig. 1).

Sequence identification abbreviations: Hxxx_Y_LLn; where *Hxxx* is abbreviated species

name; *Y* = colour variant (D = dark, P = pale); *LL* = general locality (BCS = Baja California Sur, BC = Baja California Norte, CA = California, AZ = Arizona, and NV = Nevada); *n* = sequence number for that species. For some data presentation, the following abbreviations are used: *HarN & HaNN = H. arizonensis*, both subspecies, *N* and *NN* = sequence number, *Hpin, Hspa, HobN* etc.

H. pinteri Stahnke, 1969 (1 specimen): Hpin_D_BS1 (Hpin): Isla Danzante, Baja California Sur, Mexico.

H. concolor Stahnke, 1969 (1 specimen): Hcon_P_BN1 (Hcon): Bahia de Los Angeles, Baja California Norte, Mexico.

H. hirsutus (Wood, 1863) (1 specimen): Hhir_D_BS1 (Hhir): Cabo San Lucas, Baja California Sur, Mexico.

H. spadix Stahnke, 1940 (1 specimen): Hspa_D_NV1 (Hspa): vicinity of Hawthorne, Mineral Co., Nevada, USA.

H. obscurus Williams, 1970 (4 specimens) (note that both dark and pale variants of this species are included): Hobs_D_CA1 (Hob1): Anza-Borrego Desert State Park (ABDSP), Culp Valley Camp (3382'), San Diego Co., California, USA; Hobs_D_CA2 (Hob2): ABDSP, Borrego Springs (787'), San Diego Co., California, USA; Hobs_P_CA3 (Hob3): ABDSP, Pinyon Mountain Road (3585'), San Diego Co., California, USA; Hobs_P_CA4 (Hob4): ABDSP, Indian Gorge Canyon (1085'), San Diego Co., California, USA.

H. arizonensis Ewing, 1928 (11 specimens) (dark and pale variants of *H. arizonensis* were identified as subspecies *H. a. arizonensis* and *H. a. pallidus* respectively): Hari_D_CA1 (Har1): Death Valley National Monument, Inyo Co., California, USA; Hari_D_AZ2 (Har2): Yuma, Yuma Co., Arizona, USA; Hari_D_AZ3 (Har3): Dateland, Yuma Co., Arizona, USA; Hari_P_CA4 (Har4), Hari_P_CA5 (Har5), Hari_P_CA6 (Har6): ABDSP, Borrego Springs, San Diego Co., California, USA; Hari_P_CA7 (Har7): Indio, Riverside Co., California, USA; Hari_D_CA8 (Har8): Soda Springs, near Baker, San Bernardino Co., California, USA; Hari_D_AZ9 (Har9): Gila Bend, Maricopa Co., Arizona, USA; Hari_P_CA10 (Ha10): Algodones, Imperial Co., California, USA; Hari_D_CA11 (Ha11): Shoshone, Inyo Co., California, USA.

DNA techniques

We used a standard protocol as described in Gantenbein *et al.* (1999). Total DNA was extracted from live or preserved scorpions. Extracted DNA was amplified by the polymerase chain reaction (PCR) in the Perkin Elmer 2400 PCR Thermocycler at the standard conditions (annealing temperature, 50°C). Mitochondrial LSU (large ribosomal subunit) 16S rRNA PCR primers were synthesized in the DNA Core Facility at Marshall University (Huntington, West Virginia). These primers have sequences CGATTTGAACTCAGATCA (forward, 18-mer) and GTGCAAAGGTAGCATAATCA (reverse, 20-mer) which correspond to the positions 12,867–12,887 and 13,218–13,310 in the *Drosophila yakuba* mitochondrial genome. The forward primer is a scorpion-specific version of the "universal" primer 16Sbr, or LR-J-12887, while the reverse primer has a scorpion-specific sequence designed by one of the authors (V.F.). The resulting PCR product was verified on 1% agarose electrophoretic gel and purified by Ultrafree MC 30 000 cellulose filters (Millipore, Inc.). Automated Sanger dideoxy sequencing of the double-stranded PCR product was performed at the Molecular Genetics Instrumentation Facility, University of Georgia (Athens, GA), on the ABI 9600 Sequencer. All DNA sequences were deposited in the GenBank Database with the accession numbers: AF312267 (Hpin), AF312268 (Hcon), AF312269 (Hhir), AF312270 (Hspa), AF318505- AF318508 (Hob1-Hob4), AF325509-AF325519 (Har1-Ha11).

Phylogenetic analysis software

Software packages *PAUP** version 4 (beta) (Swofford, 1998) and *Hennig86* version 1.5 (Farris, 1988) were used for Maximum Parsimony (MP) analysis of morphological characters. *Winclada* version 0.9.3 (Nixon, 1999) was used to generate the resulting *PAUP** MP cladogram showing distribution of all characters and their states. For molecular sequence analysis, *PAUP** was used to perform several techniques: Genetic distance calculation, MP with bootstrap resampling, Unweighted-Pair Group Method Arithmetic Mean (UPGMA) analysis, Neighbor-Joining (NJ) analysis, and Maximum Likelihood (ML) analysis. For the selection of an appropriate gene substitution model for ML analysis we used

Modeltest version 3.0 (Posada, 2000). In addition, we used *POY* 2.0 (Gladstein & Wheeler, 1999) to perform analysis by direct optimization (DO) (or generalized optimization: Wheeler, 1996), where both nucleotide bases and insertions/deletions (indels) are treated as observable evolutionary events, eliminating multiple sequence alignment methods. DO also allows a consensus of both molecular and morphology data as a single data matrix (i.e., simultaneous analysis). Cladograms and phylograms for the molecular sequences from *PAUP** and *POY* were generated by *TreeView* (Win 32) version 1.5.2 (Page, 1998). *MALIGN* version 2.7 (Wheeler & Gladstein, 1998) was used to analyse hand aligned sequences using parsimony-based criteria. All sequences evaluated by *PAUP** were exported in an interleaved format. Sequences exported to *POY* were stripped of all indels and organized in an non-interleaved GenBank format.

Cladistic analysis

Outgroup Considerations

The ideal outgroup for "hirsutus" group analysis would be a member of its presumed sister group, the "aztecus" group (either *H. aztecus* or *H. gertschi*), but neither species was available for DNA sampling. However, for morphology we did conduct a detailed analysis of all *Hadrurus* species (not presented in this paper), where, by declaring the "aztecus" group as an outgroup, we were able to establish the "hirsutus" subgroup as monophyletic within the "hirsutus" group. Thus, the "hirsutus" subgroup is a legitimate outgroup for morphology analysis as presented in this paper. For initial molecular analysis, *Vaejovis spinigerus* (Wood) (Vaejovidae), a "chactoid" species, was first chosen as an outgroup to the 19 *Hadrurus* sequences. All resulting phylogenetic trees supported *H. pinteri* as an outgroup (however, monophyly of the "hirsutus" subgroup was not supported). Therefore, for final molecular analysis, we used *H. pinteri* as the outgroup.

Morphology, character analysis

Morphology based analysis involved character coding of all important structures as identified by Williams (1970) and Soleglad (1976), augmented with additional trichobothria data and new analysis re-evaluating carapace and mesosoma coloration and patterns (presented here for the first time). **Appendices 1 and 2** present character descriptions and codings, and the resulting data matrix input to *PAUP**, respectively.

