An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: A case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae)

Chris A. Hamilton [a,*], Brent E. Hendrixson [b], Michael S. Brewer [c], Jason E. Bond [a]

[a] *Auburn University Museum of Natural History, Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA*
[b] *Department of Biology, Millsaps College, Jackson, MS 39210, USA*
[c] *Division of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA*

ABSTRACT

The North American tarantula genus *Aphonopelma* provides one of the greatest challenges to species delimitation and downstream identification in spiders because traditional morphological characters appear ineffective for evaluating limits of intra- and interspecific variation in the group. We evaluated the efficacy of numerous molecular-based approaches to species delimitation within *Aphonopelma* based upon the most extensive sampling of theraphosids to date, while also investigating the sensitivity of randomized taxon sampling on the reproducibility of species boundaries. Mitochondrial DNA (cytochrome *c* oxidase subunit I) sequences were sampled from 682 specimens spanning the genetic, taxonomic, and geographic breadth of the genus within the United States. The effects of random taxon sampling compared traditional Neighbor-Joining with three modern quantitative species delimitation approaches (ABGD, P ID(Liberal), and GMYC). Our findings reveal remarkable consistency and congruence across various approaches and sampling regimes, while highlighting highly divergent outcomes in GMYC. Our investigation allowed us to integrate methodologies into an efficient, consistent, and more effective general methodological workflow for estimating species boundaries within the mygalomorph spider genus *Aphonopelma*. Taken alone, these approaches are not particularly useful – especially in the absence of prior knowledge of the focal taxa. Only through the incorporation of multiple lines of evidence, employed in a hypothesis-testing framework, can the identification and delimitation of confident species boundaries be determined. A key point in studying closely related species, and perhaps one of the most important aspects of DNA barcoding, is to combine a sampling strategy that broadly identifies the extent of genetic diversity across the distributions of the species of interest and incorporates previous knowledge into the "species equation" (morphology, molecules, and natural history).

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The ability to organize biodiversity into recognizable clusters, contemporaneously referred to as species, is part of the human condition – from early humans (Chippindale and Taçon (1998) and Guthrie (2005)) to Plato and Aristotle (Wilkins (2009)), and has been extended over the past two and a half centuries to accommodate basic biological knowledge that includes evolutionary and ecological data. Species are a fundamental component of any biological investigation, and yet as Dobzhansky noted (1976), remain one of biology's oldest and most "vexing" problems. Incorrect assumptions regarding what represent natural entities we view as species, whether through the practice of "bad taxonomy" (see Bortolus, 2008) or via the presence of cryptic species, has important consequences to our understanding of evolutionary theory, ecological processes, biodiversity estimates, biogeographical history and patterns, species conservation and management decisions, and even human health (as discussed in Bickford et al. (2007). With modern day extinction rates estimated at 1000–10,000 times higher than the background rate (Barnosky et al., 2011), effective approaches are desperately needed to accelerate species discovery and identification. New species discovery and associated diversity (e.g., molecular, morphological, ecological) is paramount to understanding evolutionary pattern and process.

* Corresponding author. Address: Auburn University Museum of Natural History, Department of Biological Sciences, 331 Funchess Hall, Auburn University, Auburn, AL 36849, USA.

E-mail addresses: chamilton@auburn.edu (C.A. Hamilton), hendrb@millsaps.edu (B.E. Hendrixson), michaelsbrewer@gmail.com (M.S. Brewer), jbond@auburn.edu (J.E. Bond).

Among spiders, species placed in the infraorder Mygalomorphae (see Bond et al., 2012) represent one of the more problematic taxonomic groups for reliable species delimitation. The group is ancient (Penney and Selden, 2011; Selden and Gall, 1992), morphologically homogenous (Bond and Hedin, 2006), and prone to genetic structuring at a microgeographical scale (first demonstrated by Bond et al., 2001); mygalomorph spiders possess life-history traits that markedly differ from their sister lineage (the infraorder Araneomorphae) and most other arthropod groups. They typically have limited dispersal abilities (although some groups are known to "balloon" (Coyle, 1983, 1985), display site and habitat fidelity, take a long time to reach sexual maturity (4–7 years), and have long life spans (15–30 years) (Bond et al., 2001; Hendrixson and Bond, 2005; Arnedo and Ferrández, 2007; Hendrixson and Bond, 2007; Starrett and Hedin, 2007; Stockman and Bond, 2007; Bond and Stockman, 2008; Cooper et al., 2011; Hedin and Carlson, 2011; Satler et al., 2011; Hedin et al., 2013). Due to this suite of unique life-history traits, these taxa are often vulnerable to stochastic processes and therefore provide ideal candidates for evolutionary, biogeographical, and conservation studies (Raven, 1980; Hedin and Bond, 2006; Hendrixson and Bond, 2007; Hamilton et al., 2011; Bond et al., 2012; Bond, 2012; Hendrixson et al., 2013; Opatova et al., 2013). Taken together these traits have necessitated more integrative approaches to species delimitation that employ molecular, geographic, and ecological data (Bond et al., 2006; Arnedo and Ferrández, 2007; Hendrixson and Bond, 2007; Starrett and Hedin, 2007; Stockman and Bond, 2007; Bond and Stockman, 2008; Cooper et al., 2011; Hamilton et al., 2011; Satler et al., 2011, 2013; Hedin et al., 2013; Hendrixson et al., 2013) rather than any single character system taken alone.

1.1. Aphonopelma Pocock, 1901

The family Theraphosidae (tarantulas, baboon spiders, earth tigers) is the most diverse lineage (Platnick, 2013) within the infraorder Mygalomorphae (Raven, 1985; Hedin and Bond, 2006; Bond et al., 2012). The tarantula genus *Aphonopelma* is distributed throughout the southern third of the United States, ranging west of the Mississippi River to California and south through Mexico and into Central America. There are presently *54* nominal species in the United States (Platnick, 2013) that are thought to have rapidly diversified following expansion into the American Southwest ~5 Ma (Hamilton et al., 2011). But despite their academic appeal (in large part due to their apparent diversity and charismatic nature), the systematics and taxonomy of *Aphonopelma* remain problematic. During the past 75 years, only four major descriptive or revisionary works (Chamberlin and Ivie, 1939; Chamberlin, 1940; Smith, 1994; Prentice, 1997) have evaluated the taxonomy of *Aphonopelma*, but none of these studies employed an explicit phylogenetic approach to delimit species or to understand evolutionary relationships. The latter is fundamental to addressing important questions regarding the role that biogeography, allopatry, ecological divergence, and ancestral interactions have played in the diversification of these lineages.

Morphology-based phylogenies of mygalomorph spiders reveal widespread patterns of homoplasy among traditional taxonomic characters (Raven, 1985; Goloboff, 1993; Bond and Opell, 2002; Hedin and Bond, 2006; Bond and Hedin, 2006; Hendrixson and Bond, 2009; Bond et al., 2012). Furthermore, the quantitative or meristic features often used to evaluate relationships among these taxa may be problematic (Bond and Beamer, 2006; Hendrixson and Bond, 2009; but see Goloboff et al., 2006). Generally, morphological approaches to species delimitation in groups similar to mygalomorphs have grossly oversimplified and underestimated diversity (Locke et al., 2010; Niemiller et al., 2011). Much of the past theraphosid descriptive work was frequently based on only one to a few

specimens, generally lacking consideration of the wide range of intraspecific and intrasexual variation noted within the group (Prentice, 1997). Structure and variation of male and female genitalia has been a heavily weighted character in delimitation of spider species, but is of limited use in *Aphonopelma* due to morphological homogeneity across the US species and may only be useful for higher-level taxonomic groups (Prentice, 1997). Male mating claspers – modifications on the first two pairs of legs in adult male mygalomorph spiders used in holding and stimulating females during copulation – have also been effective at delimiting species of mygalomorph spider (e.g., Bond, 2012), yet these also appear homogeneous across *Aphonopelma*. As a consequence of all of these factors taken together we believe that a history of overzealous taxonomy has resulted in an over-description of *Aphonopelma* species within the United States. Not surprisingly, many arachnologists have expressed dismay towards the present state of theraphosid taxonomy (Raven, 1985; Smith, 1994; Pérez-Miles et al., 1996; Prentice, 1997), with Raven (1990) declaring the group a "nomenclatural and taxonomic nightmare".

1.2. DNA barcoding

Since the advent of PCR, the increased use of molecular information in systematic studies has been instrumental in uncovering tremendous evolutionary diversity previously unrecognized (cryptic species) using traditional approaches (e.g., morphology) (Hedin, 1997; Bond et al., 2001; Hebert et al., 2004a; Bickford et al., 2007; Bond and Stockman, 2008). DNA barcoding (Hebert et al., 2003a, 2003b) was proposed a decade ago as a means for quickly aiding species discovery and identification, wherein a single gene region from the animal mitochondrion could be employed for making species-level identifications or revealing cryptic diversity within lineages. By sequencing a fragment of cytochrome *c* oxidase subunit I (*CO1*), investigators can take advantage of the protein-coding gene's putative conserved nature while capitalizing on faster evolving 'silent' substitutions in the third codon position. As a consequence of inherent degeneracy in the genetic code, these third positions contain species-level information while limiting signal obfuscation due to saturation (Simon et al., 1994; Folmer et al., 1994).

DNA barcoding has its fair share of proponents and detractors (Lipscomb et al., 2003; Seberg et al., 2003; Tautz et al., 2003; Hebert et al., 2003a, 2003b, 2004a, 2004b; Will and Rubinoff, 2004; Barrett and Hebert, 2005; Hebert and Gregory, 2005; Will et al., 2005; Brower, 2006; Meier et al., 2006) with regards to its universal applicability and its ability to accurately discriminate among species. Hebert et al. (2003a, 2003b) suggest traditional taxonomy in morphologically conserved groups can lead to incorrect identifications and may fail to recognize cryptic taxa when morphological characters are uninformative or conflict with each other. Whereas Meyer and Paulay (2005) highlight how DNA barcoding can be useful for identification of species that belong to thoroughly sampled and well-understood groups, but recognize that delimitation of closely-related species in taxonomically understudied groups is problematic. The highly variable results that DNA barcoding produces across the Tree of Life (Brower, 2006; Meier et al., 2006; Astrin et al., 2006; Huber and Astrin, 2009; Bergsten et al., 2012) emphasizes the need for comprehensive and integrative approaches to identifying and delimiting species.

The DNA barcode has shown to be useful in separating and identifying species of spider from across all spiders (Barrett and Hebert, 2005; Arnedo and Ferrández, 2007; Longhorn et al., 2007; Petersen et al., 2007; Robinson et al., 2009; Kuntner and Agnarsson, 2011; Hendrixson et al., 2013). Hamilton et al. (2011) employed molecular characters on a smaller subset of

the North American *Aphonopelma* species, in an attempt to abrogate known problems in morphological-based taxonomy. This approach established an effective 'barcode gap' at 6% that distinguished clearly identifiable morphological groups and recognized putative cryptic species lineages within the genus. In addition to the genetic distance criterion, their study also emphasized consideration of phylogenetic placement as further evidence that reciprocal monophyly indicated strong support for a lack of gene flow among lineages.

As mentioned above, the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae) provides one of the greatest challenges to species delimitation and downstream identification in spiders because traditional morphological characters appear ineffective for evaluating limits of intra- and interspecific variation in the group (Prentice, 1997). Consequently, approaches that take advantage of data derived from sources other than morphology may provide more reliable methods for delimiting species in these tarantulas. The objectives of this study are threefold: (1) to evaluate the efficacy of molecular-based species delimitation methods on the identification of known and unknown *Aphonopelma* in the United States; (2) to test the sensitivity of random taxon sampling on the reproducibility of species boundaries; and (3) to integrate methodologies in a way such that an efficient, consistent, and more effective DNA barcoding strategy can be employed for delimiting species of *Aphonopelma* in the United States.

