1   **Bridging multilocus species delimitation and DNA barcoding through target enrichment of**

2   **UCEs: a case study with Mexican highland frogs**

3

4   Eugenia Zarza[1], Elizabeth M. Connors[1], James M. Maley[1], Whitney L.E. Tsai[1], Peter Heimes[2],

5   Moises Kaplan[3] & John E. McCormack[1,*]

6

7   [1] Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041

8   [2] Playa Miramar 441, Colonia Marte 08830, Mexico City, Mexico

9   [3] 1508 Brooklyn Ave, Ann Arbor, Michigan 48104

10

11

12   *Corresponding author: mccormack@oxy.edu

13

Abstract

Species delimitation has been divided by two approaches: DNA barcoding that focuses on standardization of the genetic marker and multilocus methods that place a premium on genomic coverage and conceptual rigor in modeling the divergence process. Most multilocus methods fail as barcodes, however, because few assay the same marker set and are therefore not readily comparable across studies and databases. We introduce ultraconserved elements (UCEs) as potential genomic barcodes that allow rigorous species delimitation and a bridge to DNA barcoding database to allow both rigorous species delimitation and standardized identification of delimited taxa. UCEs query thousands of loci across the nuclear genome in way that is replicable across broad taxonomic groups (i.e., vertebrates). We apply UCEs to species delimitation in a species complex of frogs found in the Mexican Highlands. *Sarcohyla* contains 24 described species, many of which are critically endangered and known only from their type localities. Evidence suggests that one broadly distributed member of the genus, *S. bistincta*, might contain multiple species. We generated data from 1,891 UCEs, which contained 1,742 informative SNPs for *S. bistincta* and closely related species. We also captured mitochondrial genomes for most samples as off-target bycatch of the UCE enrichment process. Phylogenies from UCEs and mtDNA agreed in many ways, but differed in that mtDNA suggested a more complex evolutionary history perhaps influenced by reticulate processes. The species delimitation method we used identified eight putative species (which we call lineages pending further study) within *S. bistincta*. Being able to compare linked mtDNA data to existing sequences on Genbank allowed us to identify one of these lineages nested within *S. bistincta* as an already-described species, *S. pentheter*. Another lineage nested within *S. bistincta* is currently being described as a new species (referred to here as *sp. nov.*). The remaining six lineages fell into two non-sister clades, one containing the core *S. bistincta* mostly in Oaxaca and Guerrero, and another in the Transvolcanic Belt. The latter clade, at 10% divergence in mtDNA and paraphyletic with respect to other *S. bistincta*, is a clear candidate for species status. Our study

40   demonstrates not only that UCEs can be used as effective genomic DNA barcodes, but that

41   combining multilocus genomic data with mtDNA is a powerful approach for both delimiting

42   species and identifying them in poorly described and phenotypically challenging groups.

43

44   Keywords: ultraconserved elements, genomics, population genetics, phylogeography,

45   phylogenetics, systematics, species limits, Hylinae

Introduction

Species are the fundamental biological unit and their identification and description is critical to ecology, evolution, and conservation (DeSalle and Amato 2004; Isaac et al. 2004; Mace 2004; Wiens 2007). Identifying species has always been a challenging issue in biology, especially when lineages are near the cusp of speciation (De Queiroz 2007). Yet, for the vast majority of researchers working on poorly described groups under conservation threat, conceptual debates over identifying the precise moment when one species becomes two have been superseded by the practical need for methods to quickly assess and identify the basic units of biodiversity (Hey et al. 2003; Tautz et al. 2003; Bickford et al. 2007). As a result, species delimitation methods have multiplied over the last ten years, especially those using DNA data (Pons et al. 2006; Knowles and Carstens 2007; O'Meara 2009; Yang and Rannala 2010; Fujita et al. 2012; Leaché et al. 2014).

While there are now a wide variety of methods that employ different DNA sources and algorithms to delimit species, these methods can be broadly placed into two frameworks. DNA barcoding focuses on species delimitation through the application of thresholds of sequence divergence, usually at a single mitochondrial DNA (Hebert et al. 2003) or chloroplast locus (Taberlet et al. 2007). DNA barcoding is relatively cheap and easy to implement, and by focusing on a single marker, a large database has been built by researchers, which allows for comparison of newly delimited species to already-identified species that have associated DNA. With its large user base, DNA barcoding has made substantial contributions to taxonomy and conservation of poorly known groups and geographic areas (Hebert et al. 2004; Ward et al. 2005; Witt et al. 2006; Lahaye et al. 2008). The very simplicity of DNA barcoding has made it controversial since its inception (Moritz and Cicero 2004; DeSalle et al. 2005; Ebach and Holdrege 2005; Will et al. 2005). Critics argue that it oversimplifies complex divergence processes with arbitrary thresholds that vary from one organismal group to another. Also, the history of a single gene will often fail to reflect the evolutionary history of populations and

4

71 species (Edwards and Bensch 2009; Galtier et al. 2009), thereby misleading delimitation efforts

72 in some cases.