Coloration/Patterns: Williams (1970) relied heavily on coloration and patterns of the carapace and mesosoma. As will be seen below, some of the defined subspecies that were based primarily on coloration and the patterns do not have genetic support from the DNA data. This fact, coupled with the recent discovery of a pale form of *H. obscurus* from the south of Anza-Borrego Desert State Park, California, has led us to re-evaluate the patterns of three species complexes. The three complexes all show a common theme in their coloration: each species complex has a dark and pale counterpart. In the "arizonensis" complex, both subspecies exhibit a crescent-shaped melanic pattern interconnecting the lateral and median eyes (Figs. 2–5). Depending on the subspecies, as defined by Williams (1970), this crescent-shaped pattern can extend to the posterior edge of the carapace and to most of the mesosoma (Figs. 2–3). In either case, the crescent-shaped pattern is always present. Similarly, for the "spadix" complex, we have a wedge-shaped pattern interconnecting the lateral and median eyes, which is always present (Figs. 6–9), whether it is either masked out by the totally melanic pattern as seen in *H. spadix* (Fig. 6) or just the pattern itself, as exhibited by the pale form of *H. obscurus* (Fig. 9). It is interesting to point out that the *H. obscurus* pattern, as exhibited by central California topotypical specimens, extends more into the interocular area (Fig. 7). We suggest here that the minimal pattern that represents this species complex is the pale form of *H. obscurus* (Fig. 9). In contrast to the "arizonensis" subgroup, the "hirsutus" complex does not have a pattern interconnecting the lateral and median eyes (Figs. 10–12); we consider this an important observation. All pigmentation, if present, never reaches the lateral eyes and in general is present only at the median boundary (Figs. 10–11). As exhibited in the "arizonensis" subgroup, there is a great variability in the pigmentation on the posterior half of the carapace and mesosoma. *H. concolor* populations occurring in sand environments (Fig. 12), typically in Baja California Norte, are usually devoid of any melanic pigmentation (thus
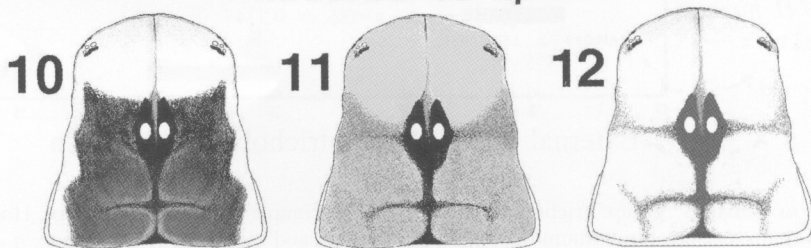
"arizonensis" complex
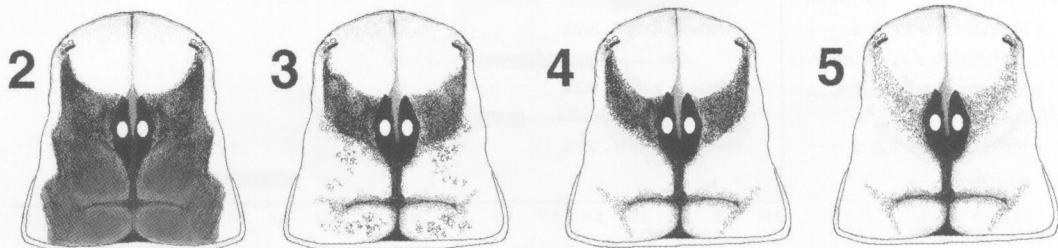
"spadix" complex
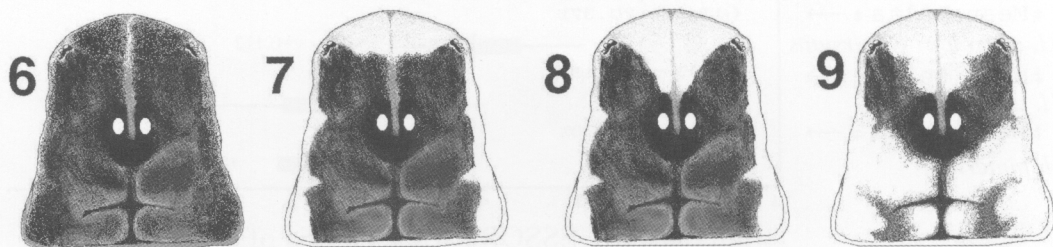
"hirsutus" complex

Figs. 2–12: *Hadrurus* carapace patterns. **2** *H. arizonensis arizonensis*, dark form, Soda Springs, California. **3** *H. a. arizonensis*, intermediate form, Soda Springs, California. **4** *H. a. pallidus*, typical crescent-shaped pattern, Borrego Springs, California. **5** *H. a. pallidus*, minimal crescent-shaped pattern, Borrego Springs, California. **6** *H. spadix*, typical form from northern Arizona. **7** *H. obscurus*, typical of topotypical form, Kern Co., California. **8** *H. obscurus*, dark form, Borrego Springs, California. **9** *H. obscurus*, pale form, southern Anza-Borrego Desert State Park (ABDSP), California. **10** *H. hirsutus*, Cabo San Lucas, Baja California Sur, Mexico. **11** *H. concolor*, mahogany form, Santa Rosalia, Baja California Sur, Mexico. **12** *H. concolor*, pale form, Bahia de Los Angeles, Baja California Norte, Mexico.

its name); other populations, especially those occurring in volcanic areas, have a variety of pigmentation and overall coloration. Some of these patterns are nearly identical to that found in the Cape region of Baja California populations of *H. hirsutus* (Fig. 10) where the carapace posterior and mesosoma are pigmented. This similarity even extends to the unusual melanic patterns found on the lateral and ventral aspect of metasomal segment V, unmatched in any other species. Yet, others exhibit a more overall mahogany coloration with a slight indication of the posterior pattern on the carapace (Fig. 11). It is proposed here that only the patterns interconnecting the lateral and median eyes are of prime importance in a phylogenetic sense. Other pigmentation found on the carapace posterior and/or mesosoma may be, in some cases, associated with local geographic
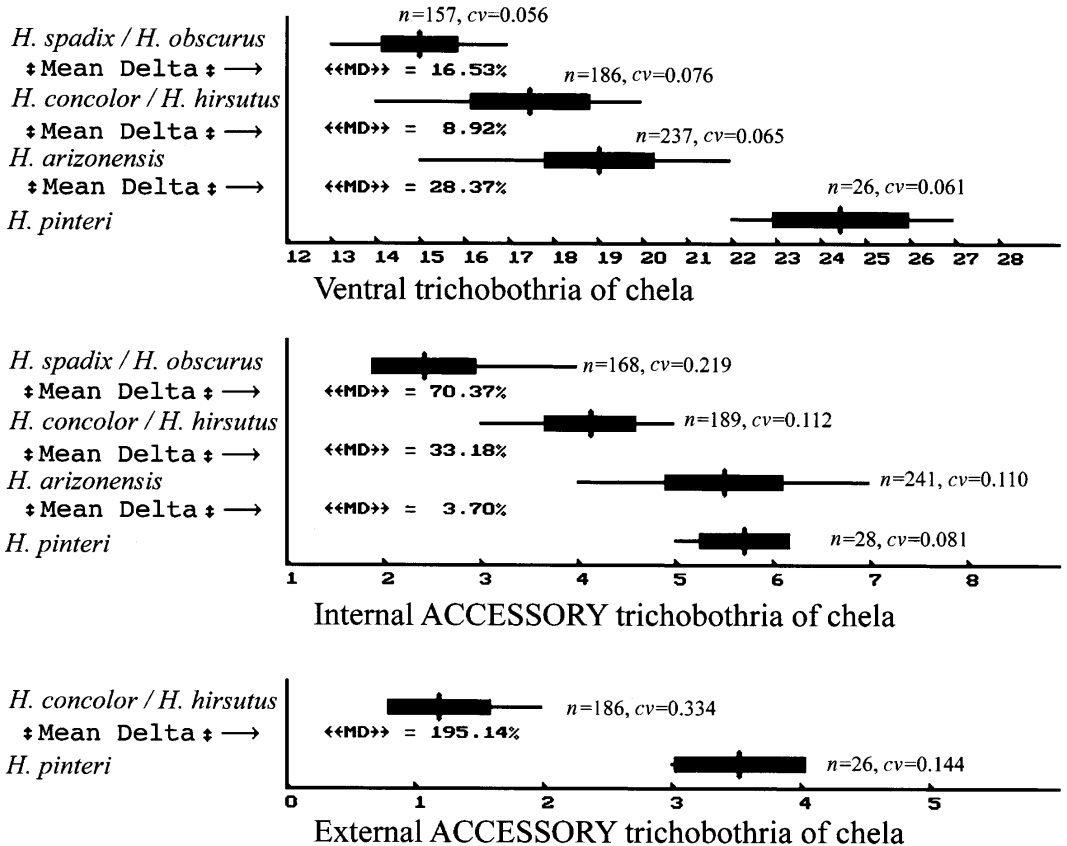
Fig. 13: *Hadrurus* "hirsutus" group. Trichobothria counts, mean comparisons as a percentage. Horizontal bar: minimum, maximum, corrected minimum/maximum (mean-sdev and mean+sdev), and mean; *n* = number of samples, *cv* = coefficient of variability (sdev/mean).

and/or microhabitat environments, and therefore cannot be used reliably in phylogenetic analysis. This hypothesis is supported, in part, by the molecular analysis presented below.
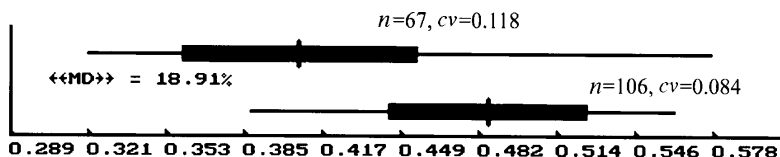
Trichobothria: The trichobothrial patterns of *Hadrurus* are quite exceptional for the order Scorpiones, exhibiting major neobothriotaxy ("major" and "minor" refer to the extent of additive neobothriotaxy). Large numbers of accessory trichobothria have evolved on the pedipalp, both on the patella, ventral and external faces, as well as on the ventral aspect of the chela. In addition, in many species, accessory trichobothria also occur on the internal and/or external aspects of the chela. In general, major neobothriotaxy on the chela is quite rare for "chactoids" (families Chactidae, Euscorpiidae, Scorpiopidae,

Superstitioniidae, Troglotayosicidae and Vaejovidae) (Vachon, 1974). Out of 43 "chactoid" + iurid genera (Fet *et al*., 2000), it occurs only in *Alloscorpiops*, *Euscorpius* (in part), *Anuroctonus*, and *Paravaejovis*; minor neobothriotaxy also occurs in the genera *Teuthraustes* (in part), *Pseudouroctonus* (in part) and *Alacran*. The accessory trichobothria only occur on the ventral aspect of the chela palm. The exceptions are *Alacran*, where they also occur on the external aspect of the palm (Francke, 1982, fig. 10) and *Teuthraustes*, where they also occur on its internal aspect (Vachon, 1974, fig. 186). The unprecedented occurrence of accessory trichobothria on these surfaces in *Hadrurus* was the basis for the group and subgroup definitions proposed by Soleglad (1976). In the present study, the