2. Methods

2.1. Taxon sampling and data collection

To our knowledge, the research presented herein is derived from the most focused and comprehensive sampling of a single theraphosid genus to date. Through our own extensive fieldwork and a citizen-based science program (in association with the American Tarantula Society, see http://www.atshq.org/articles/found.html), we have accumulated more than 1600 recently collected specimens of *Aphonopelma* from throughout their distribution in the southwestern United States (i.e., every state they are native). Sampling for this analysis comprises putative species that include: (1) numerous individuals sampled from multiple populations across a species' entire distribution (whether widespread or highly localized); (2) "singleton" species (i.e., species known from only a single specimen); and (3) "unique" species (i.e., species known from only a single sampling locality – the terms "singleton"

and "unique" are defined in Lim et al. (2011)). In total, we sequenced mtDNA from 682 of these specimens for the animal barcoding gene cytochrome *c* oxidase subunit I (*CO1*) (Fig. 1). The vast majority of specimens used for this study were opportunistically collected throughout the southwestern United States, but we also made every attempt to gather "topotypic" material from (or near) the type localities of all 54 species of *Aphonopelma* currently recognized in the United States (Appendix A). Of these, we were unable to obtain fresh material for only two of the targeted species. *Aphonopelma phasmus* Chamberlin, 1940 is known from a single adult male taken near Phantom Ranch at the "base" of the Grand Canyon. This locality is difficult to access and no attempts were made to collect the species. *Aphonopelma radinum* (Chamberlin and Ivie, 1939) is likewise known only from a single male, but was collected near Manhattan Beach, California. We have reason to believe that the type locality was mislabeled (see Prentice, 1997) but the species also may have alternatively been extirpated from the area. Manhattan Beach is a highly developed coastal suburb of Los Angeles and the habitat does not appear conducive for supporting tarantula populations; populations of other ground-dwelling mygalomorph spiders in the Los Angeles Basin likely have become extinct due to urbanization (see Bond et al. 2006). Numerous field expeditions into southern California by the authors and others (Thomas Prentice, personal communication) have failed to locate this species. Specimens from the type localities of four previously synonymized species were also sampled in order to evaluate whether nomenclatural changes made in Prentice (1997) were warranted. All material was preserved in 80% ethanol and assigned a unique voucher number (APH0000). Specimens will be deposited in the California Academy of Sciences, American Museum of Natural History, and Auburn University Museum of Natural History collections.

2.2. Molecular protocols and alignment

Tissue samples were collected from specimens by removing the third leg on the right side of the spider followed by preservation in 100% ethanol or RNA*later*™ (Qiagen, Valencia, CA, USA) and storage at −80 °C. Muscle tissue was extracted from the leg by removing ~25 mg of tissue and genomic DNA extracted using the Qiagen DNeasy Tissue Kit™ (Qiagen, Valencia, CA, USA). The concentration quality of the extracted DNA was quantified with a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, Wilmington, DE, USA) or visualized via agarose gel electrophoresis.
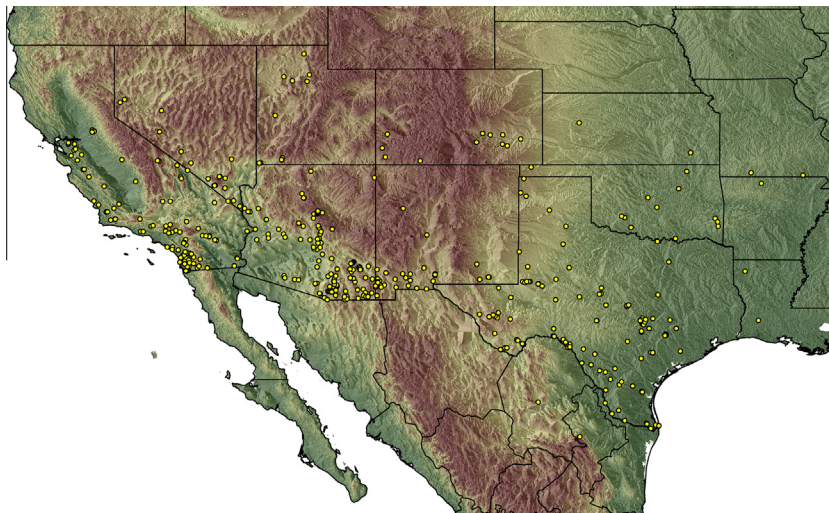

Fig. 1. General distribution map displaying the breadth of *Aphonopelma* haplotype sampling across the United States.

PCR and direct sequencing primers used for the *CO1* barcoding fragment are listed in Hamilton et al. (2011). PCR protocol followed initial denaturation at 95 °C for 2 min; 30 cycles of denaturation at 95 °C for 45 s, annealing at 48 °C for 45 s, elongation at 72 °C for 1 min; followed by 5 min of a final elongation at 72 °C. The 5′ primer LCO1490 (Folmer et al., 1994) and its derivations were tested for uniformity across *Aphonopelma*. Initial amplification and sequencing were generally successful across the group, but due to mutations in the LCO1490 binding site some primer modifications became necessary (see Table 1 in Hamilton et al., 2011). The primer C1-J-1751 "SPID" (Hedin and Maddison, 2001) provided the most consistent amplification and high-quality sequencing. PCR products were purified using ExoSAP-IT (USB Corporation; Cleveland, OH, USA) and then sequenced with an ABI 3130 Genetic Analyzer (Applied Bio-systems, Foster City, CA, USA) using the ABI Big Dye Terminator version 3.2 Cycle Sequencing Ready Reaction Kit.

All sequences were manually edited using the program Sequencher (ver. 4.1.2, Genecodes, Madison, WI, USA). Sequences were aligned with MUSCLE version 3.6 (Edgar, 2004) using default parameters, followed by minor adjustment in MESQUITE version 2.73 (Maddison and Maddison, 2011) if needed. Amino acid translations of the target gene region were examined to ensure the absence of stop codons in the alignment. The alignments were unambiguous and for consistency, sequences were trimmed to 900 bp. All *CO1* sequences have been deposited in GenBank (Appendix B), the full DNA alignment and associated phylogenetic tree have been deposited in TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S13957), and all phylogenetic data matrices, accompanying tree files, and scripts have been deposited in figshare (http://dx.doi.org/10.6084/m9.figshare.769358).

2.3. Traditional DNA barcoding

Classic DNA barcoding (Hebert et al., 2003a) calculates a genetic distance between specimens using Kimura's 2-parameter distance (K2P or K80) (Kimura, 1980) and assigns a cutoff value (the 'barcode gap') to divide OTUs into species. Uncorrected genetic distances (uncorrected *p*-distance) have also been used to define this cutoff; Srivathsan and Meier (2011) find that the use of K2P is inappropriate when employing it for closely related taxa. In order to compare the variability of each method on an extensively sampled dataset, the intra/interspecific variation of 682 specimens (representing the 54 nominal species) was assessed. Both K2P and uncorrected distances were investigated; little difference was seen, therefore we chose to follow the recommendation of Srivathsan and Meier (2011) and use uncorrected distances. An initial 'barcode gap' of 6% (Hamilton et al., 2011) was applied to the dataset in order to evaluate effectiveness and reliability across a broader evolutionary scale. Mega 5 (Tamura et al., 2011) was used to group OTUs into putative species groups and measure the mean intra/interspecific distances for each species hypothesis. This evaluation was unable to consistently split all species we had previously identified as putative species based upon morphology or biogeography. We reevaluated this cutoff by incorporating previous morphological knowledge, adding biogeographical and/or behavioral information, and the species groupings from the other methodologies (below) to establish new species hypotheses and search for a universal 'barcode gap'.

Neighbor-Joining trees were inferred using uncorrected genetic distances for the full and haplotype OTU datasets. The identified 'barcode gap' cutoff value (see below) was then applied across the trees to determine the number of species. Putative species groups were named based upon two classifications: specimens resided within a group/clade also holding topotypic localities of nominal species, or specimens possessed a defining character unique to that clade (geography, morphology, or cryptic species).

2.4. Automatic Barcode Gap Discovery

The Automatic Barcode Gap Discovery (ABGD) method (Puillandre et al., 2012) quantitatively evaluates intraspecific divergence by calculating all pairwise distances within a dataset and ordering them as ranked values. A sliding window is used to calculate a local slope function (at one-tenth its starting value) across these values; this is used to identify the first statistically significant peak where the 'barcode gap' represents a sudden increase in slope. The dataset is then recursively repartitioned into finer and finer groupings until no further gaps can be detected. A range of differing parameters was evaluated (including the defaults of Pmin = 0.001 and Pmax = 0.10); no differences in species delimitations were seen. We chose the parameters: Pmin = 0.0001, Pmax = 0.200, Steps = 10, X = 1, Nb bins = 20, due to a slightly higher *p*-value significance for these delimited groups (*p* = 0.0158) than the defaults (*p* = 0.0282). ABGD analyzes data through a web-based interface (http://www.abi.snv.jussieu.fr/public/abgd/) or through a Unix command-line version – our analyses used the latter. A large number of variable parameters were evaluated for their effects on species delimitation outcomes, with very little variation seen. ABGD was recently applied in a smaller subset of US *Aphonopelma* (Hendrixson et al., 2013); their determinations agree with our current stance of species boundaries within the *mojave* group.

2.5. Tree-based species delimitation

To identify and evaluate species hypotheses, a Maximum Likelihood tree was inferred by employing RAxML-7.2.8 (Stamatakis, 2006) on both the full dataset of 682 OTUs and the 358 OTU haplotype dataset. Due to the size of the datasets, parameters for the analyses incorporated the GTRCAT model of nucleotide site substitution based on 1000 random addition sequence replicates (RAS); branch support values were computed via 1000 non-parametric bootstrap replicates. The full and haplotype datasets were partitioned by codon, though a single partition was applied for all of the GMYC subsampling schemes (below). Aside from this particular study, we have evaluated the effects of a partitioned versus non-partitioned *Aphonopelma* dataset on the change to topology and node support; species clades remain the same and support varies little between the two approaches on this particular DNA dataset. Initial haplotypes were designated by RAxML, and later compared to TCS 1.21 designations (Clement et al., 2000) – where more taxa were collapsed due to ambiguous calls.

We employed a version of Wiens and Penkrot (2002), also implemented in Hamilton et al. (2011), to delineate species by producing a phylogenetic tree and evaluating the amount of lineage isolation – assuming no gene flow occurs or can occur between species (based on the absence of shared haplotypes between populations). This W&P method identifies divergent monophyletic clusters and uses previous morphological knowledge to identify known species on the tree. An *Aphonopelma* specimen (APH3022) from Compostela in the Nayarit state of Mexico (the type locality of *A. nayaritum*) was chosen as a divergent sister outgroup lineage.

2.6. P ID(Liberal) species boundary delimitation

To assess species boundary hypotheses across the ML gene tree, the Species Delimitation plugin (Masters et al., 2011) within Geneious Pro v5.5.4 (Biomatters; http://www.geneious.com) was investigated for effectiveness when compared to the traditional barcoding methods. This quantitative approach allows differing species boundary hypotheses to be investigated by enabling the

user to *a priori* assign taxa to putative species groups on a phylogenetic tree. As a way to measure species group distinctiveness, phylogenetic exclusivity – the probability of an unidentified specimen being correctly placed into an *a priori* species group, was calculated. This calculation evaluates the probability that membership within a clade arose by chance in a random coalescent process. Quantitative approaches to species delimitation are important when employing mtDNA data in a phylogenetic context because species boundaries will frequently be identified by deep divergences in the tree. Rigorous evaluations of species boundaries using a statistical assessment of diverging lineages is critical for limiting the under or over-splitting of species.

Cladistic structuring can lead to an incorrect perception of cryptic species, as long branches within a panmictic population can arise simply due to the stochastic nature of gene coalescence (Irwin, 2002; Kuo and Avise, 2005). Rodrigo et al. (2008) attempted to resolve this issue by devising a statistic to distinguish unique clades by measuring a ratio of the distance from a species-defining node to the tips of the tree, and the distance from that same node to its immediate ancestor – the P(Randomly Distinct) statistic (referred to as P ID(Strict) in Geneious). As a continuation of the Rodrigo et al. (2008) statistic, Ross et al. (2008) discovered that the ratio of intraspecific genetic difference to that of the nearest putative species group (Intra/Inter ratio) was a better predictor of species group identification than the traditional 'barcode gap'. This statistic, P ID(Liberal) in Geneious, represents the probability of making a correct identification of an unknown specimen by measuring the genetic variation found within its putative species group and comparing that to the species group with which it is most likely to be confused. P ID(Liberal) has also been shown to correctly identify taxa at a rate similar to using genetic distances or BLAST against a reference database (Masters et al., 2011).