73   Multilocus species delimitation, on the other hand, addresses many of the perceived

74 weaknesses of DNA barcoding by incorporating a process of gene inheritance (coalescence) and

75 by using information from many genomic locations to more accurately reflect the history of

76 lineages and how they diverged. Multilocus species delimitation, however, has been criticized

77 on the practical grounds that large amounts of genomic data are usually not needed to delimit

78 species, and the marginal gains more loci might afford are outstripped by the computational

79 burden of modeling complex speciation processes with large data sets (Collins and Cruickshank

80 2014). Another key criticism, which we focus on specifically, is that the lack of standardized

81 multilocus marker sets inhibits the broad adoption of multilocus delimitation methods because

82 newly delimited taxa cannot be readily compared to an existing database to determine if the

83 putative new species have already been identified and named and, if not, how they relate to

84 other species in an evolutionary context.

85   Our goal is to introduce ultraconserved elements (UCEs) as a candidate for a

86 standardized genomic marker set for multilocus species delimitation and identification and to

87 show how this kind of multilocus data can be even more powerful when linked to DNA

88 barcoding-type data (i.e., mtDNA). UCEs are appealing as genomic barcodes because the same

89 loci are found across major branches of the tree of life, where they act as anchors for variable

90 DNA in flanking regions (Faircloth et al. 2012). For instance, it is possible to capture the same

91 set of 1000 or more UCEs across all mammals (McCormack et al. 2012), all reptiles (Crawford et

92 al. 2012), or hundreds of UCEs across arachnid lineages separated by hundreds of millions of

93 years (Starrett et al. 2016). While the power of UCEs for deep-level systematics is clear, their

94 utility at shallower scales around the species level has only recently been demonstrated (Smith

95 et al. 2014; McCormack et al. 2016; Zarza et al. 2016). An added benefit of the UCE enrichment

96 process (and all so-called "sequence capture" methods) is that whole mtDNA genomes are often

97   captured as off-target "bycatch" (do Amaral et al. 2015), allowing for each individual to have

98   associated nuclear and mtDNA data (e.g., Zarza et al. 2016).

99   We apply UCEs as genomic barcodes to species delimitation in treefrogs in the genus

100   *Sarcohyla* in the northern Mexican Highlands. This genus, which was recently split from

101   *Plectrohyla* to reflect those species west of the Isthmus of Tehuantepec (Duellman et al. 2016),

102   contains 24 described species, many of them critically endangered and many that have never

103   been seen after their original discovery (references compiled in Stuart et al. 2008). Some species

104   are thought to be in serious decline or extinct (Lips et al. 2004). Species limits and relationships

105   within *Sarcohyla* are not well known because few species have associated DNA sequences, and

106   the lack of voucher material means that the tadpole stages and geographic ranges for each

107   species are not well characterized (Duellman 2001; Faivovich et al. 2009; Duellman et al. 2016).

108   We focus specifically on the *Sarcohyla bistincta* complex and closely related species. *Sarcohyla*

109   *bistincta* is one of the more broadly distributed members of the genus and might comprise

110   multiple species. Especially in its early developmental stages, *S. bistincta* looks similar to closely

111   related species, meaning that identification from tadpoles can be challenging. Thus, *Sarcohyla*

112   frogs present a good test case of whether UCE data are effective at delimiting species, and how

113   linking UCE and mtDNA data is a powerful approach for helping identify delimited taxa by

114   comparison to existing genetic databases.