Ratio 1

H. spadix / H. obscurus
✸Mean Delta✸ ⟶
H. arizonensis

$n=67, cv=0.118$

⟨⟨MD⟩⟩ = 18.91%

$n=106, cv=0.084$

0.289 0.321 0.353 0.385 0.417 0.449 0.482 0.514 0.546 0.578

Ratio 2

H. obscurus
✸Mean Delta✸ ⟶
H. spadix

$n=20, cv=0.182$

⟨⟨MD⟩⟩ = 24.27%

$n=41, cv=0.115$

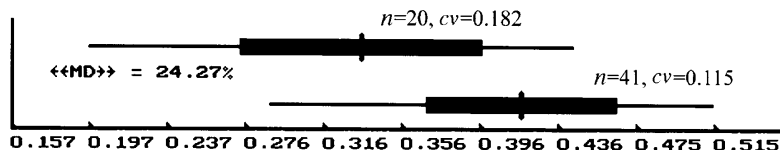0.157 0.197 0.237 0.276 0.316 0.356 0.396 0.436 0.475 0.515

Fig. 14: "arizonensis" subgroup. Trichobothria distance ratios, mean comparisons as a percentage. Ratio 1 = distance from trichobothrium *Est* to palm edge/palm length; Ratio 2 = distance between trichobothria *esb-eb/esb-Et5*. See Figure 13 for definitions of terms.

trichobothrial statistics presented in Soleglad (1976) have been augmented considerably with the data from additional specimens, especially from those species within the "hirsutus" group, increasing the original data set by over 50%; more than 600 individual data samples per chela area are included for the genus. The two *Hadrurus* groups are separated by the presence or absence of internal accessory trichobothria, present in the "hirsutus" group and absent from the "aztecus" group. Germane to this paper, and a key to the definition of the two subgroups of the "hirsutus" group, is the presence or lack of external accessory trichobothria: they are present in the "hirsutus" subgroup and lacking in the "arizonensis" subgroup. Further species grouping within these two subgroups is based on numbers of accessory trichobothria on the three chelal surfaces; the significant differences in mean values for these species are shown in Figure 13. Two morphometric ratios are defined involving relative trichobothria distances, specifically relevant within the "arizonensis" subgroup (Fig. 14). All of these diagnostic characters are discussed and illustrated in Soleglad (1976).

Other structures: Characters such as the aculear glands found on sexually mature males of some species, setation of the metasoma, and differences in pectinal tooth counts, as reported by Williams (1970), are also represented in our analysis. Of particular interest are aculear glands: the geographical distribution of species that exhibit this curious structure is unusual, two occurring in Baja California, *H. pinteri* and

H. concolor, and one in southern Mexico, *H. aztecus*.

Biogeography: Three groups discussed by Williams (1970) had essentially disjunct geographic ranges. Although character trends were provided for each, nothing significant or compelling was established as discrete diagnostic characters. In a sense, Williams's suggested phylogeny was dependent, in part, on associated biogeographical affinities. The phylogeny proposed by Soleglad (1976) was based on trichobothrial patterns, and also paralleled these same three disjunct geographic areas, two of which are pertinent to this study: (1) the southwestern United States and (2) Baja California, Mexico (the third is southern Mexico with the "aztecus" group). The two subgroups of interest are allopatric: the "hirsutus" subgroup in Baja, and "arizonensis" subgroup in the United States (with extensions into northern Baja and mainland Mexico). *H. pinteri* presents an interesting situation because it is the only *Hadrurus* species that occurs sympatrically with other species. Its range overlaps that of *H. concolor* by more than 50% and also that of *H. arizonensis* at its extreme southern range in Baja Norte, albeit *H. pinteri* is isolated in part by its unique microhabitat niche, volcanic terrain. Considering such distribution phylogenetically important and connected to presumed speciation events, we have represented these area-taxon associations as characters in our data matrix (see **Appendices 1 and 2**). In addition, larger areas were broken down into parapatric

subareas and microareas further refining geographic distribution of taxa.

Treatment of meristic and continuous data

Some of the data involving trichobothria, both counts (i.e. meristic) and morphometric ratios (i.e. continuous), involve certain amount of filtering prior to coding into discrete character state values. Although somewhat elegant techniques exist for the *a posteriori* establishment of relative gaps between these values (*sensu* gap weighting, Thiele, 1993) we have opted to use *a priori* gap coding methods based on the overall comprehensiveness of the statistical data itself. The gap weighting technique defined by Thiele (1993) requires that the character states be ordered so as to establish a relative cost for transition from one state to another; unfortunately, this ordering requirement also forces evolutionary assumptions on the linkage between the character states, an assumption we do not think is necessarily true for this data. Our gaps, of equal weight (one) and unordered, were established by two criteria: (1) mean differences were 15% or greater, and (2) corrected minimum and maximum (i.e. corrected by one standard error) overlap was less than 33%.

Molecular analysis

Several phylogenetic analyses were applied to the mtDNA sequences: genetic distance measurements, cluster analysis, maximum likelihood, maximum parsimony, and direct optimization.

Sequence alignment: We began with a hand alignment of all 19 sequences which was analysed with *PAUP** using a 3:1:1 transversion, transition, and indel (insertion/deletion) weight ratio (tv | tt | indel). We then experimented with various *MALIGN* alignments again analysing the resulting aligned sequences with *PAUP**, always using parsimony as our deciding factor. We tried realignments with and without SPR and TBR branch swapping and with and without a topology group specification. As our suggested topology, we used a tree generated by *POY*. Without the topology, *MALIGN* created three equally parsimonious sequence sets, with a cost of 158. Using a specified topology, one alignment was isolated with a cost of 160. Interestingly, without doing any swapping and no topology, *MALIGN* provided a single sequence, again with the cost of 158. Each of these sequences were analysed with

PAUP*; all five plus the hand aligned produced the same results: out of 369 characters, 31 were parsimonious informative, 127 steps, CI = 0.8504, CIE = 0.7164 (CI excluding uninformative characters), RI = 0.8882, and the same topology. Therefore, we decided to use the original hand alignment for further MP, UPGMA, and NJ analysis. For ML analysis we created a hand aligned set of sequences that had all sites removed that contained one or more ambiguous codes or indels (complete deletion, Swofford *et al.*, 1996: 453). For Direct Optimization analysis we converted the hand aligned sequences to non-interleaved GenBank format with indels stripped out.

Maximum Likelihood (ML) analysis: For ML analysis we used sequences stripped of all indels and ambiguous values. This allowed us to determine the unconstrainted likelihood of a tree (-ln L) which cannot be calculated if ambiguous base values are present. We calculated hierarchical likelihood ratio tests to determine the most appropriate gene substitution model using the *Modeltest* package 3.0 (Posada, 2000) as described in Posada & Crandall (1998). Input to *Modeltest* was *PAUP**-generated output of 56 sets of likelihood scores based on combinations of fourteen gene substitution models with and without invariable site and gamma distribution specifications. From this test the Hasegawa *et al.* (1985) gene substitution model was recommended with gamma distribution (HKY+G). Using the derived gamma shape parameter (*alpha*) estimate from the ML analysis above, we initiated several NJ sequences using a gamma rate qualifier and genetic substitution model HKY.

Most Parsimonious (MP) analysis: We experimented with various shape values on each side of the estimated shape parameter, analysing the resulting topologies as the rate of site heterogeneity increased and decreased. These topologies were compared to those derived from other molecular-based analyses. We experimented with 3:1:1, 3:1:4, and 5:1:4 tv | tt | indel weight ratios, which exhibited no significant differences in the resulting topology. We decided on the 3:1:1 weight ratio since it produced the least number of MPTs (two). From these sequences, we initiated eleven separate bootstrap samples, one of 1000 replicates and ten at 400 replicates each. The eleven bootstrap results were compared,
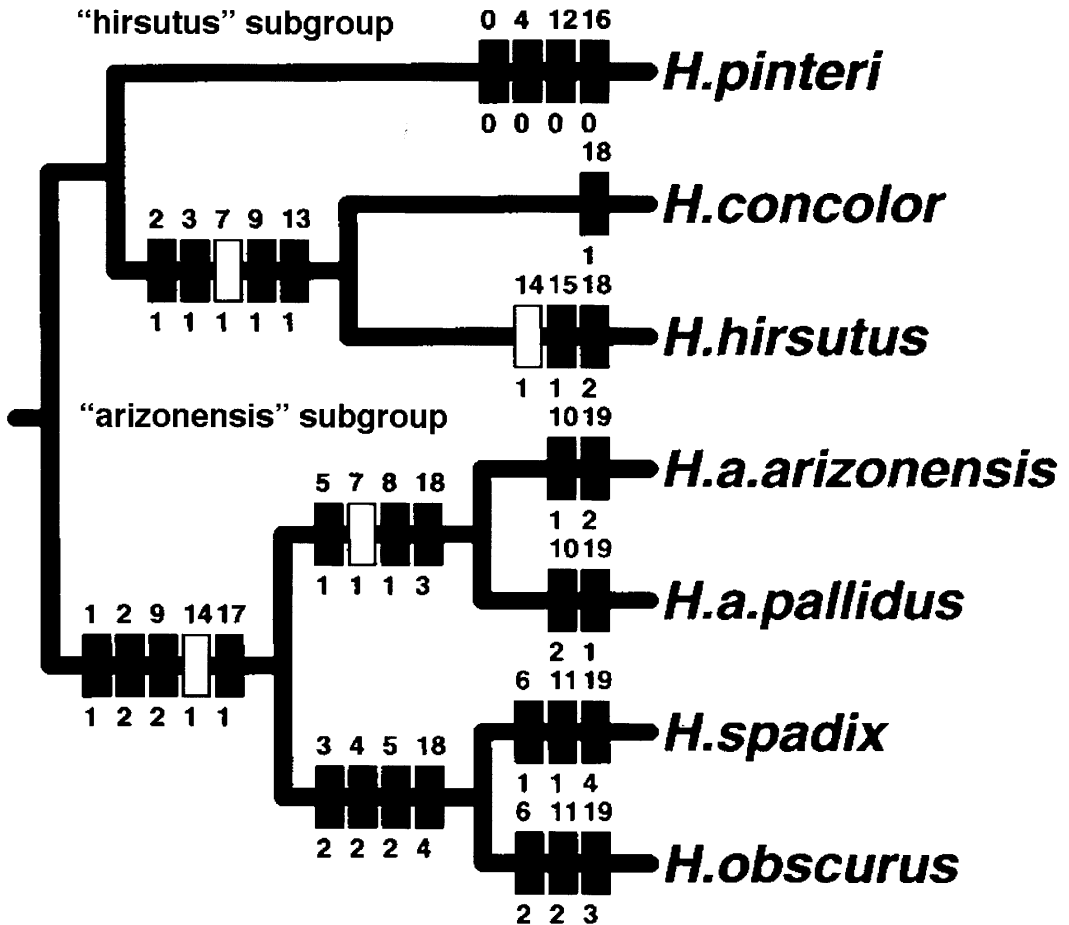
Fig. 15: Morphology-based cladogram: *Hadrurus* "hirsutus" group, showing "hirsutus" and "arizonensis" sub-groups. All characters and states shown with delayed optimization (DELTRAN). Character number above rectangle and character state below rectangle; black rectangles represent non-homoplasious characters and white rectangles represent homoplasious characters.