The full dataset ML gene tree was used to assign the putative species groups, based on the presence of divergent, monophyletic clades (described above). Differing species boundary hypotheses were tested by collapsing/expanding clades and singletons with low P ID(Liberal) values into their monophyletic sister species' clades until a broadly fitting, highly supported, consistent pattern was found across the tree.

2.7. Taxon sampling effects on GMYC species delimitation

The Generalized Mixed Yule Coalescent (GMYC) model (Pons et al., 2006), as implemented in the R package "splits" (Ezard et al., 2009), is a species delimitation method that starts with an inferred gene tree rather than actual sequence data and attempts to statistically model (through maximum likelihood calculations) the point on a time calibrated (ultrametric) phylogeny where within species population-level processes of molecular evolution shift to between species coalescent processes. During a GMYC analysis, single (Pons et al., 2006) and multiple threshold models (evolutionary processes occur with differing rates across the tree, Monaghan et al., 2009) are applied and the best fitting model is chosen.

To test the effects of taxon sampling on the reproducibility of a single locus species delimitation method, we performed multiple iterations of random taxon jackknifing on the *COI* haplotype matrix (358 OTUs) using a novel Python script developed for this project (Appendix C) – GMYC does not allow investigations with redundant haplotypes. The script initially removes line breaks from the FASTA header lines only, resulting in a single line for each taxon comprising the header, a temporary placeholder, and the sequence. The outgroup taxon (APH3022) was then removed from the dataset before randomization and reinserted afterwards as the first sequence in the file. The remaining ingroup taxa lines were randomized using the Python function "random.shuffle". Nine different 'taxon inclusion' groups were defined as a subset of the original

haplotype dataset (at 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, and 10%). The OTUs (and subsequent sequence data) within these groups were extracted from the randomized dataset and the temporary placeholders were replaced with line breaks, creating a FASTA file. This process was performed 100 times for each 'taxon inclusion' group. These FASTA files were converted to PHYLIP files using the Fasta2Phylip.pl script (http://indra.mullins.microbiol.washington.edu/index.html; under 'Sequence manipulator'). Maximum Likelihood tree searches were performed on the 900 permutated datasets using the program RAxML 7.2.8 (Stamatakis, 2006) comprising 100 random addition sequence replicates (RAS). The resulting best trees for each of the 'taxon inclusion' groups were used in the subsequent GMYC analyses.

To facilitate the GMYC evaluation of large numbers of trees produced from the jackknifed data matrices, a simple R script was employed to automate the process across all 900 trees, as well as the 100% haplotype tree (Appendix D). A single tree was read into R using the "read.tree" command as implemented in the package "ape" (Paradis et al., 2004; Paradis, 2006). The tree was converted to ultrametric using the command "chronopl" and made fully dichotomous using the command "multi2di" to resolve the phylogeny using zero length branches. The GMYC analysis was performed using the command "gmyc", saving the summary for further analyses, and the next tree was loaded. This process was repeated for all the trees within each permutated dataset. Using each summary file, the number of 'ML clusters' and 'ML entities' were extracted from the data of each 'taxon inclusion group'. A 'cluster' represents an independently evolving clade – 'singletons' or exceptionally long branches are not counted; an 'entity' represents all clades and all singletons deemed as independently evolving – 'singletons' are considered a putative cluster given more population representatives.

2.8. Taxon sampling effects on Neighbor-Joining, ABGD, and P ID(Liberal)

To compare the GMYC results with those of the other methodologies, we randomly sampled 27 ML trees and their associated datasets (alignments) from each 'taxon inclusion group', choosing pairs corresponding to the lowest, a midpoint, and highest number of GMYC species delimited. Using the same methodology and cutoffs outlined above for each methodology, Neighbor-Joining trees were inferred from each of these 27 datasets, with the determined 'barcode gap' (see Section 3) applied across each tree, and number of species counted. In order to quantitatively test species designations, both the ABGD and the P ID(Liberal) statistic were applied across the same 27 sampled datasets and ML trees respectively.

3. Results

Two datasets using the animal barcoding gene, *cytochrome c oxidase subunit I* (CO1), tested species boundaries and identification of United States *Aphonopelma*: a 682 full specimen dataset and a 358 unique haplotype-only dataset. The full dataset contained 356 parsimony-informative sites, 454 identical sites (50.4%), 39 singleton sites, 89.3% pairwise identity, a nucleotide base composition of: A = 19.8%, C = 11.9%, G = 24.6%, T = 43.8%, and a GC content of 36.4%. The unique haplotype only dataset contained 350 parsimony-informative sites, 505 identical sites (56.1%), 45 singleton sites, 89.2% pairwise identity, a nucleotide base composition of: A = 19.7%, C = 11.9%, G = 24.6%, T = 43.8%, and a GC content of 36.8%.

3.1. DNA barcoding

Following application of the initial 6% 'barcode gap' species hypotheses and identification of its failure to consistently identify known species, groupings were reevaluated to investigate the applicability of a universal 'barcode gap' across the dataset. Clades were collapsed or split until a broadly fitting consistent pattern was found, using prior taxonomic knowledge and information unique to particular *Aphonopelma* lineages (i.e., biogeography, ecology, or behavior). A 'barcode gap' was identified at 5%, where intraspecific distances (⩽4.0/3.9% (K2P/uncorrected)) never equaled or exceeded the interspecific distances (⩾5.5/5.2% (K2P/uncorrected)), thereby delimiting 34 total species, 32 of which reside within the United States and 2 from Mexico (sp. *nayaritum* – the outgroup, and sp. "Coahuila MX" – a cryptic species originally identified as a member of *Aphonopelma moderatum* (Chamberlin and Ivie, 1939)) (Table 1; Fig. 2). This approach identified 16 nominal species, 7 newly discovered, and 9 cryptic species. When applying the 'barcode gap' across the full OTU and haplotype NJ trees, 37 and 35 species are delimited respectively (Fig. 2). Though it should be noted that within the full OTU inferred tree, three clades sit at the divergence cutoff with deep branch splits identified as separate species (based on a strict application of this cutoff criterion). Utilization of our prior knowledge of the specimens just outside these splits would lower this number to 34 total (32 in the US), consistent with the 'barcode gap' species hypotheses and the P ID(Liberal) approach (see below). The ABGD method delimited 46 species in both the full OTU and haplotype datasets. The groupings delimited by this method were consistent between the full and haplotype datasets as well as the 'barcode gap', tree-based, and P ID(Liberal) approaches, though ABGD appears to over-split certain lineages – based upon our knowledge (morphology, ecology, and biogeography) of these independent lineages. Species hypotheses that were over-split represent lineages where deep divergences occur between certain populations across their distributions (*hentzi*, *iodius*, *brunnius*, *eutylenum*, sp. *paloma* nov 1, sp. *mojave* nov E, sp. *mojave* nov W, and sp. nov G.

3.2. Tree-based delimitation

The W&P method was applied across the Maximum Likelihood phylogenetic inferences for both the full and haplotype datasets, delineating 41 species (39 US species plus two species from Mexico) (Fig. 2). These species hypotheses were based upon major clade monophyly and highly divergent clades that appear, based upon topological structuring, to be experiencing no present or very little recent gene flow (initial species hypotheses summarized in column 1 of Table 1). Topologies of both the full and haplotype datasets were largely congruent, identifying the same species clades. Like ABGD, species hypotheses that appear to be over-split represent lineages where deep divergences occur between certain populations (generally lineages that are highly divergent and sister to all other lineages in the clade) across their distributions.

3.3. P ID(Liberal) species boundary delimitation

Independent of the traditional barcoding approach, but following the same guidelines, the tree-based hypotheses were reevaluated for species hypothesis testing. All delimited species except one possess a P ID(Liberal) value ⩾0.93 (Table 1). *A.* sp. nov C, comprising only three specimens, possesses a P ID(Liberal) value of 0.81, and represents a morphologically distinct species from its sister species, sp. nov D. A cutoff was set where all delimited species would possess a P ID(Liberal) value >0.80 (an 80% probability of correctly placing an unknown specimen into its *a priori* designated species), revealing a pattern that mirrors the traditional barcoding

approach – 34 species delimited, 32 of which are found in the United States (Table 1; Fig. 2). P ID(Liberal) consistently identified the same 16 nominal species, 7 newly discovered, and 9 cryptic species as determined in the traditional barcoding approaches.

3.4. GMYC and taxon sampling effects

The performance of the single and multiple threshold models indicated the two models were not significantly different from each other; therefore the outcomes of the single threshold model were selected. After applying GMYC across the ML inferred tree from the 358 OTU haplotype dataset, 83 'clusters' and 114 'entities' were identified as independently evolving lineages (Fig. 2). Random taxon sampling provided an opportunity to evaluate the effects of collecting effort (i.e., which individuals or populations are sampled), as well as evolutionary history on the GMYC method's ability to consistently identify the independently evolving lineages we are identifying as species. Within and between species estimates exhibited considerable variance, with estimates ranging from all sampled OTUs represented as a single species (found in the 10%, 20%, and 30% datasets), to all or nearly all OTUs being delimited as separate species (e.g. 36 species in the 10% 'entities' dataset and 210 species in the 60% 'entities' dataset) (Table 2). The average number of species across all 100 replicates within a 'taxon inclusion' group ranged from 5.79 species (10% 'clusters') and 25.81 species (10% 'entities'), to 60.21 species (90% 'clusters') and 83.65 species (90% 'entities') (Table 2). Fig. 3a and b illustrates the frequency distributions for how many times a particular number of delimited species occurred within each 'taxon inclusion group'. Our regression analyses indicate the number of species delimited in each iteration was positively correlated with the actual number of OTUs being analyzed – 'entities' ($R^2 = 0.4013$; $p < 0.001$ in Fig. 4a) and 'clusters' ($R^2 = 0.7389$; $p < 0.001$ in Fig. 4b).

The GMYC results were compared to Neighbor-Joining inference, and the quantitative approaches ABGD and P ID(Liberal) by randomly sampling three alignments (or the associated tree) within each GMYC 'taxon inclusion group' (the lowest, a midpoint, and highest number of species delimited). Using each of those datasets, a NJ tree was inferred followed by the application of the 'barcode gap' (5%). NJ species numbers ranged from 19 (10% low) to 37 (70% high) (Table 3). The same ABGD methodology and parameters (outlined above) were employed on each sampled alignment. Scores, while higher, were generally consistent, showing an initial increase and leveling off as sampling is increased (Fig. 5). Across the random sampling, ABGD species numbers ranged from 20 (10% low) to 34 (30% low, 30% high, 40% low, 40% high), to 44 (30% medium, 80% high, 90% medium), with three extreme outliers (91 at 50% low; 100 at 60% medium; 83 at 70% high) (Table 3). Each associated ML tree had the P ID(Liberal) species boundary cutoff (⩾0.80) applied and number of species counted. P ID(Liberal) quantitative species trended similarly with NJ, ranging from 17 to 21 (low to high) across the 10% group, to 32 for all three 90% groups (Table 3). As can be seen in Fig. 5, the NJ and P ID(Liberal) approaches converge around the 34 delimited species estimated from our integrative methodological workflow and remain remarkably stable and consistent once 40% of the data has been included. The differences between GMYC (extreme variation) and the other methodologies (relative stability) are also visualized in Fig. 6.

4. Discussion

Molecular markers, like *CO1*, may possess effective species boundary information within certain taxonomic groups and consequently have the potential to be a rapid and efficient means to delineate and identify species. But how do we determine the best

Table 1
Summary of initial species hypotheses, species delimitations, and the various metrics used in the hypothesis-testing framework and methodological pathway.