115   **Methods**

116   *Sampling and Ingroup Determination*

117   MK collected tadpoles from January to June 2004 across most of the range where *Sarcohyla*

118   *bistincta* are known to exist (Duellman 2001) in the Transvolcanic Belt of Michoacán, Morelos,

119   and the state of México, the Sierra Madre del Sur of Guerrero, and the highlands of Oaxaca

120   stretching into Puebla and Veracruz (Fig. 1; Table 1). Unsampled parts of the *S. bistincta* range

121   include the far west Transvolcanic Belt in Michoacán and Jalisco, the far northwest in the Sierra

6

122   Madre Occidental (Nayarit, Durango, and Sinaloa), and the far northeast in Hidalgo (see Fig. S1

123   for sampled and unsampled locations and known ranges of all *Sarcohyla* species). Tadpoles

124   were targeted to improve sampling efficiency. After collection of many tadpoles from a

125   sampling location with a dip net, they were separated by species based on morphology to the

126   extent possible and reared to subadults in the laboratory prior to vouchering. Species

127   identification was based on the most recent diagnosis of *S. bistincta* and other closely related

128   species (Duellman 2001). One tadpole was chosen for the tissue voucher, while the other

129   individuals became physical vouchers with museum catalog numbers. Thus, we provide both

130   field numbers and catalog numbers in Table 1 to provide a link to both the exact genetic

131   material and the associated voucher for that genotype. Before limiting our taxonomic sampling

132   to 38 *S. bistincta* samples and two close outgroups, we ran preliminary phylogenetic analyses

133   also including samples from 45 *Sarcohyla* individuals and the outgroup genus *Exerodonta* to

134   ensure we had correctly identified the ingroup and closest outgroups (Table S1).

135   *Sequence capture and next-generation sequencing*

136   We extracted genomic DNA from tissue using a Qiagen (Valencia, CA) DNAeasy Blood and

137   Tissue extraction kit. We visualized extractions on an agarose gel to ensure fragments were

138   larger than 200 base pairs (bp) and quantified the resulting double-stranded DNA using a Qubit

139   2.0 Fluorometer (Carlsbad, CA). For each sample, we sheared 100 μl of 20ng/μl concentration

140   DNA to a size distribution with its peak between 400 and 600 bp using a Bioruptor

141   ultrasonicator (Diagenode). We prepared libraries for each sheared sample with a KAPA

142   (Boston, MA) LTP library preparation kit for the Illumina platform, attaching custom indexing

143   tags (Faircloth and Glenn 2012) to each sample to allow sample pooling.

144   We enriched pools of eight samples using a set of synthetic RNA probes that target 5,060

145   tetrapod UCEs (MYbaits_Tetrapods-UCE-5K kit, Mycroarray) following the standard UCE

146   enrichment protocol (Faircloth et al. 2012) with one modification. Amphibians have large and

147   variable genome sizes with a high percentage of repetitive DNA (Olmo 1991). While we do not

7

148 have information about the genome size and composition of *Sarcohyla* specifically, we wanted to

149 decrease the potential risk of the probes hybridizing to repetitive elements (McCartney-Melstad

150 et al. 2016). We thus increased by 6X the amount of the Cot-1 blocker, a synthetic DNA derived

151 from chicken that binds to repetitive regions. After enrichment and recovery PCR, we verified

152 the library size range with an Agilent 2100 Bioanalyzer (Palo Alto, CA). We quantified the

153 enriched pools using qPCR and combined them in equimolar ratios before sequencing on an

154 Illumina HiSeq 2000 lane (100-bp paired-end cycle) at the University of California Santa Cruz

155 Genome Technology Center.

156 *Bioinformatics of next-generation sequencing data*

157 We demultiplexed the Illumina raw reads and converted them to FASTQ format with the

158 program bcl2fastq v.1.8.4 (Illumina, Inc.). We created quality control reports with FASTQC

159 v0.10.1 to detect overrepresented sequences and adapter contamination. To eliminate adapter

160 contamination and low quality bases, we trimmed the FASTQC output using illumiprocessor

161 (Faircloth 2012). We trimmed and assembled these reads into contigs with Trinity (Haas et al.

162 2013) and ABySS (Simpson et al. 2009), both of which are built into the PHYLUCE pipeline

163 (Faircloth 2015). PHYLUCE uses LASTZ (Harris 2007) to align all assembled contigs to UCE

164 probe sequences in order to isolate only UCE contigs and to identify and eliminate any paralogs

165 (MATCH_CONTIGS_TO_PROBES script). We then turned this filtered set of UCE contigs into a

166 data matrix describing which UCE loci are present in which samples (GET_MATCH_COUNTS

167 script).

168 *Phylogenetic trees from concatenated UCE data*

169 We extracted contigs into a single FASTA file (GET_FASTAS_FROM_MATCH_COUNTS script)

170 and aligned the output for each locus (SEQCAP_ALIGN script) using MAFFT (Katoh et al.