including both individual branch support and overall topology.

Direct Optimization (DO) analysis: With *POY* we exercised four separate tests each using a 3:1:1 tv|tt|indel weight ratio: (1) analysis of the sequences alone; (2) the sequences plus a data matrix as used in the morphology analysis; (3) the sequences plus the data matrix weighted by three; and 4) the sequences with a reduced morphology data matrix. In each test, the resulting tree set was further analysed with four *PAUP** consensus sequences. Note for test 3, since the sequences with *PAUP** yielded 31 cladistically informative

characters (out of 369) and the morphology analysis provided 12 informative characters, applying a weight of three to the latter placed the two data sets roughly on an equal basis. Actual data input to *POY* was a stepmatrix, the 19 GenBank format DNA sequences, and the morphology data matrix, full and altered, expanded for all 19 terminal taxa. The stepmatrix (or Sankoff character) specified a 3:1:1 tv|tt|indel weight ratio. **Appendix 3** contains the actual *POY* input line, the *Hennig86* compatible data matrix, and stepmatrix.

Hari_A (2) = Hari_D_CA1 + Hari_D_CA8
Hari_B (5) = Hari_D_AZ2 + Hari_D_AZ3 + Hari_P_CA4 + Hari_P_CA5 + Hari_D_AZ9
Hari_C (7) = **Hari_B** + Hari_P_CA6 + Hari_P_CA7
Hari_D (2) = Hari_P_CA10 + Hari_D_CA11

	Absolute Distance (Range)	Transversions (Range)	Uncorrected Percentage (Range)
H. spadix v. H. obscurus (4)	8–10	1–2	2.2–2.7
H. obscurus (4) [dark & pale]	0–3	0–1	0.0–0.8
Hari_A (2) [dark]	0	0	0.0
Hari_B (5) [dark & pale]	0	0	0.0
Hari_C (7) [dark & pale]	0–2	0	0.0–0.5
Hari_D (2) [dark & pale]	5	4	1.4
Hari_A (2) *v.* Hari_C (7)	4	0	1.1
Hari_A (2) *v.* Hari_D (2)	6–8	1–3	1.6–2.3
Hari_C (7) *v.* Hari_D (2)	2–5	1–3	0.5–1.4
H. arizonensis (11)	0–8	0–4	0.0–2.3
"ariz-comp" *v.* "spa-comp"	21–29	4–9	5.7–8.1
H. pinteri v. H. concolor	45	9	12.3
H. pinteri v. H. hirsutus	42	12	11.5
H. concolor v. H. hirsutus	23	7	6.3
H. pinteri v. "ariz-comp"	41–45	10–13	11.2–12.3
H. pinteri v. "spa-comp"	34–37	8–9	9.3–10.1
H. concolor v. "ariz-comp"	26–28	3–6	7.1–7.6
H. concolor v. "spa-comp"	23–26	5–7	6.3–7.1
H. hirsutus v. "ariz-comp"	23–26	6–9	6.3–7.3
H. hirsutus v. "spa-comp"	25–26	8–10	6.9–7.1

Table 1: Absolute genetic distance matrix: selective groupings.

Results

Morphology

Twenty characters are defined, ten of which are multivalued and none are weighted. Of the twenty characters, twelve are parsimony-informative. For the *PAUP** run, the "hirsutus" subgroup was declared the outgroup for rooting purposes. Because of the small number of taxa (seven) *PAUP**'s exact algorithm (Branch and Bound) was used in finding the most parsimonious trees (in this case, a single tree). Figure 15 illustrates this cladogram with the distribution of all characters and their states with delayed (DELTRAN) character optimization (which minimizes character reversals). Overall tree length was 36 steps (two extra steps), an overall consistency of 94.44% (CI = 0.9444, CIE = 0.9091 and RI = 0.8824), exhibiting only two homoplasious characters which are represented as independent losses on the cladogram: Character 7, setation between inferior median carinae, was lost independently on "hirsutus" and "arizonensis" complexes; Character 14, the loss of aculear glands, occurred independently on *H. hirsutus* and the "arizonensis" subgroup. This robust result was further supported by *PAUP** bootstrap sampling (based on ten separate samples, 10,000 replicates each), exhibiting 74–93% support on all clades. For confirmation we ran the same data matrix, character typing, and outgroup specification with *Hennig86*, obtaining the same results as above, a single tree of 36 steps with characters 7 and 14 both homoplasious (each occurring on the tree twice). It is particularly important to note that the resulting cladogram, which reflects both Williams's and Soleglad's diagnostic characters, supports the monophyly of the two subgroups of the "hirsutus" group as defined by Soleglad, and within the "arizonensis" subgroup, the "arizonensis" and "spadix" species complex clades are also preserved. It should also be noted

that combining both morphological characters and biogeographical associations created no incongruent conditions.

Genetic Distance analysis

We measured the genetic distance of all 19 sequences on a taxon-pair basis, adjusting resulting absolute distances based on several nucleotide substitution models. Table 2 shows the overall statistics of base value occurrence percentages and the ratios of transitions and transversions for these sequences. Table 3 presents distance data based on six gene substitution models for a select set of sequences. We discuss these results (Table 1) based on absolute distance (as

modified by factoring in percentages of indels and ambiguity codes on a taxon pair basis as described in Swofford *et al*. (1996: 454-455) and implemented in *PAUP**). The genetic distance is also reflected in the cluster analysis-generated phylograms (i.e. UPGMA and NJ) (Figs. 16–17).

"*arizonensis*" *subgroup:* Distances within the "spadix" complex (*H. spadix* and four *H. obscurus* sequences) are quite small, ranging from 2.2 to 2.7%. Within *H. obscurus*, zero to 0.8% distances are observed. It is important to note that Hobs_D_CA2 and Hobs_P_CA3, dark and pale forms respectively, exhibit zero differences. The 11 sequences of *H. arizonensis* provide interesting distance groupings. Based just on distances and preliminary results from UPGMA and

Number of bases per sequence (369) for each taxon (19)

	A	C	G	T	Amb	?/N	GAP	Compared
Hpin	139:0.377	71:0.192	42:0.114	116:0.314	0:0.000	0:0.000	1:0.003	0.997
Hcon	144:0.390	67:0.182	39:0.106	118:0.320	0:0.000	0:0.000	1:0.003	0.997
Hhir	142:0.385	64:0.173	40:0.108	120:0.325	0:0.000	2:0.005	1:0.003	0.992
Hspa	143:0.388	66:0.179	39:0.106	120:0.325	0:0.000	0:0.000	1:0.003	0.997
Hob1	141:0.382	64:0.173	39:0.106	124:0.336	0:0.000	0:0.000	1:0.003	0.997
Hob2	142:0.385	64:0.173	38:0.103	123:0.333	0:0.000	1:0.003	1:0.003	0.995
Hob3	142:0.385	64:0.173	39:0.106	123:0.333	0:0.000	0:0.000	1:0.003	0.997
Hob4	142:0.385	64:0.173	39:0.106	123:0.333	0:0.000	0:0.000	1:0.003	0.997
Har1	141:0.382	70:0.190	39:0.106	118:0.320	0:0.000	0:0.000	1:0.003	0.997
Har2	141:0.382	70:0.190	39:0.106	118:0.320	0:0.000	0:0.000	1:0.003	0.997
Har3	140:0.379	70:0.190	39:0.106	118:0.320	0:0.000	1:0.003	1:0.003	0.995
Har4	140:0.379	70:0.190	39:0.106	118:0.320	0:0.000	1:0.003	1:0.003	0.995
Har5	141:0.382	70:0.190	38:0.103	118:0.320	0:0.000	1:0.003	1:0.003	0.995
Har6	141:0.382	71:0.192	38:0.103	117:0.317	0:0.000	1:0.003	1:0.003	0.995
Har7	141:0.382	69:0.187	38:0.103	119:0.322	0:0.000	1:0.003	1:0.003	0.995
Har8	141:0.382	69:0.187	38:0.103	118:0.320	0:0.000	2:0.005	1:0.003	0.992
Har9	141:0.382	70:0.190	39:0.106	117:0.317	0:0.000	1:0.003	1:0.003	0.995
Ha10	142:0.385	69:0.187	39:0.106	118:0.320	0:0.000	0:0.000	1:0.003	0.997
Ha11	132:0.358	67:0.182	43:0.117	115:0.312	0:0.000	11:0.030	1:0.003	0.967