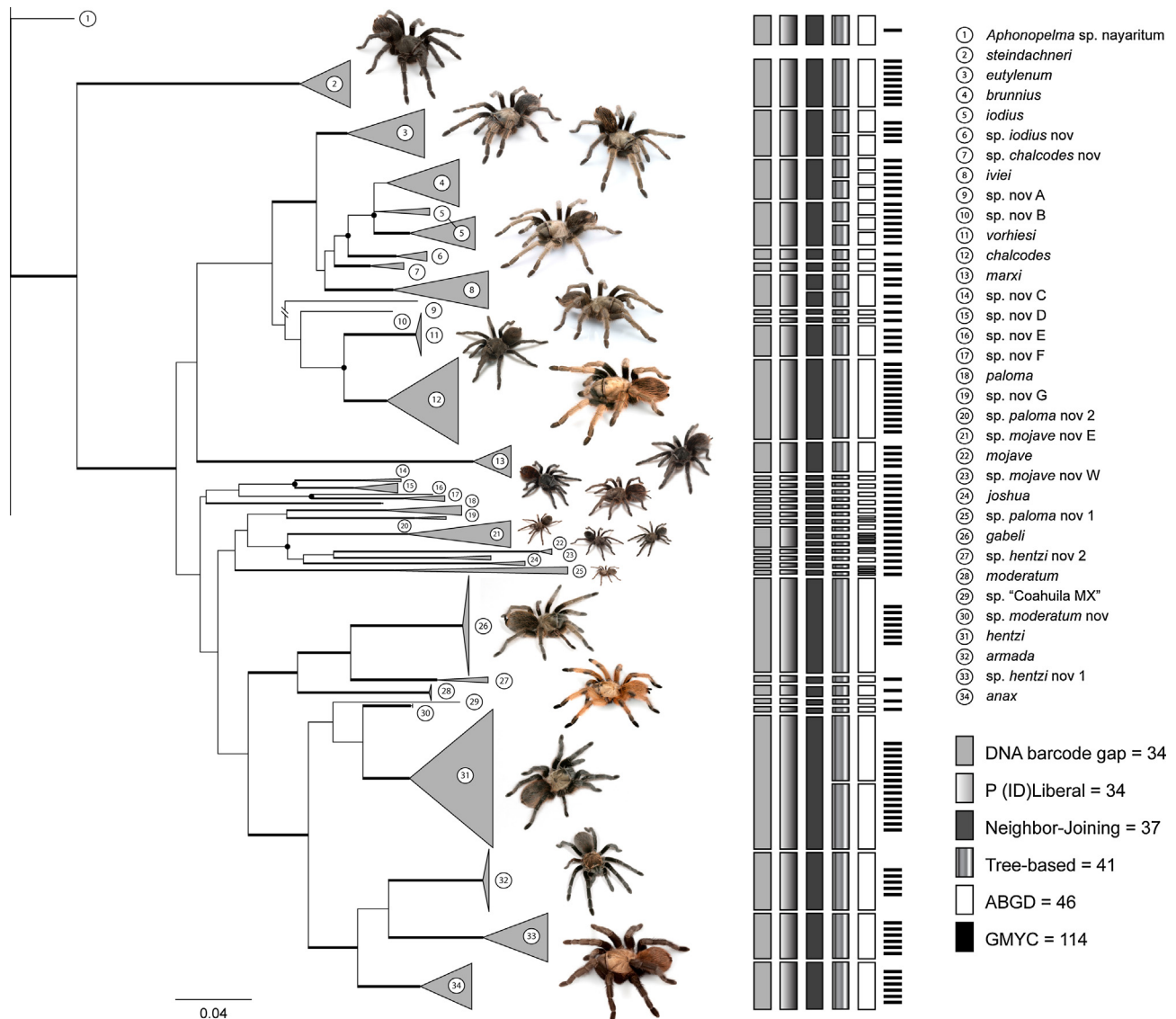| Initial tree-based species hypotheses (*n* = # of OTUs)	Barcode gap designated species (*n* = # of haplotypes)	ML species bootstrap support (full/haplotype)	Intraspecific *p*-dist. – K2P (uncorr.)	Closest interspecific *p*-dist. – K2P (uncorr.)	P ID(Liberal) – prob of correct id (CI)	Closest P ID(Liberal) species
anax (*n* = 34)	*anax* (*n* = 17)	100/100	0.019 (0.018)	0.08 (0.075) – sp. *hentzi* nov 1	0.97 (0.95, 1.0)	*armada*
armada (*n* = 46)	*armada* (*n* = 14)	100/100	0.002 (0.002)	0.083 (0.078) – sp. *hentzi* nov 1	1.00 (0.98, 1.0)	*anax*
brunnius (*n* = 35)	*brunnius* (*n* = 16)	47/53 – (clade 1 = 100/100; clade 2 = 24/47)	0.03 (0.029)	0.055 (0.052) – *iodius*	0.98 (0.95, 1.0)	*iodius*
chalcodes (*n* = 63)	*chalcodes* (*n* = 42)	95/93	0.019 (0.018)	0.056 (0.054) – *vorhiesi*	0.97 (0.94, 1.00)	*vorhiesi*
eutylenum (*n* = 36)	*eutylenum* (*n* = **18**)	99/98	0.028 (0.027)	0.054 (0.052) – sp. *iodius* nov	0.97 (0.94, 1.00)	sp. *iodius* nov
gabeli (*n* = 73)	*gabeli* (*n* = 15)	100/100	0.002 (0.002)	0.078 (0.073) – sp. *hentzi* nov 2	1.00 (0.98, 1.0)	sp. *hentzi* nov 2
hentzi (*n* = 100)	*hentzi* (*n* = 46)	100/100	0.017 (0.016)	0.056 (0.054) – sp. *moderatum* nov	0.98 (0.95, 1.0)	sp. *moderatum* nov
iodius (*n* = 31)	*iodius* (*n* = 18)	2 clades (clade 1 = 99/99; clade 2 = 27/28)	0.031 (0.03)	0.055 (0.052) – *brunnius*	0.96 (0.93, 0.98)	*brunnius*
iviei (*n* = 28)	*iviei* (*n* = 18)	100/100	0.031 (0.03)	0.065 (0.062) – sp. *iodius* nov	0.97 (0.94, 0.99)	sp. *iodius* nov
joshua (*n* = 4)	*joshua* (*n* = 3)	100/100	0.024 (0.024)	0.1 (0.098) – sp. nov C	0.96 (0.85, 1.0)	sp. *mojave* nov W
marxi (*n* = 24)	*marxi* (*n* = 16)	100/100	0.014 (0.013)	0.111 (0.103) – *chalcodes*; 0.112 (0.103) sp. nov A	0.99 (0.96, 1.0)	sp. nov B
moderatum (*n* = 13)	*moderatum* (*n* = 2)	100/100	<0.001 (<0.001)	0.085 (0.08) – sp. *hentzi* nov 2	1.00 (0.96, 1.0)	sp. *hentzi* nov 2
mojave (*n* = 5)	*mojave* (*n* = 2)	100/100	0.007 (0.007)	0.107 (0.099) – sp. nov D; 0.111 (0.102) – sp. *mojave* nov W	0.98 (0.88, 1.0)	sp. *mojave* nov W
paloma (*n* = 2)	*paloma* (*n* = 1)	100/n/a	–	0.091 (0.086) – sp. *paloma* nov 1; 0.095 (0.089) – sp. nov D	0.98 (0.83, 1.0)	sp. nov C
steindachneri (*n* = 35)	*steindachneri* (*n* = 27)	100/100	0.016 (0.016)	0.101 (0.093) – *vorhiesi*; 0.105 (0.097) – sp. *nayaritum*	0.99 (0.96, 1.0)	sp. *nayaritum*
vorhiesi (*n* = 32)	*vorhiesi* (*n* = 13)	100/100	0.002 (0.002)	0.056 (0.054) – *chalcodes*	1.00 (0.98, 1.0)	*chalcodes*
sp. "Big Bend TX"	Included into *hentzi*	–	–	–	–	–
sp. Coahuila MX (*n* = 1)	sp. Coahuila MX (*n* = 1)	–	–	0.064 (0.061) – sp. *moderatum* nov	0.96 (0.83, 1.0)	sp. *moderatum* nov
sp. "iod-cha CANVAZ"	Included into *iodius*	–	–	–	–	–
sp. "NE CA"	Included into *iviei*	–	–	–	–	–
sp. "new dwarf AZ"	Included into sp. *paloma* nov 1	–	–	–	–	–
sp. "new dwarf NM"	Included into sp. nov G	–	–	–	–	–
sp. "SE CA 2"	Included into *brunnius*	–	–	–	–	–
sp. "SE CA"	Included into *eutylenum*	–	–	–	–	–
sp. *chalcodes* nov (*n* = 8)	sp. *chalcodes* nov (*n* = 8)	100/100	0.018 (0.017)	0.06 (0.057) – sp. *iodius* nov	0.93 (0.87, 1.00)	sp. *iodius* nov
sp. *hentzi* nov 1 (*n* = 36)	sp. *hentzi* nov 1 (*n* = 18)	100/100	0.015 (0.014)	0.08 (0.075) – *anax*	0.98 (0.96, 1.0)	*armada*
sp. *hentzi* nov 2 (*n* = 5)	sp. *hentzi* nov 2 (*n* = 2)	100/100	0.028 (0.028)	0.078 (0.073) – *gabeli*	0.97 (0.86, 1.0)	*gabeli*
sp. *iodius* nov (*n* = 6)	sp. *iodius* nov (*n* = 10)	90/89	0.016 (0.016)	0.054 (0.052) – *iodius* & *eutylenum*	0.96 (0.85, 1.0)	sp. *chalcodes* nov
sp. *moderatum* nov (*n* = 4)	sp. *moderatum* nov (*n* = 1)	100/n/a	–	0.056 (0.054) – *hentzi*	0.98 (0.87, 1.0)	*hentzi*
sp. *mojave* nov E (*n* = 20)	sp. *mojave* nov E (*n* = 12)	100/100	0.04 (0.039)	0.103 (0.095) – sp. nov D & sp. nov C	0.97 (0.95, 1.0)	sp. *mojave* nov W
sp. *mojave* nov W (*n* = 4)	sp. *mojave* nov W (*n* = 4)	100/100	0.038 (0.037)	0.11 (0.101) – sp. *mojave* nov E; 0.11 (0.102) – *paloma*; 0.111 (0.102) – *mojave*	0.93 (0.82, 1.0)	*mojave*
sp. *nayaritum* (*n* = 1)	sp. *nayaritum* (*n* = 1)	–	–	0.09 (0.085) – *chalcodes*	0.96 (0.83, 1.0)	*steindachneri*
sp. nov A (*n* = 1)	sp. nov A (*n* = 1)	–	–	0.102 (0.095) – sp. *moderatum* nov; 0.103 (0.096) – sp. nov G	0.96 (0.83, 1.0)	sp. nov B
sp. nov B (*n* = 1)	sp. nov B (*n* = 1)	–	–	0.072 (0.068) – *chalcodes*	0.96 (0.83, 1.0)	*chalcodes*
sp. nov C (*n* = 3)	sp. nov C (*n* = 3)	100/100	0.028 (0.028)	0.06 (0.057) – sp. nov D	**0.81 (0.67, 0.96)**	sp. nov D
sp. nov D (*n* = 8)	sp. nov D (*n* = 8)	100/100	0.019 (0.018)	0.06 (0.057) – sp. nov C	0.94 (0.87, 1.0)	sp. nov C
sp. nov E (*n* = 1)	sp. nov E (*n* = 1)	–	–	0.087 (0.081) – sp. nov F	0.96 (0.83, 1.0)	sp. nov F
sp. nov F (*n* = 5)	sp. nov F (*n* = 5)	100/100	0.014 (0.013)	0.087 (0.081) – sp. nov E	0.96 (0.86, 1.0)	sp. nov E
sp. nov G (*n* = 8)	sp. nov G (*n* = 5)	100/100	0.031 (0.029)	0.098 (0.09) – sp. *paloma* nov 2	0.93 (0.87, 1.00)	sp. *paloma* nov 2
sp. *paloma* nov 1 (*n* = 6)	sp. *paloma* nov 1 (*n* = 6)	100/100	0.04 (0.038)	0.098 (0.091) – sp. nov G & *anax*	0.94 (0.84, 1.0)	sp. *paloma* nov 2
sp. *paloma* nov 2 (*n* = 3)	sp. *paloma* nov 2 (*n* = 3)	100/100	0.019 (0.018)	0.098 (0.09) – sp. nov G; 0.097 (0.09) – sp. *nayaritum*	0.93 (0.78, 1.0)	sp. nov G
Total species = 41	Total species = 34				Total species = 34	

Fig. 2. Maximum Likelihood inferred *CO1* gene tree for the full 682 specimen dataset of United States *Aphonopelma*. Species delimitations are based on the integrative methodological approach. The horizontal bars illustrate the species delimitations for each individual method. All 32 of the redefined US species were reciprocally monophyletic and possessing high ML bootstrap support (⩾90), with two exceptions – the paraphyletic *A. iodius* grouping (5) and its weakly monophyletic sister species *A. brunnius* (4). Thickened branches indicate nodal support with ML bootstrap values ⩾90%. Support for major nodes with ML bootstrap values >80% and <90% are denoted by a black dot. Species names are indicated with their associated number. Not all 32 species are represented in the photographs. Image size of each spider is relative to the other species that are represented.

approach for discovering those boundaries? And how sensitive to sampling is a particular approach? In this study, we have investigated how single-marker molecular species delimitation approaches fare in their ability to converge on the same answer within the focal taxon *Aphonopelma*. Our work integrates methodologies into a hypothesis-testing framework that proposes and evaluates the number of both known and unknown species of tarantula within the United States. Using prior taxonomic knowledge of this group, six independent investigations were carried out to propose and test species hypotheses: traditional DNA barcoding ('barcode gap' & NJ), modern statistical DNA barcoding (ABGD), a tree-based phylogenetic approach, and quantitative boundary delimitation approaches P ID(Liberal) & GMYC.