171 2005). We then required that 75% of the samples needed to have data for a given locus to be

172 included in the final concatenated matrix (GET_LOCI_WITH_MIN_TAXA script). We

173   converted the NEXUS file into PHYLIP format (FORMAT_NEXUS_FILES_FOR_RAxML script)

174   and constructed a maximum-likelihood (ML) tree in RAxML v8.0.19 (Stamatakis 2014) under

175   the GTRGAMMA model of evolution with 100 bootstrap searches, followed by a search for the

176   tree with the highest likelihood.

177   *Mitochondrial DNA assembly and analysis*

178   We identified and assembled mtDNA genomes from off-target, trimmed Illumina reads using

179   the reference genome of a closely related species, *Hyla annectans* (Genebank accession number

180   KM271781; Ye et al. 2016). We used MITObim 1.7 (Hahn et al. 2013), a Perl wrapper for MIRA

181   4.0.2 (Chevreux et al. 1999), that takes a baiting and iterative mapping approach for assembly.

182   We conducted *de novo* annotation of the assembled mtDNA regions with the MITOchondrial

183   genome annotation Server, MITOS (Bernt et al. 2013). We selected for phylogenetic analysis

184   only those individual genomes with MIRA quality score grater than 30. We aligned each

185   protein-coding region separately in Geneious vR8 (Kearse et al. 2012) using the integrated

186   MUSCLE (Edgar 2004) plugin. We corrected the alignments manually when necessary and

187   constructed a concatenated mtDNA matrix, which we also ran in RAxML v8.0.19.

188   Additionally, we melded our mtDNA data with existing *Sarcohyla* and *Plectrohyla*

189   mtDNA data on Genbank to determine whether any of the lineages we uncovered in *S. bistincta*

190   relate to already-described species. We determined that *cytochrome b* is the best-represented on

191   Genbank in this group. We downloaded all existing *cytochrome b* sequences from *Sarcohyla* and

192   *Plectrohyla* taxa. We combined these sequences with those from a subset of our *S. bistincta*

193   samples, choosing the individual with the most raw reads from each major genetic lineage in

194   the UCE tree. We used *Exerodonta* as an outgroup. Here, we aligned the trimmed, filtered reads

195   for each individual to a *Sarcohyla cytochrome b* reference sequence. We formed a consensus

196   sequence for each individual from the mapped reads. We then created an alignment and

197   generated a phylogeny using BEAST v2.4.2 (Bouckaert et al. 2014). Because we were particularly

9

198   interested in how *S. pentheter* related to our lineages, but there are no *cytochrome b* sequences for

199   *S. pentheter* on Genbank, we also compared our samples to Genbank samples using the *16S*

200   gene.

201   *Calling SNPs from UCE loci*

202   We called SNPs from UCE loci so that we could run genetic clustering tests and infer a species

203   tree. Calling SNPs requires a reference sequence, and we chose the sample with the most UCE

204   contigs recovered within the ingroup (UMMZ 239727). We then used BWA (Li and Durbin

205   2009) to map the reads of each sample to this reference. We used SAMtools (Li et al. 2009) to

206   sort the reads, and Picard (available at http://broadinstitute.github.io/picard) to identify and

207   remove PCR duplicates. We realigned the mapped reads to minimize mismatched bases due to

208   indels, and we removed indels using the Genome Analysis Toolkit 3.2 (GATK; McKenna et al.

209   2010) and a custom script (indelrealigner.sh), as suggested by the Best Practices workflow

210   (DePristo et al. 2011; van der Auwera et al. 2013).

211   There is no SNP database available for treefrogs, so we followed best practices for base

212   recalibration for non-model organisms suggested by GATK (McKenna et al. 2010). This consists

213   of (1) doing an initial round of calling SNPs on the original, uncalibrated data, (2) selecting the

214   SNPs with the highest confidence (a minimum emission and call quality of 40 or more), and (3)

215   using these SNPs as the database of known SNPs. We executed four rounds of base

216   recalibration on the original data to filter out systematic error using a custom script (genotype-

217   recal.sh). We called genotypes on the last recalibrated BAM file. We used vcf-tools (Danecek et

218   al. 2011) to select one SNP per UCE and produce two data sets, one allowing 25% missing data

219   for STRUCTURE v 2.3.4 (Pritchard et al. 2000), and one with no missing data, which is a

220   requirement for SNAPP (Bryant et al. 2012) species tree analysis, implemented in BEAST v2.2.1

221   (Bouckaert et al. 2014).