A_Base Values: 2676 (0.384) (Amb|GAPs factored in)
C_Base Values: 1289 (0.185)
G_Base Values: 744 (0.107)
T_Base Values: 2261 (0.324)
Unknown Values: 22 (0.003)
GAP Values: 19 (0.003)

295   Positions are IDENTICAL / 74 Positions mismatched
64   Positions mismatched 1 base-value (2 values for this site)
10   Positions mismatched 2 base-values (3 values for this site)

Base_Freqencies:
BF = $1 - (af^2 + cf^2 + gf^2 + tf^2) = 0.70179$

Global Transition/Transversion Ratio:
Transitions (2398) / Transversions (715) = 3.354

Transitions: A<—>G: 728
                    C<—>T: 1670

Transversions: A<—>C: 160
                    A<—>T: 398
   C<—>G: 51
                    G<—>T: 106

Table 2: mtDNA (16S rRNA) sequence statistics.

Taxon		[base-pairs-compared | mismatches]			
TT	**=**	**Transitions (Ti)/Transversions (Tv) (Ti/Tv)**			
UC	**=**	**Uncorrected distance (p-distance, or dissimilarity (D) distance)**			
JC	**=**	**Jukes & Cantor (1969); equal base frequency, one substitution type**			
F81	**=**	**Felsenstein (1981); unequal base frequency, one substitution type**			
K2P	**=**	**Kimura (1980); equal base frequency, unequal Ti:Tv**			
HKY	**=**	**Hasegawa, Kishino & Yano (1985); unequal base frequencies, unequal Ti:Tv**			
TAMNEI	**=**	**Tamura & Nei (1993); unequal base frequency, three substitution types**			

	Hcon	Hhir	Hspa	Hob1	Harl
Hpin	[369|45]	[369|42]	[369|37]	[369|35]	[369|42]
TT	36/9 (4.004)	30/12 (2.502)	29/8 (3.621)	25/10 (2.500)	32/10 (3.202)
UC	(0.12264)	(0.11465)	(0.10067)	(0.09516)	(0.11429)
JC	(0.13391)	(0.12442)	(0.10810)	(0.10176)	(0.12400)
F81	(0.13475)	(0.12515)	(0.10864)	(0.10225)	(0.12469)
K2P	(0.13729)	(0.12635)	(0.11009)	(0.10302)	(0.12642)
HKY	(0.14051)	(0.12858)	(0.11207)	(0.10449)	(0.12883)
TAMNEI	(0.14051)	(0.12869)	(0.11211)	(0.10463)	(0.12886)
Hcon		[368|23]	[368|23]	[368|25]	[368|27]
TT		16/7 (2.297)	18/5 (3.600)	18/7 (2.571)	24/3 (8.000)
UC		(0.06306)	(0.06250)	(0.06793)	(0.07337)
JC		(0.06587)	(0.06526)	(0.07121)	(0.07721)
F81		(0.06608)	(0.06547)	(0.07147)	(0.07750)
K2P		(0.06633)	(0.06595)	(0.07183)	(0.07870)
HKY		(0.06693)	(0.06669)	(0.07260)	(0.08001)
TAMNEI		(0.06807)	(0.06681)	(0.07292)	(0.08084)
Hhir			[368|25]	[368|26]	[368|23]
TT			17/8 (2.135)	16/10 (1.605)	17/6 (2.840)
UC			(0.06830)	(0.07114)	(0.06278)
JC			(0.07162)	(0.07475)	(0.06557)
F81			(0.07187)	(0.07503)	(0.06577)
K2P			(0.07212)	(0.07511)	(0.06614)
HKY			(0.07282)	(0.07576)	(0.06678)
TAMNEI			(0.07334)	(0.07649)	(0.06727)
Hspa				[368|9]	[368|21]
TT				7/2 (3.500)	17/4 (4.250)
UC				(0.02446)	(0.05707)
JC				(0.02486)	(0.05935)
F81				(0.02489)	(0.05952)
K2P				(0.02496)	(0.05999)
HKY				(0.02506)	(0.06062)
TAMNEI				(0.02522)	(0.06107)
Hob1					[368|23]
TT					17/6 (2.833)
UC					(0.06250)
JC					(0.06526)
F81					(0.06547)
K2P					(0.06582)
HKY					(0.06647)
TAMNEI					(0.06728)

Table 3: Distance matrix pairwise comparisons based on several gene substitution models.

NJ analysis, we recognize three genetic subclades within the "arizonensis" complex: Hari_A, Hari_C, and Hari_D as defined in Table 1 and shown in the map in Figure 1. The two sequences in Hari_A, both from dark forms, exhibit zero distances from each other. Similarly, of the seven sequences in clade Hari_C, five (Hari_B) exhibit zero distances, involving two pale forms from
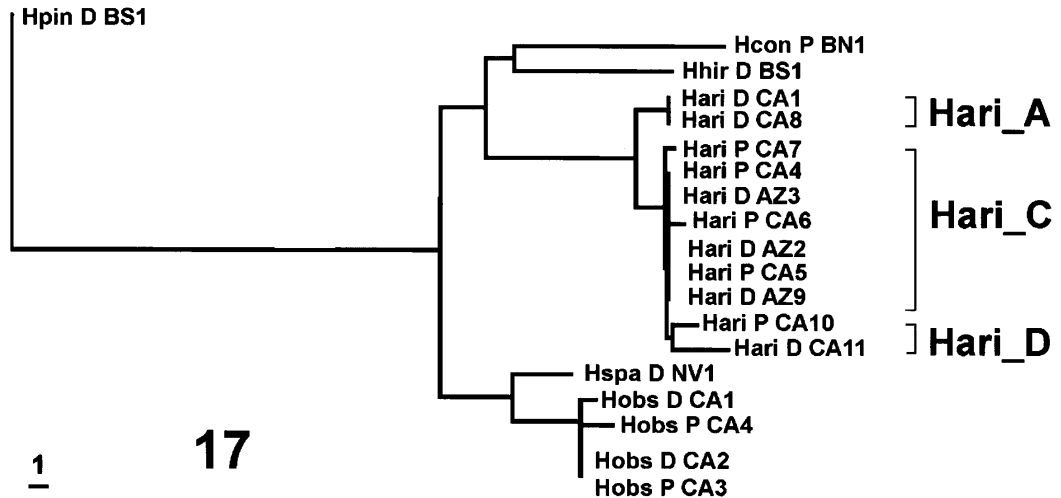
Figs. 16–17: Molecular phylograms, cluster analysis: tv | tt | indel weight ratio = 3:1:0. **16** Unweighted Pair Group Method Arithmetic Mean (UPGMA). **17** Neighbor-Joining (NJ).

southern California and three dark forms from southern Arizona. These five sequences are from specimens that correspond morphologically to and occur in the geographical areas of *H. a. arizonensis* and *H. a. pallidus*, as stated in the diagnoses of Williams (1970). In fact, specimens Hari_P_CA4 and Hari_D_AZ9 are separated approximately by 180 miles, each occurring well in their respective ranges as specified by Williams. Distance between Hari_A and Hari_C is 1.1% with no transversions. The third clade,