Based on an integrative summary of those methods deemed effective, our approach to species delimitation indicates that a robust estimate of species diversity is likely to total 34 species given our current sampling regime (Fig. 2). These species include 16

previously recognized nominal species, 7 putative newly discovered, and 9 previously unrecognized species (i.e. cryptic) (see species hypotheses in Table 1 and Fig. 2). Our findings indicate that out of the 54 nominal species of US *Aphonopelma* recognized prior to this research endeavor, 30 appear to be in need of synonymy, with six likely *nomen dubia*, and two nominal species unsampled (*A. phasmus* and *A. radinum*, mentioned above) (Appendix A). The details of the phylogeny and the discussion of species relationships will be outlined in a future taxonomic revision of the genus by Hamilton, Hendrixson, and Bond (in preparation).

4.1. Effective delimitation approaches

The initial methodological step to understanding species boundaries in this taxon group evaluated the usage of a 'barcode gap' to identify species boundary hypotheses. After reviewing both K2P and uncorrected *p*-distances across the full dataset, the

Table 2
Summary of the number of species delimited in each GMYC 'taxon inclusion group', the range of variation, and the average number of species within each 100 perturbations for both 'entities' and 'clusters'.

GMYC species delimitation – 'taxon inclusion groups' (%)	Range of spp. #	Avg. # of spp.
'Entities'		
10	1–36	25.81
20	1–72	20.5
30	4–107	21.82
40	8–142	23.04
50	8–178	36.47
60	9–210	45.84
70	10–85	56.53
80	12–104	67.19
90	12–118	83.65
100	114	–
'Clusters'		
10	1–13	5.79
20	1–19	7.04
30	1–26	8.73
40	2–35	13.97
50	2–43	24.52
60	5–49	32.46
70	9–63	41.79
80	10–69	49.67
90	11–79	60.21
100	83	–

presence of a universal 6% 'barcode gap' was not found to reliably and consistently identify nominal species. New species hypotheses were generated through the implementation of ABGD and a Maximum-Likelihood tree-based delimitation approach. The tree-based approach inferred a phylogeny, and somewhat subjectively delineated 41 species based upon major clade monophyly and high nodal support (Fig. 2). Based upon prior taxonomic knowledge and incorporation of the other delimitation approaches herein, ABGD appeared to over-split the 34 'barcode gap' species into 46 groupings (Fig. 2). Those species hypotheses that were over-split in both

the ABGD and tree-based approaches (also seen in the GMYC analyses) represent species where deep divergences occur between certain populations across their distributions (*hentzi*, *iodius*, *brunnius*, *eutylenum*, sp. *paloma* nov 1, sp. *mojave* nov E, sp. *mojave* nov W, and sp. nov G). Of these groupings, *hentzi*, *iodius*, *brunnius*, and *eutylenum* represent lineages that potentially experienced rapid expansions (evidence for *hentzi* in Hamilton et al., 2011). The lineages of sp. *mojave* nov E, sp. *mojave* nov W, sp. *paloma* nov 1, and sp. nov G represent species classified as 'dwarf' or 'pygmy' tarantulas. Due to their size and lack of dispersal ability, these 'dwarf' species exhibit genetic structuring very similar to what has been seen in other mygalomorph spiders (Bond et al., 2001; Hendrixson and Bond, 2005; Arnedo and Ferrández, 2007; Hendrixson and Bond, 2007; Starrett and Hedin, 2007; Bond and Stockman, 2008; Cooper et al., 2011; Satler et al., 2011; Hedin et al., 2013; Hendrixson et al., 2013). By scrutinizing the outputs of all approaches and applying prior taxonomic knowledge based on morphology, geographic distribution, and unique ecology or behavior, a 5% genetic distance cutoff was discovered that consistently and effectively recognized putative species boundaries of 16 nominal species while identifying 16 unknown species (Fig. 2). It has been noted in the past that the presence of a 'barcode gap' could reflect incomplete sampling across a species distribution (Moritz and Cicero, 2004). Our sampling scheme indicates that this is likely not the case in *Aphonopelma*, with the distinct clustering of populations and species truly representing a lack of past gene flow, with evolutionary events separating these independently evolving lineages.

Species hypothesis testing can be further extended by merging closely related OTUs or clades into their sister group or expanding and breaking up these groupings, statistically evaluating these hypotheses until a broadly-fitting consistent pattern is found across the tree. P ID(Liberal) recognized the same 16 nominal taxonomic species as the above barcoding and tree-based approaches, confirmed the same 7 newly discovered species, while identifying the same 9 cryptic species. P ID(Liberal) species designation probabilities were found to be highly significant ($\geqslant 93\%$) for all but one
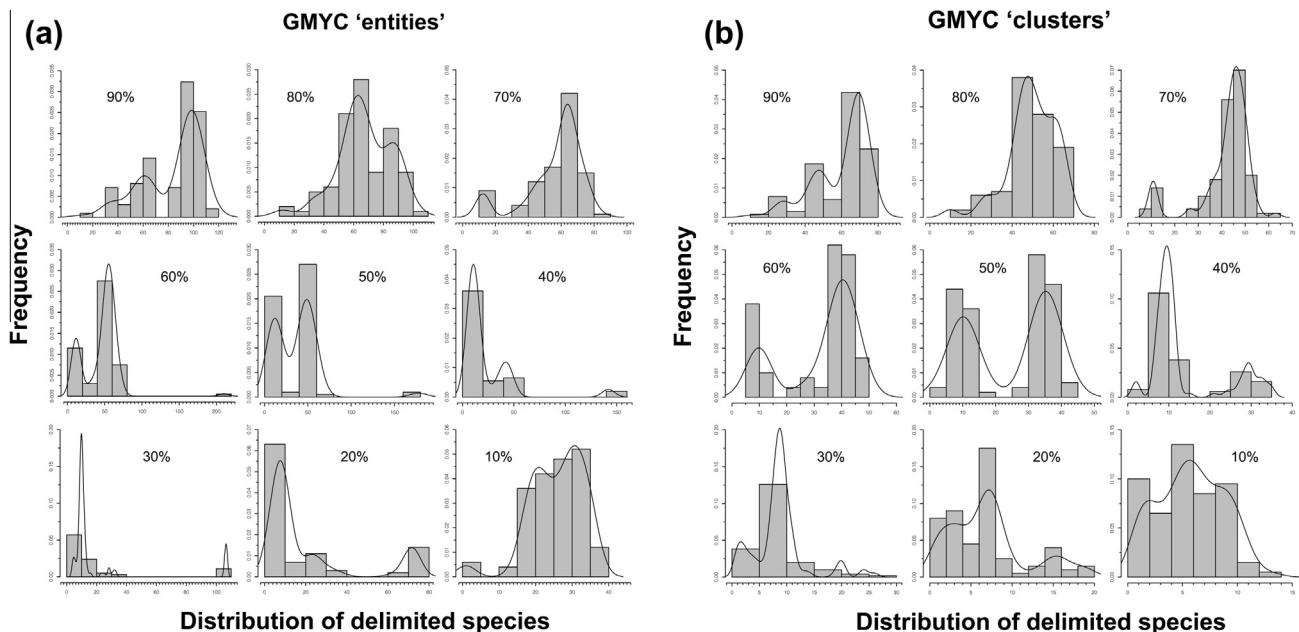

Fig. 3. Histograms representing the effects of random taxon sampling on Generalized Mixed Yule Coalescent (GMYC) delimitation outcomes – 'entities' (a) and 'clusters' (b). Each permutation was plotted to visualize the distribution of number of species delimited per 'taxon inclusion' group and the frequency with which the number of species delimited occurred. 'Entities' are considered a more realistic representation of the number of species in an analysis by representing all clades and all singletons deemed as independently evolving.
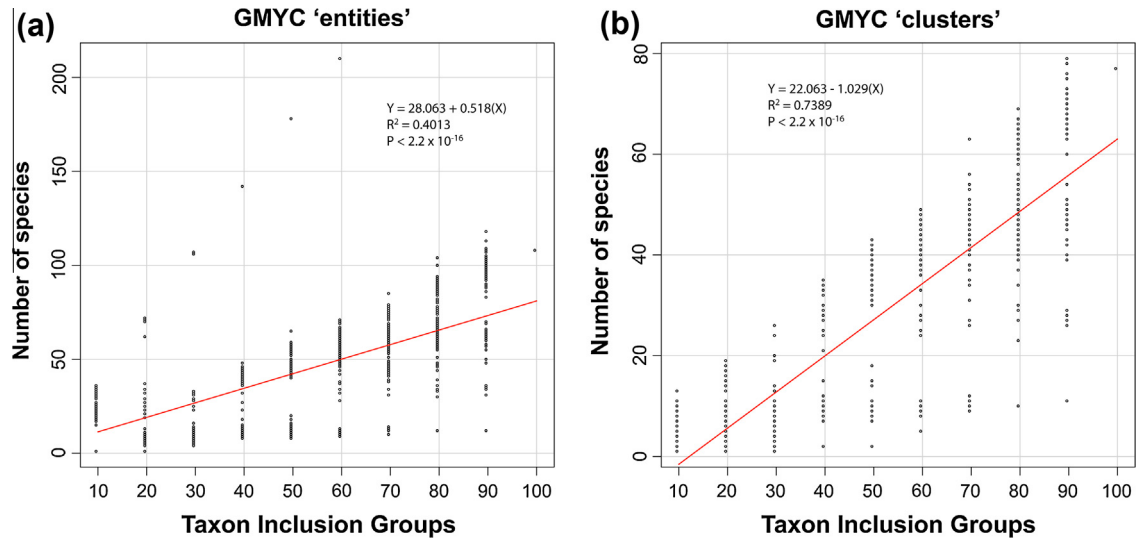
Fig. 4. Scatterplots representing the effects of random taxon sampling on GMYC delimitation outcomes. (a & b) Illustrates the variation of species delimited within each 'taxon inclusion group', in particular the extreme outliers. 'Entities' are considered a more realistic representation of the number of species in an analysis by representing all clades and all singletons deemed as independently evolving. The number of GMYC species is correlated to the total number of OTUs sampled in each 'taxon inclusion' group.

Table 3
Summary of the effects of random taxon sampling between Generalized Mixed Yule Coalescent (GMYC), Neighbor-Joining, and the quantitative approaches ABGD and P ID(Liberal). The lowest (low), a midpoint (medium), and highest (high) number of species delimited in each 'taxon inclusion group' were compared across analyses to evaluate sensitivity and congruence.