222   *STRUCTURE analyses*

10

223   The species delimitation method we used requires some *a priori* information about possible

224   clustering of individuals into species. We used STRUCTURE v2.3.4 as an unbiased way to

225   assess the limits of fine-scale genetic structure in our data. Thus, our intent was not to

226   determine the single most likely number of genetic clusters. Rather, our goal was to determine

227   the maximum number of genetic clusters in our data. We began by analyzing all individuals of

228   *S. bistincta* plus two outgroup species *S. chryses* and *S. hazelae* under $K=4$, reasoning that this

229   would likely split out the two outgroups as well as revealing one division within the ingroup.

230   After this, each identified genetic cluster was further analyzed at $K=2$ until no further structure

231   was evident.

232   *SNAPP tree and species delimitation*

233   We generated a species tree from the SNP matrix using SNAPP 1.1.10 (Bryant et al. 2012). This

234   analysis included all putative *S. bistincta* samples and one outgroup, *S. chryses.* For this run, we

235   made no a priori assumptions about how individuals grouped into species and allowed each

236   individual to be considered its own "species" (i.e., terminal tip). We ran two instances of

237   SNAPP for seven million generations using default priors. We combined tree and parameter

238   files from both runs with LogCombiner 2.1.3 and displayed the full set of likely species trees

239   with Densitree v2.2.1 (Bouckaert et al. 2014).

240   We then applied Bayes Factor Species Delimitation (BFD; Grummer et al. 2014; Leaché et

241   al. 2014) to SNAPP analyses to determine the most likely number of species. We used the

242   Structure results to generate hypotheses for how individuals might cluster into species. We

243   tested five scenarios with BFD that split or lumped populations identified in the Structure

244   analysis, including the hypothesis that all putative *S. bistincta* group into a single species. We

245   assessed which scenario (2, 6, 7, 9, or 10 species) was favored by comparing Bayes factors using

246   marginal likelihoods from the SNAPP analysis.

247   **Results**

248   *NGS summary statistics*

249   Detailed summary statistics for each of the 38 ingroup samples and two outgroups are

250   described in Table 1. ABySS produced longer contigs than Trinity, and a higher number of UCE

251   loci, so we used ABYSS contigs in all downstream analyses. Reads per sample ranged from

252   17,052 to 3,423,330 with an average of 1,185,165 reads. The number of identified UCEs ranged

253   from 381 to 2,444 with an average of 1,976 UCEs. The mean length of individual UCE loci per

254   individual ranged from 222 to 717 bp with an average of 522 bp. On average, 18% of the

255   assembled contigs corresponded to unique UCE loci.

256   For SNP calling, across 40 samples of *S. bistincta* and outgroups, 9% of the trimmed

257   reads mapped to our designated reference individual. The SNP read depth ranged from 2.4 to

258   35.0 with an average depth of 21.2. The recalibration and quality control steps resulted in an

259   initial matrix of 16,578 SNPs. After removing non-biallelic loci, selecting one SNP for every

260   UCE, and allowing 25% missing data, there were 1,742 SNPs in the STRUCTURE data set, while

261   the 100% complete data matrix for SNAPP contained 399 SNPs.

262   *UCE phylogeny from concatenated data*

263   Our more taxonomically inclusive data set with all available *Sarcohyla* and outgroup *Exerodonta*

264   *xera* (Table S1) contained 1,866 UCE loci and 1,030,450 bp for a concatenated analysis. The

265   resulting ML tree (Fig. S2) showed strong support for monophyly of *Sarcohyla*, and identified *S.*

266   *arborescandens* and *S. cyclada* as sister species that together form a clade sister to the rest of the

267   *Sarcohyla* included in the study. We thus limited further analyses to a smaller data set of 40

268   samples with *S. hazelae* as the outgroup (Table 1). This focal data set contained 1,891 UCE loci

269   and 1,038,600 bp. The ML tree of these 40 samples found strong support for many clades within

270   the species currently described as *S. bistincta*, conforming to distinct geographic areas (Fig. 2a).