Hari_D, is interesting because its two sequences are the only *H. arizonensis* that exhibit transversion mismatches with the other sequences, ranging from one to three. This subclade also involves both dark and pale forms. The overall distance range within the 11 sequences of the "arizonensis" complex is from 0–2.3%, a little less than that for the "spadix" complex. The distance range between the two complexes, involving a total of 16 sequences, is from 5.7–8.1%, forming two
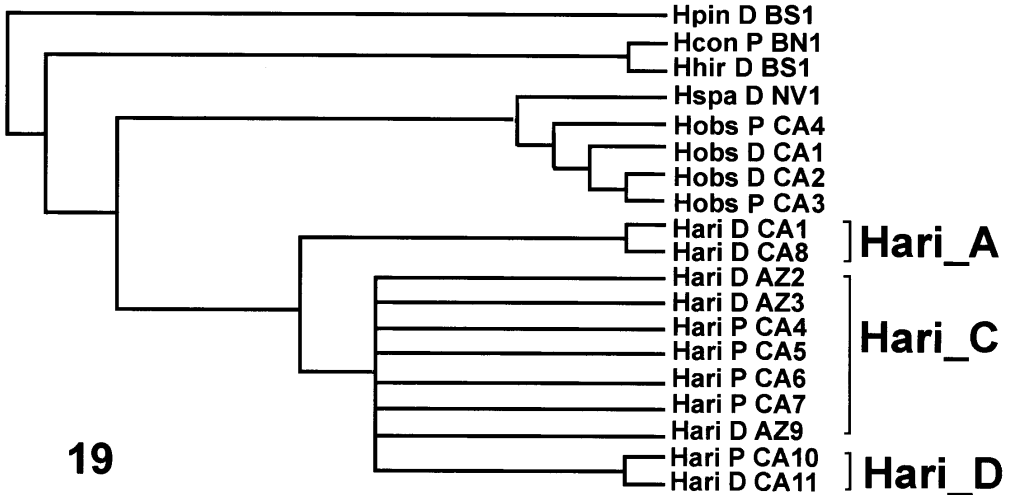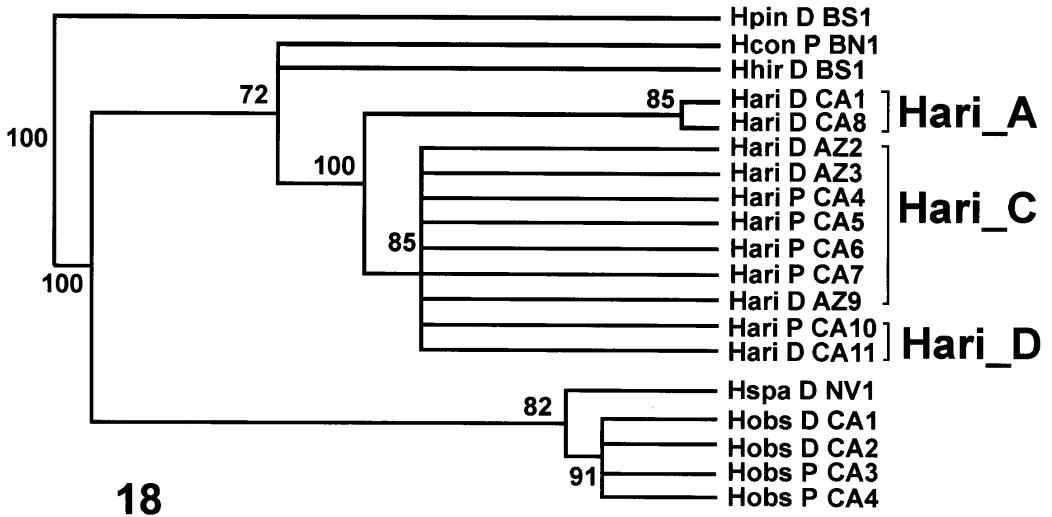
Figs. 18–19: Molecular and/or morphology cladograms, tv l tt l indel weight ratio = 3:1:1. **18** Maximum Parsimony (MP), bootstrap support, 1000 replicates. **19** Direct Optimization (DO) analysis.
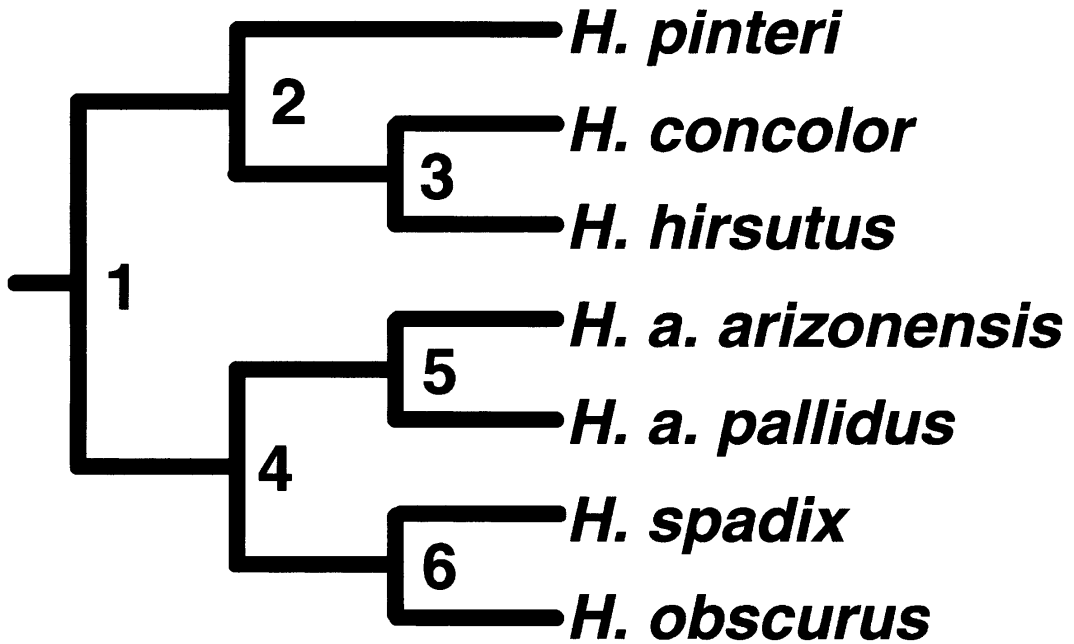
well-defined clades within the "arizonensis" subgroup.

"hirsutus" subgroup: Genetic distancses found within this subgroup are much greater than those found within the "arizonensis" subgroup. *H. pinteri* exhibits the most distance within this subgroup, ranging from 11.5–12.3%, involving 9–12 transversions. Interestingly, although *H. pinteri* shows the most distance in general within the "hirsutus" group (9.3–12.3%, 8–13 transverions), the other two species of this sub-

group, *H. concolor* and *H. hirsutus*, are also quite distant from each other as well as to *H. pinteri* and, exhibit roughly the same distance from the two species complexes in the "arizonensis" subgroup.

Cluster Analysis

UPGMA and NJ (Figs. 16–17): UPGMA and NJ analyses were conducted with a 3:1:0 tv l tt l indel weight ratio using an absolute distance parameter (we also conducted UPGMA and NJ

```
┌──────────────────────── H. pinteri
┌─ 2 ┤
│    │    ┌─────────────── H. concolor
│    └─ 3 ┤
│         └─────────────── H. hirsutus
─ 1 ┤
│         ┌─────────────── H. a. arizonensis
│    ┌─ 5 ┤
│    │    └─────────────── H. a. pallidus
└─ 4 ┤
     │    ┌─────────────── H. spadix
     └─ 6 ┤
          └─────────────── H. obscurus
```

| | | NODE Support Phylogenetic Analysis Component | | | | | |
| | Morph | Molecular | | | | | |
Node	MP	GenDis	UPGMA	NJ	ML	MP	DO + Morph
1	yes	und	yes-L	NO	NO	NO	yes-L
2	yes	und	yes	NO	NO	NO	yes
3	yes	und	yes	yes	yes	yes	yes
4	yes	yes	yes	yes	yes	yes	yes
5	yes	NO	NO	NO	NO	NO	NO
6	yes	yes	yes	yes	yes	yes	yes

Fig. 20: Node support, all phylogenetic analyses. Yes = full support; Yes-L = supported but ladderized; NO = no support; und = undetermined.

using all six gene substitution models as described in Table 3, resulting in no topological differences). Inclusion of this analysis here is interesting for two reasons: (1) it provides a visual microview of the distances discussed above for the two clades within the "arizonensis" subgroup, showing all subtle differences, and (2) it illustrates the problems that algorithms of this type have with sequences that clearly do not demonstrate equal gene substitution rates across all sites for all taxa. For example, with closely related taxa, as reflected in the "arizonensis" subgroup, all distances and resulting topologies are reasonable and endorsed by all other analyses: the three major genetic subclades of *H. arizonensis* are well delineated, including the subtle distance differences of Hari_P_CA6 and Hari_P_CA7 from other members in Hari_C. In contrast, the three taxa in the "hirsutus" subgroup are clearly distant from the other subgroup and, evidently, from each other as well, especially *H. pinteri*. It is interesting to note that, for the latter, UPGMA and NJ do not agree as to the placement of *H. concolor* and *H. hirsutus*. However, with either algorithm, note that *H. concolor* and *H. hirsutus* do form a single clade.

Maximum Likelihood (ML) analysis

The most interesting aspect of this analysis was studying the impact on NJ topology using different values of the shape parameter (*alpha*) for gamma distribution modeling of rate heterogeneity across sites. NJ generated phylograms, based on the Hasegawa *et al.* (1985) gene substitution model with gamma distribution (as recommended by *Modeltest*), demonstrated this dramatically. Using the estimated shape parameter derived from ML analysis (*alpha* = 0.181546), we produced a degraded topology where taxon *H. spadix* split from *H. obscurus* and joined the "arizonensis" and "hirsutus" complexes. By increasing *alpha* to 0.402306 we were able to stablize the topology, as shown in Figure 18. Note that a value as high as 0.402305 still produced a degraded phylogram. As *alpha* gets very large, all sites approach a relative value of 1 (thus an even site evolution rate), for small *alphas* (less than 1), the rate of site heterogeneity increases (Swofford *et al.*, 1996: 444). Therefore, based on ML analysis, the estimated shape parameter for the gamma distribution was small suggesting that the overall rate of evolutionary change over sites is uneven.