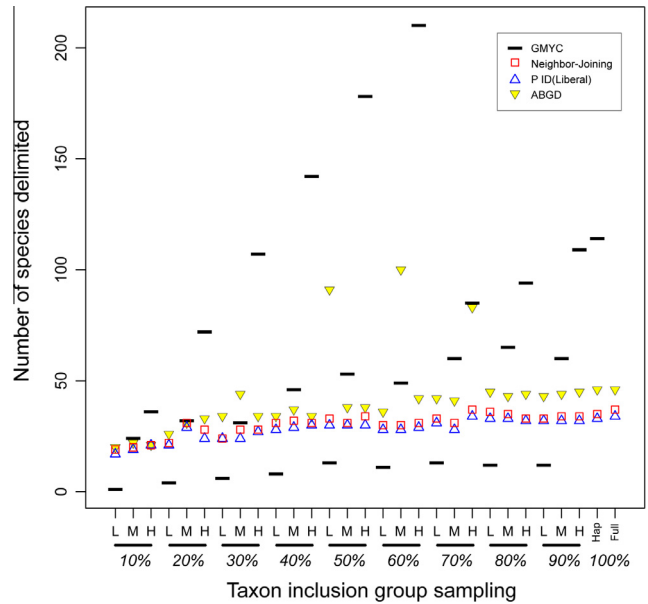| Sampled tree	GMYC 'entities'	NJ species	P(ID) species	ABGD species
10% low	1	19	17	20
10% medium	24	20	19	22
10% high	36	21	21	21
20% low	4	22	21	26
20% medium	32	31	29	31
20% high	72	28	24	33
30% low	6	24	24	34
30% medium	31	28	24	44
30% high	107	28	27	34
40% low	8	31	28	34
40% medium	46	32	29	37
40% high	142	31	30	34
50% low	13	33	30	91
50% medium	53	31	30	38
50% high	178	34	30	38
60% low	11	30	28	36
60% medium	49	30	28	100
60% high	210	31	29	42
70% low	13	33	31	42
70% medium	60	31	28	41
70% high	85	37	34	83
80% low	12	36	33	45
80% medium	65	35	33	43
80% high	94	33	32	44
90% low	12	33	32	43
90% medium	60	34	32	44
90% high	109	34	32	45
100% haplotype	114	35	33	46
100% full	–	37	34	46

Fig. 5. Comparison of the effects of random taxon sampling between GMYC, Neighbor-Joining, the quantitative approaches ABGD and P ID(Liberal). The relative stability of these alternative approaches lie in stark contrast to the divergent estimates seen across the suite of GMYC analyses. The sampled datasets and trees are represented as the lowest (L), a midpoint (M), and highest (H) number of species delimited in each 'taxon inclusion group'.

sampled specimens in the sp. nov C clade. Two other species possess high intra/inter ratios (*A. iviei* Smith, 1994 = 0.35; *A. iodius* (Chamberlin and Ivie, 1939) = 0.54), but both are wide-ranging, commonly encountered species throughout the southwestern US and were heavily sampled in this analysis. The southwestern US is an area well known for its complex geologic history, particularly during the Pliocene and Pleistocene, where populations of numerous taxa were heavily influenced by past fragmentation, restriction into refugia and subsequent expansion (Lamb et al., 1989; Riddle et al., 2000; Jaeger et al., 2005; Wood et al., 2012; Hedin et al., 2013; Graham et al., 2013). On the other hand, sp.
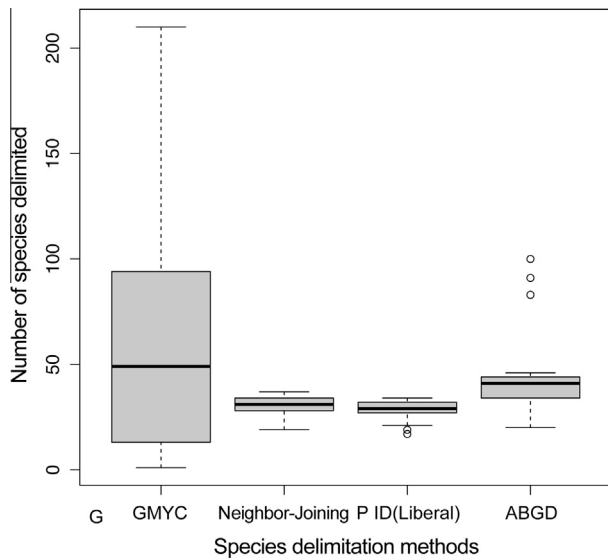
of the 32 redefined species (sp. nov C = 81% (CI = 67–96%) (Table 1). This particular outcome appears to be due to the relatively high intra/inter distance ratio (0.34) and the small number of

Fig. 6. Box plot comparison of the effects of random taxon sampling between GMYC, Neighbor-Joining, the quantitative approaches ABGD and P ID(Liberal).

nov C is a rarely collected spider with a very limited distribution in the Arizona "sky islands". The inclusion of only three specimens of this species in our sampling likely skews this inter/intra ratio, lowering the confidence of correctly placing an unknown specimen into this group.

All 32 US species that we ultimately determined to be independently evolving species were reciprocally monophyletic with high ML bootstrap support (⩾90) with two exceptions, the paraphyletic *A. iodius* grouping and the weakly monophyletic *A. brunnius* Chamberlin, 1940 lineage (Fig. 2) – both are members of the southwestern US "eutylenum group". This species complex contains 15 nominal taxa that are morphologically homogeneous. The paraphyletic *A. iodius* consists of two lineages: a clade including the *A. iodius* topotype – 99/99 (full/haplotype) bootstrap support, and a clade which includes the *A. nevadanum* Chamberlin, 1940 topotype (synonymized with *A. iodius* by Prentice (1997)) – low bootstrap support (27/28). The *brunnius* species group also possesses low bootstrap support (47/53) unless it is broken into two groups: a large, well supported lineage (100/100) inhabiting the southern Coastal Ranges of California, and a smaller poorly supported lineage that is sister to all other lineages in the clade (24/47) within the western Mojave desert (Table 1; Fig. 2). Two alternative hypotheses were tested using the P ID(Liberal): (1) the *iodius* lineage was broken into two monophyletic species groups (*iodius* & *nevadanum*) and (2) the *iodius* and *brunnius* lineages were collapsed into one monophyletic species group. P ID(Liberal) significance was indicated for all three of these hypotheses: *iodius* = 96% (CI = 94–99%); *nevadanum* = 88% (CI = 78–98%); *iodius*–*brunnius* = 96% (CI = 93–98%). The *iodius*–*brunnius* lineage has only moderate 89/88 (full/haplotype) bootstrap support. High P ID(Liberal) support does not ensure that a particular grouping represents a species, only that this grouping represents an independently evolving lineage; one which could be made up of a number of closely related sister species. Because of this, we reject these alternative groupings based on the violation of the established genetic distance limits and biogeographic knowledge that the divergent lineages of these species are derived from an area (the Mojave desert) that has been highly affected (population fragmentation) during past geologic and climatic events (Wood et al., 2012; Graham et al., 2013; Hendrixson et al., 2013), which may cause this discordance.

Of particular note is that the "*eutylenum* group" (*A. eutylenum* Chamberlin, 1940, *A. brunnius*, *A. iviei*, *A. iodius*, *A.* sp. *iodius* nov, *A.* sp. *chalcodes* nov) putatively represents the only cryptic species complex in the US tarantulas. While other cryptic species have been found in the US, this group is unique because it comprises a large group of closely related morphologically homogeneous sister lineages where, at this time, molecular data and geographic distribution represent the truly informative characters. This phenomenon has been found in other mygalomorph spiders (Hendrixson and Bond, 2005; Starrett and Hedin, 2007; Stockman and Bond, 2007; Bond and Stockman, 2008; Satler et al., 2011), but this is the first time in theraphosids. Problems with species boundary delimitation in this group could be a reflection of our usage of a single mtDNA locus. Unfortunately, effective nuclear markers with species-level phylogenetic information are lacking in theraphosids – although this will likely change in the future, as new high-throughput sequencing methodologies have become available for non-model organisms (Lemmon et al., 2012).

4.2. GMYC and sampling effects

Employment of the GMYC approach across the 358 OTU haplotype dataset identified 114 'entities' as independently evolving lineages. Following reevaluation of species delimitation no logical pattern can be seen in the GMYC designations (morphological, ecological, biogeographical, etc.) that would indicate to us these delineations are credible (Fig. 2), unlike the other investigative approaches. Do these results indicate the true diversity in *Aphonopelma* has been vastly underestimated in the United States? In order to investigate this question, the dataset was perturbed, randomly sampled, and GMYC was employed on each of the 900 datasets to test for effects of taxon sampling on delimitation outcomes. Our goal was to assess whether GMYC accurately and consistently delimited the same number of "species" independent of sampling.

GMYC delimitations were used as the benchmark for comparing dataset sampling sensitivity. When sampling was incomplete or biased, GMYC provided divergent outcomes, sometimes wildly, not only in number of species delimited within and between 'taxon inclusion groups' (Fig. 4a and b; Table 3), but also the number of times a particular number of species were delimited within the 'taxon inclusion group' subsampling replicates (Fig. 3a and b). Based solely on which OTUs have been sampled from across our dataset, the range of delimited species revealed tremendous variation, from 1 to 36 (in the 10% group), 9 to 210 (60%), and 12 to 118 (90%) (Table 3). As these delimitations are determined by the amount of genetic divergence found between specimens in a dataset and their evolutionary relationships, a sampling bias could arise simply from poor population sampling across a species' distribution, or from ecological and behavioral variation that allow for certain species to be found more easily (i.e. spatial (niche preference) and temporal (activity) differences). Comparisons to other methodologies indicate NJ, ABGD, and P ID(Liberal), particularly when employed in an integrative fashion and when sampling is robust, were much more stable approaches to understanding species boundaries regardless of which OTUs had been randomly sampled. As can be seen in Fig. 5 and Table 3, once sampling increases to just 40% of the total dataset, the number of species delimited using either NJ or P ID(Liberal) approaches only slightly fluctuated as sampling is increased. ABGD was relatively similar in its consistency, though generally higher in number of species delimited – subsampling did cause some instability with a few extreme outliers. ABGD appears to also be affected when individual species possess a mix of deep and very shallow divergences. The stability of these alternative approaches lie in stark contrast to the divergent estimates seen across the suite of GMYC analyses (Figs. 3a, 4a, and 5; Table 3).

GMYC has previously performed poorly when a small number of species are being investigated (O'Meara, 2010), when a large number of OTUs possess excessively long branches ('singleton' or 'unique' species) in relation to others are included in an analysis (Lohse, 2009), or recent and rapid divergences (Reid and Carstens, 2012). Reid and Carstens (2012) also indicate that GMYC performs best under scenarios where divergences are deep and sampling is incomplete – a scenario that does not generally reflect modern species-level and phylogeographic studies.

Deep genetic structuring (i.e. long branches) is known to affect the estimation of the coalescence point, causing significant overestimation in species numbers (see Lim et al., 2011), and has been observed in this group of spiders before (Hamilton et al., 2011; Hendrixson et al., 2013). As an example, in Hamilton et al. (2011) GMYC appeared to perform rather well though one lineage was problematic – *A. hentzi* split into nine entities. *A. hentzi* represents a species with a large effective population size, deep genetic structuring in the lineages that are sister to all other lineages in the clade, and rapid expansion in a number of other lineages.

GMYC assumes reciprocal monophyly, that the gene tree is the true tree (unlikely), and that there is no geographic substructure within species (highly unlikely). As pointed out by Lohse (2009), Papadopoulou et al. (2009), and Fujisawa and Barraclough (2013), the assumption of no geographic substructuring within a species is most likely always violated and will lead to hierarchical clustering, with population clusters being identified within species clusters thereby overestimating species entities. Unfortunately, GMYC does not incorporate information regarding the amount of time needed for lineages to sort, which can be substantial. Particularly problematic is the use of a Yule speciation model, as it does not represent a biologically realistic prediction of the evolutionary history of the lineages in question. An assumption of no extinction is likely violated in every dataset analyzed under GMYC.

Inter and intraspecific variation, speciation and extinction rates, degree of isolation, effective population size, and the migration rate are critical for accurate GMYC delimitation (Papadopoulou et al., 2009). Esselstyn et al. (2012) investigated the effects of effective population size and speciation rate on GMYC accuracy; their findings indicated that larger effective population sizes ($N_e > 10^5$) skew delimitations into an overestimation of species, with deeper coalescent events influencing the transition point on an ultrametric tree; likely what is occurring within the *A. hentzi* lineage in Hamilton et al. (2011). Further simulation work by Fujisawa and Barraclough (2013) indicates the main factor in the accuracy of delimitation corresponds with the N_e of the species relative to the divergence times between them – as the mean and variance in N_e increases, delimitation accuracy declines (e.g. when N_e increases above 10^4, the accuracy of GMYC fell to below 20%).

Simulations attempt to represent the differing effects of evolution on lineages, but due to the random nature of evolution and the large number of parameters that may be involved in shaping the species and populations we investigate, we may not be able to accurately represent the true evolutionary history of lineages under investigation. Our results appear to contradict Fujisawa and Barraclough (2013) that increased sampling of a particular species will improve performance. The variation seen within our 'taxon inclusion groups' is alarming (Figs. 3–6), particularly because a large number of the species sampled possesses heavy representation in the original dataset (see Table 1) – generally much more than the five specimens per species in the Fujisawa and Barraclough (2013) simulations. While the Fujisawa and Barraclough (2013) findings reflect the qualities of their particular simulated datasets possessing optimal parameters, datasets where the true evolutionary history of the group may fall outside these (e.g. a Goldilocks zone, where N_e is low and species divergence times

are high – similar to Reid and Carstens (2012)), are likely to be unsuitable for GMYC analysis.

Because of a mix of differing coalescent times, lineage sorting could be particularly problematic for a single-gene coalescent approach like GMYC in taxa comparable to *Aphonopelma* where a number of species (e.g., *A. hentzi, A. eutylenum, A. iodius, A. smithi*) possess short branch lengths relative to their large distributions and putative large N_e as a result of rapid and recent divergence while also harboring deep divergences within the lineages that are sister to all other lineages in those species. Simply stated, it is not just the random sampling of individuals within a species that determines the clustering, but the evolutionary events that led to the population structuring within a species, where a random sampling of species with differing values of intra- and interspecific divergence and N_e appears to weigh heavily on GMYC delimitation outcomes. While other studies (Papadopoulou et al., 2009; Esselstyn et al., 2012; Reid and Carstens, 2012; Fujisawa and Barraclough, 2013) have evaluated the effects of differing parameters on the estimation ability of GMYC, none have looked at the effects of randomly sampling taxa from the same dataset (where species and their distributions have been heavily sampled) on species delimitation consistency.

4.3. Methodological summary

A key point in studying closely related species, and perhaps one of the most important aspects of 'barcoding', is to combine a broad sampling strategy that identifies the extent of genetic diversity across the distribution of the species of interest (Funk, 1999), while building previous taxonomic and biological knowledge into the "species equation" (e.g. morphology, molecules, and natural history). Species delimitation does not exist in a vacuum, nor should it. Taken alone, these approaches are not particularly useful, especially in the absence of prior knowledge of the focal taxa, but the integration of multiple lines of evidence will likely guide the understanding of species boundaries. The consistent convergence of these methods (excluding GMYC) towards similar species number estimates supports the idea that we have effectively assessed species diversity within the context of the data available. In order to obtain confident species-level assignments, a comprehensive 'barcoding' profile within the group being investigated needs to be generated (a modified view of Hebert et al., 2003a). To understand the breadth of genetic variation within the focal taxa, comprehensive sampling will be required, which is particularly important when preliminary evidence suggests the taxonomy of the group of interest (sensu stricto *Aphonopelma*) is likely problematic or may harbor cryptic species (Hedin, 1997; Bond et al., 2001; Bond and Stockman, 2008; Satler et al., 2011; Hamilton et al., 2011; Bond, 2012; Hendrixson et al., 2013).

Following the integrative approach laid out above, an optimal path to understanding species boundaries starts with a likelihood tree-based framework to develop the initial species hypotheses; distinct clades are defined as those that do not share haplotypes between populations and can be identified as divergent monophyletic population clusters. Genetic distances are then calculated to assess levels of interspecific and intraspecific divergence for those initially defined species clades, allowing for subsequent iterative reassessment of species group membership. Neighbor-Joining inference can then be used to visually validate species group membership and highlight populations that should be investigated more closely for morphological, ecological, and/or biogeographic differences. Use of this prior biological knowledge is particularly useful, as it aids in identification of nominal species, identifies putative cryptic lineages, and confirms or refutes initial species boundary hypotheses. Species hypotheses are then statistically evaluated for species group membership through the implementa-

tion of ABGD and calculation of the P ID(Liberal) value. Questionable species groups can be altered and statistically reassessed, bearing in mind the previous biological knowledge, until a consistent fit is seen across the taxon of interest. Seeking congruence across multiple methods should allow for this type of integrative approach to be broadly applicable across taxonomic groups, not just within the mygalomorph spider genus *Aphonopelma* using mtDNA barcode data.

4.4. Potential issues

Meier et al. (2008) argued in favor of choosing the smallest interspecific genetic distance for determining the 'barcode gap'. Although DNA barcoding is an imperfect approach, our sampling (including both genetic and geographic breadth of most of these species) identifies a consistent pattern of divergence across species that are recognizably different (morphology, behavior, and ecology) from other species in their immediate region, thus leading to an assumption that the pattern of mean genetic distance ('barcode gap') could be effectively extended out to the other closely-related species (i.e., cryptic). Our investigation assessed the ability of popular molecular species delimitation methodologies to distinguish what we, as systematists, determine to be species – and thus they represent species hypotheses. While we agree with DNA barcoding opponents that there is no "right" way to determine species, we feel this marker is quite effective at distinguishing morphologically distinct lineages of *Aphonopelma*.

Mygalomorph spiders, particularly tarantulas, are known for their male-biased dispersal when large numbers of males can be seen crossing roads during mating seasons, while females are known to be highly philopatric. Due to the nature of mtDNA, our evaluations of GMYC and ABGD "over-splitting" lineages could very well be due to male-biased dispersal and the subsequent genetic signature of female population structuring. Effective single copy, phylogenetically-informative nuclear markers, when available (currently in development), will likely address such issues in the future.

Talavera et al. (2013) examine a number of factors that can affect GMYC performance: tree reconstruction method, taxon sampling coverage and taxon level, as well as geographic sampling and scale. According to their findings, analyses that are frequently employed due to their ease of use and speed (e.g. RAxML and the "chronopl" function for converting a tree to ultrametric) performed poorly when compared to a time-calibrated (or relative date) Bayesian inference in BEAST (Drummond et al., 2012). Future investigations should be aware of this computational trade-off and incorporate the appropriate methodologies.

Phylogenetic uncertainty (i.e. low node support) (Reid and Carstens, 2012) and incorrectly estimated/represented branch lengths (Talavera et al., 2013) can lead to unreliable GMYC species delimitation. While all (or mostly all, see above) of our delimited species clades possess high nodal support, there is uncertainty as one moves closer towards intraspecific tips as well as towards the root of the tree (Fig. 2). Reid and Carstens (2012) found that by looking at differences in the number of nucleotides (i.e. longer loci equal more information), the limited information residing in a single locus could lead to branch length uncertainty and significantly affect the outcome of GMYC. Future work should evaluate uncertainty in the GMYC delineation by comparing the AIC model-averaging approach, as proposed by Powell (2012) and Fujisawa and Barraclough (2013), or by employing the Bayesian GMYC method developed by Reid and Carstens (2012).

These findings do not suggest that GMYC never be used. Because both GMYC and ABGD start by identifying boundaries from the sequence data, they provide objective theory-based methods for identifying species boundaries. If it does not appear that taxa under investigation fit the required effective parameters, the methods could be used as the initial step to identifying putative independently-evolving lineages and therefore generate species hypothesis to be reevaluated through the incorporation of the methodology laid out above.

4.5. Conclusions

If the amount of diversity on Earth has been vastly undescribed, often due to cryptic diversity, action must be taken to accelerate sampling and the discovery process. DNA barcoding was initially proposed as a method to quickly aid overcoming this hurdle (Hebert et al., 2003a, 2003b), but taken alone it serves little utility in species delineation (only identification once a comprehensive taxonomy is produced). The common problem with the application of traditional DNA barcoding deals with questions relating to the amount of divergence required to correctly and consistently identify distinct species, as well as the extent of sampling that must be carried out to accurately assess this divergence. The literature is rife with criticisms on the use of DNA barcoding in species discovery and identification (Lipscomb et al., 2003; Seberg et al., 2003; Will and Rubinoff, 2004; Will et al., 2005; plus many more). Unfortunately, barcoding is not the panacea to the species problem; its shortcomings often mirror those that rely strictly on morphological identifications (lack of informative characters, homoplasy, etc.).

As the new age of genomics continues to evolve and vast amounts of whole genome data become available, one is left wondering if single marker DNA barcoding analyses will remain relevant. However, the ease at which these data can be collected and analyzed for hundreds or thousands of individuals supports the notion that there is still value in utilizing such an approach if it is effective for species delimitation; that is, the larger genomic data sets may be unnecessary for particular questions if species can be effectively delineated. DNA barcoding can be effective if it is not wielded as a blunt instrument, but instead in an informed manner that considers the group's taxonomy, ecology, biogeography, and population genetics in an integrative decision making process.

Appendix A. Supplementary material

All supplementary material, including data files, trees, and scripts, have been archived at figshare (http://dx.doi.org/10.6084/m9.figshare.769358). Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.11.007.

References

Arnedo, M., Ferrández, M., 2007. Mitochondrial markers reveal deep population subdivision in the European protected spider *Macrothele calpeiana* (Walckenaer, 1805) (Araneae, Hexathelidae). Conserv. Genet. 8, 1147–1162.

Astrin, J.J., Huber, B.A., Bernhard, M., Kluetsch, C.F.C., 2006. Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluation of species identification methods using CO1 and 16S rRNA. Zool. Scripta 35, 441–457.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the Earth's sixth mass extinction already arrived? Nature 470, 51–57.

Barrett, R., Hebert, P., 2005. Identifying spiders through DNA barcodes. Can. J. Zool. 83, 481–491.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The effect of geographical scale of sampling on DNA Barcoding. Syst. Biol. 61, 851–869.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155.

Bond, J., 2012. Phylogenetic treatment and taxonomic revision of the trapdoor spider genus *Aptostichus* Simon (Araneae, Mygalomorphae, Euctenizidae). ZooKeys 252, 1–209.

Bond, J.E., Hedin, M., Ramirez, M., Opell, B., 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. Mol. Ecol. 10, 899–910.

Bond, J.E., Opell, B., 2002. Phylogeny and taxonomy of the genera of south-western North American Euctenizinae trapdoor spiders and their relatives (Araneae: Mygalomorphae, Cyrtaucheniidae). Zool. J. Linn. Soc.-Lond. 136, 487–534.

Bond, J.E., Beamer, D., 2006. A morphometric analysis of mygalomorph spider carapace shape and its efficacy as a phylogenetic character (Araneae). Invertebr. Syst. 20, 1–8.

Bond, J.E., Beamer, D.A., Lamb, T., Hedin, M., 2006. Combining genetic and geospatial analyses to infer population extinction in mygalomorph spiders endemic to the Los Angeles region. Anim. Conserv. 9 (2), 145–157.

Bond, J.E., Hedin, M., 2006. A total evidence assessment of the phylogeny of North American euctenizine trapdoor spiders (Araneae, Mygalomorphae, Cyrtaucheniidae) using Bayesian inference. Mol. Phylogenet. Evol. 41, 70–85.

Bond, J.E., Stockman, A.K., 2008. An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. Syst. Biol. 57, 628–646.

Bond, J.E., Hendrixson, B.E., Hamilton, C.A., Hedin, M., 2012. A reconsideration of the classification of the spider infraorder Mygalomorphae (Arachnida: Araneae) based on three nuclear genes and morphology. PLoS ONE 7, e38753.

Bortolus, A., 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. AMBIO 37, 114–118.

Brower, A., 2006. Problems with DNA barcodes for species delimitation: "ten species" of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodivers. 4, 127–132.

Chamberlin, R.V., 1940. New American tarantulas of the family Aviculariidae. Bull. Univ. Utah 30, 1–39.

Chamberlin, R.V., Ivie, W., 1939. New Tarantulas from the Southwestern States. Bull. Univ. Utah 29, 1–17.

Chippindale, C., Taçon, P., 1998. The Archaeology of Rock-Art. Cambridge University Press, Cambridge.

Clement, M., Posada, D., Crandall, K., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.

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

Coyle, F., 1983. Aerial dispersal by mygalomorph spiderlings (Araneae, Mygalomorphae). J. Arachnol. 11, 283–286.

Coyle, F., 1985. Ballooning mygalomorphs: estimates of the masses of *Sphodros* and *Ummidia* balloonery (Araneae: Atypidae, Ctenizidae). J. Arachnol. 13, 291–296.