Most Parsimonious (MP) analysis (Heuristic alogorithm): Using a 3:1:1 weight ratio, we generated two maximum parsimonious trees (MPTs) with 127 steps, CI = 0.8504, CIE = 0.7164 and RI = 0.8882; a 3:1:4 weight ratio resulted in six MPTs with 158 steps, CI = 0.8608, CIE = 0.7215 and RI = 0.8878; and with a 5:1:4 weight ratio, we generated four MPTs with 206 steps, CI = 0.8738, CIE = 0.7426 and RI = 0.8952. All three weight ratios produced essentially the same topology and therefore, based on the smaller number of MPTs, we decided on the 3:1:1 ratio for the bootstrap analysis. The results from the eleven separate bootstrap samples (one at 1000 replicates and ten at 400 replicates) produced the same topology and relative branch support. A polytomy was created for taxa *H. concolor*, *H. hirsutus*, and the "arizonensis" complex. Support for this polytomy was the weakest, ranging from 68 to 75% across the eleven samples. Other support was better, as shown in Figure 18. Finally, *PAUP** did not accept the "hirsutus" subgroup as the outgroup since it was not monophyletic, consequently we declared *H. pinteri* as our outgroup for reasons given earlier in this paper. The topology reflected in

this analysis is the same as exhibited in NJ and ML analyses.

Direct Optimization (DO) analysis

Test 1: Analysing just the DNA sequences, *POY* isolated seven trees with a cost of 125, essentially exhibiting what we refer to as the "accepted topology" for molecular data (Fig. 19): the "hirsutus" subgroup ladderized, the "arizonensis" subgroup forming a single clade with the "spadix" and "arizonensis" species complexes forming two defined subclades. There was slight variability within the four sequences of *H. obscurus* and the *H. arizonensis* clade split into the same three subclades as those described for genetic distances (Hari_A, Hari_C, and Hari_D, see Table 1). All four *PAUP** consensus tests (i.e. strict, semi-strict, majority-rule, and Adams) agreed with the topology.

Test 2: This interesting test illustrated dramatically the conflict of molecular sequences with morphology for two subspecies of *H. arizonensis*: twenty trees were produced with a cost of 162, exhibiting the same minor variability in *H. obscurus* but showing major variability within the "arizonensis" species complex. Except for Hari_D_CA1 and Hari_D_CA8 (Hari_A in Fig. 18), the topology was not consistent. *POY* attempted to group by colour phase, dark or pale, as dictated by morphology, and therefore did not form the standard topology as seen in the other tests. Furthermore, the four *PAUP** consensus tests did not agree.

Test 3: We weighted the entire morphology matrix by three which resulted in less variability, the weighting forced *POY* to group *H. arizonensis* as three clades, endorsing dark *v.* pale sequences (i.e. giving morphology more weight offset the impact of the sequences which does not group by colour). This test produced 20 trees with a cost of 234.

Test 4: Based on results of the three prior tests and the genetic distance analysis discussed above (Hari_B), *as a hypothesis*, we questioned the validity of *H. a. pallidus* as a valid subspecies, and consequently, altered two characters in the morphology matrix that distinguished this taxon: character 10, the distinction of the carapacial patterns and character 19, the biogeographical breakdown into two microareas. This resulted in 20 trees of a cost of 159, where all four consensus tests yielded the accepted topology. Note that by

increasing the weight from 2 to 3 for transversions, POY was able to isolate clade Hari_D, the only *H. arizonensis* that exhibited transversion mismatches with the other sequences of its species.

Discussion

Figure 20 shows a summary of clade node support based on seven analyses conducted against morphology and the molecular sequences of the "hirsutus" group. In this figure the putative phylogeny of Williams and Soleglad is shown with all six nodes identified. Except for results based entirely on morphology, none of the analyses supported all clades, although DO (with and without the inclusion of a reduced morphology matrix) and UPGMA did support all upper level clades, only disagreeing with the delineation of the two subspecies of *H. arizonensis* (node 5). Also, none of the analyses based solely on molecular sequences supported these two subspecies. Using molecular analyses only, the conflict in upper clade support is quite interesting and definitely implies that the rate of gene substitution across sites between the "hirsutus" and "arizonensis" subgroups of *Hadrurus* is not equal. In general these analyses did not form a clean dichotomy between the two subgroups and exhibited inconsistent support even within the "hirsutus" subgroup, specifically nodes 1 and 2. As shown above in the analyses discussions, UPGMA and NJ gave conflicting topologies at this level. In addition, MP analysis could not adequately support the upper clades. Genetic distance analysis, though illustrating large distances between the two subgroups, could not consistently support the topology within the "hirsutus" subgroup (i.e. nodes 2 and 3). In stark contrast, delineation of the two species complexes within the "arizonensis" subgroup are well defined by all analyses (node 4). Important to note in this subgroup is that the number of transversions is quite small to nil, implying a recent divergence within the subgroup. In contrast, the "hirsutus" subgroup exhibits greater absolute distance as well as more transversions, both within the subgroup and between the two subgroups, implying an earlier divergence. This hypothesized mismatch in nucleotide substitution rate could account for the discrepancies and inconsistencies exhibited in the different molecular analyses. ML analysis using gamma distribution certainly supports this hypothesis.

We conclude that MP analysis based on morphology and DO analysis, with and without morphological considerations, give us our best view of the true phylogeny of the "hirsutus" group of *Hadrurus*, since they together incorporate *both* morphology and molecular information. The "hirsutus" and "arizonensis" subgroups are shown to be monophyletic within the "hirsutus" group of *Hadrurus* as are the "arizonensis" and "spadix" complexes within the "arizonensis" subgroup. Since the three identified genetic subclades within *H. arizonensis* (Hari_A, Hari_C, and Hari_D) are not distinguished by any consistent morphology, do not form discrete geographic populations, and only exhibit minor genetic distances, we consider all three subclades as species *H. arizonensis*.

Taxonomic emendations

Based on extensive molecular data presented above, we are declaring the subspecies *H. arizonensis pallidus* Williams, 1970 a synonym of *H. arizonensis arizonensis* Ewing, 1928.

Acknowledgements

References

FARRIS, J. S. 1988: *Hennig86, Version 1.5*. Published by the author.

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

FET, V., SISSOM, W. D., LOWE, G. & BRAUN-WALDER, M. E. 2000: *Catalog of the scorpions of the world (1758–1998)*. New York, NY: New York Entomological Society.

FRANCKE, O. F. 1982: Studies of the scorpion sub-families Superstitioninae and Typhlochactinae, with description of a new genus (Scorpiones, Chactoidea). *Ass. mex. Cave Stud. Bull.* **8**: 51–61.

GANTENBEIN, B., FET, V., LARGIADÈR, C. & SCHOLL, A. 1999: First DNA phylogeny of the genus *Euscorpius* Thorell, 1876 (Scorpiones, Euscorpiidae) and its bearing on the taxonomy and biogeography of this genus. *Biogeographica* **75**: 59–72.

GERTSCH, W. J. & SOLEGLAD, M. E. 1972: Studies of North American scorpions of the genera *Uroctonus* and *Vejovis* (Scorpionida, Vejovidae). *Bull. Am. Mus. nat. Hist.* **148**: 549–608.

GLADSTEIN, D. & WHEELER, W. 1999: *POY: phylogeny reconstruction via direct optimization of DNA Data, Version 2.0*. New York, NY: American Museum of Natural History.

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

HJELLE, J. T. 1972: Scorpions of northern California coast ranges (Arachnida: Scorpionida). *Occ. Pap. Calif. Acad. Sci.* **92**: 1–59.

JUKES, T. H. & CANTOR, C. R. 1969: Evolution of protein molecules. *In* H. N. Munro (ed.). *Mammalian protein metabolism*. New York, NY: Academic Press: 21–132.

KIMURA, M. 1980: A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. mol. Evol.* **16**: 110–120.

NEI, M. 1987: *Molecular evolutionary genetics*. New York, NY: Columbia University Press.

NIXON, K. C., 1999: *Winclada (BETA) 0.9.99i*. Ithaca, NY: Published by the author.

PAGE, R. D. M. 1998: *TreeView (win32) 1.5.2*. Glasgow, Scotland.

POSADA, D. 2000: *Modeltest Version 3.0*. Provo, UT: Published by the author.

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

SISSOM, W. D. & FET, V. 2000: Family Iuridae. *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: 409–420.

SOLEGLAD, M. E. 1976: The taxonomy of the genus *Hadrurus* based on chela trichobothria (Scorpionida: Vejovidae). *J. Arachnol.* **3**: 113–134.

STAHNKE, H. L. 1971: The redescription of the scorpion *H. aztecus* (Vejovidae). *Ent. News* **82**: 121–131.

SWOFFORD, D. L. 1998: *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. 4.0 Beta Version. 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.

THIELE, K. 1993: The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* **9**: 275–304.

VACHON, M. 1974: Étude des caractères utilisés pour classer les familles et les genres de scorpions (Arachnides). 1. La trichobothriotaxie en Arachnologie, Sigles trichobothriaux et types de trichobothriotaxie chez les scorpions. *Bull. Mus. natn. Hist. nat. Paris* **3** (140): 857–958.

WHEELER, W. 1996: Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics* **12**: 1–9.

WHEELER, W. & GLADSTEIN, D. 1998: *MALIGN Version 2.7, Parallel Version 1.5*. New York, NY: American Museum of Natural History.

WILLIAMS, S. C. 1970: A systematic revision of the giant hairy scorpion genus *Hadrurus*. *Occ. Pap. Calif. Acad. Sci.* **87**: 1–62.

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

Appendix 1. Morphology and biogeography character descriptions and taxon assignments

TRICHOBOTHRIA-based characters, the CHELA (0–6)

CHAR 0: external ACCESSORY trichobothrium on fixed finger:
  0 = yes, *H. pinteri*
  1 = no

CHAR 1: external ACCESSORY trichobothria on palm:
  0 = yes, "hirsutus" subgroup
  1 = no, "arizonensis" subgroup

CHAR 2: external ACCESSORY trichobothria NUMBER:
  0 = 3–4, *H. pinteri*
  1 = 1–2, *H. hirsutus* and *H. concolor*
  2 = none, "arizonensis" subgroup

Supporting data:
  H.pinteri 3.0–4.0 (3.54) (±0.51) [026]:{3.0–4.0}→0.144
  H.con/hir 1.0–2.0 (1.20) (±0.40) [186]:{0.8–1.6}→0.334

CHAR 3: internal ACCESSORY trichobothria NUMBER:
  0 = 5–6, *H. pinteri* and *H. arizonensis*
  1 = 4–5, *H. hirsutus* and *H. concolor*
  2 = 2–3, *H. spadix* and *H. obscurus*

Supporting data:
  H.pinteri 5.0–6.0 (5.71) (±0.46) [028]:{5.3–6.2}→0.081
  H.ariz 4.0–7.0 (5.51) (±0.61) [241]:{4.9–6.1}→0.110
  H.con/hir 3.0–5.0 (4.14) (±0.46) [189]:{3.7–4.6}→0.112
  H.spa/obs 2.0–4.0 (2.43) (±0.53) [168]:{1.9–3.0}→0.219

CHAR 4: ventral trichobothria NUMBER:
  0 = > 22, *H. pinteri*
  1 = 16–20, *H. arizonensis*, *H. hirsutus* and *H. concolor*
  2 = 14–16, *H. spadix* and *H. obscurus*

Supporting data:
  H.pinteri 22.0–27.0 (24.46) (±1.50) [026]: {23.0–26.0}→0.061
  H.ariz 15.0–22.0 (19.05) (±1.24) [237]: {17.8–20.3}→0.065
  H.con/hir 14.0–20.0 (17.50) (±1.32) [186]: {16.2–18.8}→0.076
  H.spa/obs 13.0–17.0 (15.01) (±0.84) [157]: {14.2–15.9}→0.056

CHAR 5: Est/Palm_length RATIO:
  0 = inapplicable
  1 = 0.445–0.526 (0.486), *H. arizonensis*
  2 = 0.360–0.456 (0.408), *H. spadix* and *H. obscurus*

Supporting data:
  ***** MEAN DELTA: (0.486:0.408) = 0.189
  H.ariz 0.388–0.563 (0.486) (±0.041) [106]: {0.445–0.526}→0.084
  H.spa/obs 0.321–0.578 (0.408) (±0.048) [067]: {0.360–0.456}→0.118

CHAR 6: esb-eb/esb-Et5 RATIO:
  0 = inapplicable
  1 = 0.369–0.466 (0.417), *H. spadix*
  2 = 0.275–0.397 (0.336), *H. obscurus*

Supporting data:
  ***** MEAN DELTA: (0.417:0.336) = 0.243
  H.spad 0.290–0.515 (0.417) (±0.048) [041]:{0.369–0.466}→0.115
  H.obsc 0.197–0.444 (0.336) (±0.061) [020]:{0.275–0.397}→0.182

SETAE-based character (7)

CHAR 7: numerous setae between inferior median carinae, metasomal segments I-III:
  0 = yes, *H. spadix*, *H. obscurus* and *H. pinteri*
  1 = no

COLORATION/PATTERNS-based characters (8–13)

CHAR 8: chelal fingers pigmented, reddish to black:
 0 = yes
 1 = no, *H. arizonensis*

CHAR 9: Coloration and patterns on carapace and mesosoma:
 0 = Carapace and mesosoma entirely melanic, no significant variability, *H. pinteri*
 1 = Carapace and mesosoma variable, from complete concolorous to melanic pattern on carapace posterior half and mesosoma, NEVER in interocular area, *H. concolor* and *H. hirsutus*
 2 = Melanic pattern on carapace CONNECTING median tubercle and lateral eyes, carapace posterior and mesosoma variable, *H. spadix, H. obscurus* and *H. arizonensis*

CHAR 10: Melanic crescent-shaped pattern on carapace:
 0 = no
 1 = yes, carapace posterior and mesosoma melanic, *H. a. arizonensis*
 2 = yes, carapace posterior and mesosoma without melanic pattern, *H. a. pallidus*

CHAR 11: Melanic wedge-shaped pattern on carapace:
 0 = no
 1 = yes, interocular area melanic (wedge-shaped pattern masked by total melanic pattern), *H. spadix*
 2 = yes, interocular area clear to some degree (wedge-shaped pattern partially masked by intermediate melanic pattern), *H. obscurus*

CHAR 12: Marbled melanic patterns on metasoma and chelal palms:
 0 = yes, *H. pinteri*
 1 = no

CHAR 13: Melanic pattern on lateral and ventral aspect of metasomal segment V:
 0 = no
 1 = yes, variable, dependent on local geographic races, *H. concolor* and *H. hirsutus*

ACULEAR GLANDS-based character (14)

CHAR 14: Aculear glands present on telson of sexually mature males:
 0 = yes, *H. pinteri* and *H. concolor*
 1 = no

PECTINAL TOOTH COUNT-based character (15)

CHAR 15: gross tooth count ranges for males and females:
 0 = normal ranges, male 32–44 (37), female 24–35 (29)
 1 = reduced ranges, male 28–35 (32), female 22–27 (24) (a reduction of approximately 14–17%), *H. hirsutus*

BIOGEOGRAPHICAL-based characters (16–19)

CHAR 16: Sympatric/Allopatric/Parapatric Distribution:
 0 = Sympatric, *H. pinteri* (with *H. concolor* and, to a limited degree, *H. arizonensis*)
 1 = Allopatric/parapatric (by areas), all other species

CHAR 17: General allopatric areas, DISJUNCT:
 0 = Baja area: *H. pinteri, H. concolor* and *H. hirsutus*
 1 = United States area: *H. arizonensis, H. spadix* and *H. obscurus*

CHAR 18: Specific parapatric subareas, CONNECTED:
 0 = inapplicable, *H. pinteri*
 1 = Baja area, Baja Sur subarea: *H. concolor*
 2 = Baja area, Cape Region subarea: *H. hirsutus*
 3 = United States area, CA–AZ subarea: *H. arizonensis*
 4 = United States area, CA–NV subarea: *H. spadix* and *H. obscurus*

CHAR 19: Specific parapatric microareas, CONNECTED:
 0 = inapplicable, "hirsutus" subgroup
 1 = CA–AZ subarea, California microarea: *H. a. pallidus*
 2 = CA–AZ subarea, Arizona microarea: *H. a. arizonensis*
 3 = CA–NV subarea, California microarea: *H. obscurus*
 4 = CA–NV subarea, Nevada microarea: *H. spadix*

Appendix 2. Data matrix: morphology and biogeography character mappings

```
Partial PAUP* input:
begin taxa ;
   dimensions ntax = 7 ;
   taxlabels H.pinteri
      H.concolor H.hirsutus
      H.spadix H.obscurus
      H.a.arizonensis H.a.pallidus
      ;
end ;
begin characters ;
   dimensions nchar = 20 ;
   format symbols = "01234" ;
   matrix
[                           1111111111
                   01234567890123456789 ]
   H.pinteri       00000000000000000000
   H.concolor      10111001010011001010
   H.hirsutus      10111001010011111020
   H.spadix        11222210020110101144
   H.obscurus      11222220020210101143
   H.a.arizonensis 11201101121010101132
   H.a.pallidus    11201101122010101131
   ;
end ;
outgroup H.pinteri H.concolor H.hirsutus ; [the "hirsutus" subgroup]
```

Appendix 3. POY input sequence, morphology data matrix and stepmatrix

```
POY input line
    poypc -noleading -norandomizeoutgroup -molecularmatrix step.311 -maxtrees 20
            -multibuild 20 -seed -1 -slop 2 -checkslop 5 had_fin.gen matrix.prt > poy_out
Note: file "had_fin.gen" contains GENBANK formatted DNA sequences stripped of indels:
```

Hennig86 compatible data matrix (file "matrix.prt")

```
Note: characters 10 and 19 are altered from original morphology matrix, suppressing
distinction between dark and pale forms of H. arizonensis.
    "Hadrurus modified"
    20 19
    Hpin_D_BS1   00000000000000000000
    Hcon_P_BN1   10111001010011001010
    Hhir_D_BS1   10111001010011111020
    Hspa_D_NV1   11222210020110101144
    Hobs_D_CA1   11222220020210101143
    Hobs_D_CA2   11222220020210101143
    Hobs_P_CA3   11222220020210101143
    Hobs_P_CA4   11222220020210101143
    Hari_D_CA1   11201101121010101132
    Hari_D_AZ2   11201101121010101132
    Hari_D_AZ3   11201101121010101132
    Hari_P_CA4   11201101121010101132
    Hari_P_CA5   11201101121010101132
    Hari_P_CA6   11201101121010101132
    Hari_P_CA7   11201101121010101132
    Hari_D_CA8   11201101121010101132
    Hari_D_AZ9   11201101121010101132
    Hari_P_CA10  11201101121010101132
    Hari_D_CA11  11201101121010101132
    ;
    cc -0.19 ; proc / ;
```

STEPMATRIX specifying 3:1:1 weights to transversions, transitions, and indels (file "step.311")

```
    0 3 1 3 1
    3 0 3 1 1
    1 3 0 3 1
    3 1 3 0 1
    1 1 1 1 0
```