Dobzhansky, T., 1976. Organismic and Molecular Aspects of Species Formation. Sinauer Associates, Sunderland, Massachusetts.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Esselstyn, J.A., Evans, B.J., Sedlock, J.L., Khan, F.A.A., Heaney, L.R., 2012. Single-locus species delimitation: a test of the Mixed-Yule Coalescent model, with an empirical application to Philippine round-leaf bats. Proc. R. Soc. B 279, 3678–3686.

Ezard, T., Fujisawa, T., Barraclough, T.G., 2009. SPLITS: SPecies' LImits by Threshold Statistics. R Package Version 1.0-11.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Funk, D., 1999. Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. Mol. Biol. Evol. 16, 67–82.

Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent Approach: a revised method and evaluation on simulated data sets. Syst. Biol. http://dx.doi.org/10.1093/sysbio/syt033 (in press).

Goloboff, P.A., 1993. A reanalysis of Mygalomorph spider families (Araneae). Am. Mus. Novit., 1–32.

Goloboff, P.A., Mattoni, C.I., Sebastian Quinteros, A., 2006. Continuous characters analyzed as such. Cladistics 22, 589–601.

Graham, M.R., Jaeger, J.R., Prendini, L., 2013. Phylogeography of the Arizona hairy scorpion (*Hadrurus arizonensis*) supports a model of biotic assembly in the Mojave Desert and adds a new Pleistocene refugium. J. Biogeogr. 40 (7), 1298–1312.

Guthrie, R.D., 2005. The Nature of Paleothic Art. The University of Chicago Press, Chicago.

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

Hebert, P., Cywinska, A., Ball, S., DeWaard, J., 2003a. Biological identifications through DNA barcodes. Philos. Trans. R. Soc. Lond. B 270, 313–321.

Hebert, P., Ratnasingham, S., de Waard, J.R., 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Philos. Trans. R. Soc. B 270, S96–S99.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101, 14812–14817.

Hebert, P., Stoeckle, M., Zemlak, T., Francis, C., 2004b. Identification of birds through DNA barcodes. PLoS Biol. 2, 1657–1663.

Hebert, P., Gregory, T., 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 852–859.

Hedin, M., 1997. Molecular phylogenetics at the population/species interface in cave spiders of the southern Appalachians (Araneae: Nesticidae: *Nesticus*). Mol. Biol. Evol. 14, 309–324.

Hedin, M., Bond, J., 2006. Molecular phylogenetics of the spider infraorder Mygalomorphae using nuclear rRNA genes (18S and 28S): conflict and agreement with the current system of classification. Mol. Phylogenet. Evol. 41, 454–471.

Hedin, M., Carlson, D., 2011. A new trapdoor spider species from the southern Coast Ranges of California (Mygalomorphae, Antrodiaetidae, *Aliatypus coylei*, sp. nov.), including consideration of mitochondrial phylogeographic structuring. Zootaxa, 55–68.

Hedin, M., Maddison, W., 2001. Phylogenetic utility and evidence for multiple copies of Elongation Factor-1 alpha in the spider genus *Habronattus* (Araneae: Salticidae). Mol. Biol. Evol. 18, 1512–1521.

Hedin, M., Starrett, J., Hayashi, C., 2013. Crossing the uncrossable: novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, *Antrodiaetus*) from California. Mol. Ecol. 22, 508–526.

Hendrixson, B., Bond, J., 2005. Testing species boundaries in the *Antrodiaetus unicolor* complex (Araneae: Mygalomorphae: Antrodiaetidae). Mol. Phylogenet. Evol. 36, 405–416.

Hendrixson, B., Bond, J., 2007. Molecular phylogeny and biogeography of an ancient Holarctic lineage of mygalomorph spiders (Araneae: Antrodiaetidae: *Antrodiaetus*). Mol. Phylogenet. Evol. 42, 738–755.

Hendrixson, B.E., Bond, J.E., 2009. Evaluating the efficacy of continuous quantitative characters for reconstructing the phylogeny of a morphologically homogeneous spider taxon (Araneae, Mygalomorphae, Antrodiaetidae, *Antrodiaetus*). Mol. Phylogenet. Evol. 53, 300–313.

Hendrixson, B.E., DeRussy, B.M., Hamilton, C.A., Bond, J.E., 2013. An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, *Aphonopelma*). Mol. Phylogenet. Evol. 66, 327–340.

Huber, B.A., Astrin, J.J., 2009. Increased sampling blurs morphological and molecular species limits: revision of the Hispaniolan endemic spider genus *Tainonia* (Araneae: Pholcidae). Invertebr. Syst. 23, 281–300.

Irwin, D.E., 2002. Phylogeographic breaks without geographic barriers to gene flow. Evolution 56, 2383–2394.

Jaeger, J., Riddle, B., Bradford, D., 2005. Cryptic Neogene vicariance and Quaternary dispersal of the red-spotted toad (*Bufo punctatus*): insights on the evolution of North American warm desert biotas. Mol. Ecol. 14, 3033–3048.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kuntner, M., Agnarsson, I., 2011. Biogeography and diversification of hermit spiders on Indian Ocean islands (Nephilidae: *Nephilengys*). Mol. Phylogenet. Evol. 59, 477–488.

Kuo, C., Avise, J., 2005. Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. Genetica 124, 179–186.

Lamb, T., Avise, J.C., Gibbons, J.W., 1989. Phylogeographic patterns in mitochondrial DNA of the desert tortoise (*Xerobates agassizi*), and evolutionary relationships among the North American gopher tortoises. Evolution, 76–87.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.

Lim, G.S., Balke, M., Meier, R., 2011. Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Syst. Biol. 61, 165–169.

Lipscomb, D., Platnick, N., Wheeler, Q., 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol. Evol. 18, 65–66.

Locke, S.A., McLaughlin, D.J., Marcogliese, D.J., 2010. DNA barcodes show cryptic diversity and a potential physiological basis for host specificity among Diplostomoidea (Platyhelminthes: *Digenea*) parasitizing freshwater fishes in the St. Lawrence River, Canada. Mol. Ecol. 19, 2813–2827.

Lohse, K., 2009. Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). Syst. Biol. 58, 439–442.

Longhorn, S., Nicholas, M., Chuter, J., Vogler, A., 2007. The utility of molecular markers from non-lethal DNA samples of the CITES II protected "tarantula" *Brachypelma vagans* (Araneae, Theraphosidae). J. Arachnol. 35, 278–292.

Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis.

Masters, B.C., Fan, V., Ross, H.A., 2011. Species delimitation – A Geneious plugin for the exploration of species boundaries. Mol. Ecol. Resour. 11, 154–157.

Meier, R., Shiyang, K., Vaidya, G., Ng, P., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55, 715–728.

Meier, R., Zhang, G., Ali, F., 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Syst. Biol. 57, 809–813.

Meyer, C., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, 2229–2238.

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

Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2 (10), 1529–1531.

Niemiller, M.L., Near, T.J., Fitzpatrick, B.M., 2011. Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). Evolution 66, 846–866.

O'Meara, B.C., 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59, 59–73.

Opatova, V., Bond, J.E., Arnedo, M.A., 2013. Ancient origins of the Mediterranean trap-door spiders of the family Ctenizidae (Araneae, Mygalomorphae). Mol. Phylogenet. Evol. 69, 1135–1145.

Papadopoulou, A., Monaghan, M.T., Barraclough, T.G., Vogler, A.P., 2009. Sampling error does not invalidate the Yule-Coalescent Model for species delimitation. A response to Lohse (2009). Syst. Biol. 58 (4), 442–444.

Paradis, E., 2006. Analysis of Phylogenetics and Evolution with R. Springer Verlag.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Penney, D., Selden, P.A., 2011. Fossil Spiders: The Evolutionary History of a Mega-diverse Order. Siri Scientific Press, Manchester, 128 pp.

Petersen, F.T., Damgaard, J., Meier, R., 2007. DNA taxonomy: how many DNA sequences are needed for solving a taxonomic problem? The case of two parapatric species of louse flies (Diptera: Hippoboscidae: *Ornithomya* Latreille, 1802). Arthropod. Syst. Phylogeny 65, 119–125.

Pérez-Miles, F., Lucas, S.M., da Silva, P.I., Bertani, R., 1996. Systematic revision and cladistic analysis of Theraphosinae (Araneae – Theraphosidae). Mygalomorph 1, 33–68.

Platnick, N.I. (Ed.), 2013. The World Spider Catalog Version 14.0. American Museum of Natural History. <http://research.amnh.org/iz/spiders/catalog/INTRO1.html>. http://dx.doi.org/10.5531/db.iz.0001 (last accessed 23.07.13).

Pons, J., Barraclough, T., Gomez-Zurita, J., Cardoso, A., Duran, D., Hazell, S., Kamoun, S., Sumlin, W., Vogler, A., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609.

Powell, J.R., 2012. Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. Methods Ecol. Evol. 3, 1–11.

Prentice, T., 1997. Theraphosidae of the Mojave Desert west and north of the Colorado River (Araneae, Mygalomorphae, Theraphosidae). J. Arachnol. 25, 137–176.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21 (8), 1864–1877.

Raven, R., 1980. The evolution and biogeography of the Mygalomorph spider family Hexathelidae (Araneae, Chelicerata). J. Arachnol. 8, 251–266.

Raven, R.J., 1985. The spider infraorder Mygalomorphae (Araneae) – cladistics and systematics. Bull. Am. Mus. Nat. Hist. 182, 1–184.

Raven, R.J., 1990. Comments on the proposed precedence of *Aphonopelma* Pocock, 1901 (Arachnida, Araneae) over *Rhechostica* Simon, 1892. Bull. Zool. Nomencl. 47, 126–127.

Reid, N.M., Carstens, B.C., 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. 12, 196.

Riddle, B., Hafner, D., Alexander, L., Jaeger, J., 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. Proc. Natl. Acad. Sci. 97, 14438–14443.

Robinson, E., Blagoev, G., Hebert, P., Adamowicz, S., 2009. Prospects for using DNA barcoding to identify spiders in species-rich genera. ZooKeys 16, 27–46.

Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H., Tsai, P., 2008. The perils of plenty: what are we going to do with all these genes? Philos. Trans. R. Soc. B 363, 3893–3902.

Ross, H.A., Murugan, S., Li, W.L.S., 2008. Testing the reliability of genetic methods of species identification via simulation. Syst. Biol. 57, 216–230.

Satler, J.D., Starrett, J., Hayashi, C.Y., Hedin, M., 2011. Inferring species trees from gene trees in a radiation of California trapdoor spiders (Araneae, Antrodiaetidae, *Aliatypus*). PLoS ONE 6, e25355 (1–13).

Satler, J.D., Carstens, B.C., Hedin, M., 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). Syst. Biol. 62, 805–823.

Seberg, O., Humphries, C., Knapp, S., Stevenson, D., Petersen, G., Scharff, N., Andersen, N., 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. Trends Ecol. Evol. 18, 63–64.

Selden, P.A., Gall, J.C., 1992. A Triassic mygalomorph spider from the northern Vosges, France. Paleontology 35, 211–235.

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. Entomol. Soc. Am. 87, 651–701.

Smith, A.M., 1994. Tarantula Spiders – Tarantulas of the USA and Mexico. Fitzgerald Publishing, London.

Srivathsan, A., Meier, R., 2011. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 27, 1–5.

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

Starrett, J., Hedin, M., 2007. Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae). Mol. Ecol. 16, 583–604.

Stockman, A.K., Bond, J.E., 2007. Delimiting cohesion species: extreme population structuring and the role of ecological interchangeability. Mol. Ecol. 16, 3374–3392.

Talavera, G., Dinca, V., Vila, R., 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. Methods Ecol. Evol. http://dx.doi.org/10.1111/2041-210X.12107 (in press).

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Tautz, D., Arctander, P., Minelli, A., Thomas, R., Vogler, A., 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18, 70–74.

Wiens, J.J., Penkrot, T.A., 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51, 69–91.

Wilkins, J.S., 2009. Species: A History of the Idea. University of California Press, Berkeley and Los Angeles.

Will, K., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20, 47–55.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851.

Wood, D.A., Vandergast, A.G., Barr, K.R., 2012. Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. Divers. Distrib., 1–16.