271   In brief, there were three clades on relatively long branches: one distributed across the

12

272   Transvolcanic Belt (blue clade); one inhabiting two disjunct areas along the coastal slopes of the

273   Sierra Madre del Sur in Guerrero and Oaxaca (gray clade); and one broadly distributed in the

274   Sierra Madre del Sur (red and pink clades), the Oaxaca Highlands (yellow + orange clade), and

275   one individual in the southern portion of the Transvolcanic Belt (purple). One individual that

276   nested within *S. bistincta* was labeled as a different species, *S. mykter*, from Guerrero. We

277   suspect that this sample was mislabeled and is actually a duplicate of an *S. bistincta* sample

278   already included in the study because their field numbers are similar (last two digits

279   transposed) and the two samples grouped together in all analyses. We have left this sample in

280   all analyses, but labeled it as a duplicate of *S. bistincta* UMMZ 239749.

281   *mtDNA tree*

282   Our final concatenated mtDNA matrix was 11,269 base pairs including gaps, as coverage of the

283   mtDNA genome varied from sample to sample in accordance with the non-targeted nature of

284   the DNA collection (Table 1). Relationships in the ML tree (Fig. 2b) among the 29 individuals

285   with high quality scores were similar to the concatenated UCE tree with two key differences

286   within the broadly distributed clade in Guerrero and Oaxaca: (1) in the mtDNA tree,

287   individuals from Eastern and Western Guerrero (pink and red) formed a clade, whereas they

288   were more divergent in the UCE tree; (2) in the mtDNA tree, individual UMMZ 239731 (purple)

289   was still on a relatively long branch, but that branch was nested within the Guerrero clade

290   described above instead of being sister to a much more expansive clade, as in the UCE tree.

291   *Structure analysis*

292   The first run of STRUCTURE at K=4 split the two outgroup species into distinct clusters and

293   split the remaining individuals into two clusters. Further analysis of each cluster at K=2

294   revealed nine genetic clusters (Fig. 1b), which are concordant with clades observed in the UCE

295   and mtDNA phylogeny, as well as with the delimited species (lineages) described below.

296   *SNAPP species tree and species delimitation*

297   The cloudogram of posterior species trees obtained with SNAPP (Fig. 3) is consistent with the
298   genetic clusters in the Structure analysis and with relationships in phylogenies resulting from
299   the UCE and mtDNA concatenated analyses. With respect to the discrepancies between the two,
300   the SNAPP tree agreed with the mtDNA tree that eastern and western Guerrero individuals
301   form a clade, but agreed with UCE tree in the placement of individual UMMZ 239731. Bayes
302   factor analysis of SNAPP runs with varying number of species suggested the best model
303   consists of nine species. The model with nine species was supported by a factor of 10 over other
304   models with more and fewer species (Table 2). These putative species are labeled as lineages in
305   Fig. 3 and elsewhere, pending further study.

306   *mtDNA phylogeny combining new data with Genbank sequences*

307   Using 16S sequences, we determined that our delimited Lineage 8 matched an *S. pentheter*
308   sequence on Genbank. Lineage 8 individual UMMZ 239772 had one of the lowest read counts of
309   any samples and very few mtDNA reads. However, five reads mapped to 16S covering 421 bp
310   of the 681 bp reference sequence (Genbank *S. pentheter* accession number DQ055825). Over this
311   stretch, UMMZ 239772 was identical to the reference. As a point of comparison, UMMZ 239679
312   (a member of the blue *S. bistincta* Lineage 1 in the Transvolcanic Belt) had 70 differences across
313   the 681 bp (10.3% divergence). This DNA identification of UMMZ 239772 as *S. pentheter* was
314   confirmed by re-examining the subadult specimen.

315   After confirming UMMZ 239679 as *S. pentheter*, we generated a Bayesian tree of
316   *cytochrome b* combining our samples with Genbank sequences (Fig. 4). This tree revealed not
317   only that Lineage 8 (*S. pentheter*) was nested within the current *S. bistincta*, but so was another
318   species not included in our sampling, *S. calthula*. Relationships within *S. bistincta* were
319   otherwise similar to the more taxonomically inclusive mtDNA phylogeny. For example UMMZ
320   239731 (purple) was still nested within the Guerrero samples (red and pink). The combined
321   phylogeny also helped clarify relationships outside of *S. bistincta* by supporting *S. chryses + S.*
322   *mykter* to be sister to the *S. bistincta + S. pentheter + S. calthula* complex.

14

Discussion

324   One limitation of current multilocus species delimitation methods is the lack of a standardized

325   marker set that would allow for data from delimited species to be combined with already-

326   existing data both to identify species and place them into broader evolutionary context (Collins

327   and Cruickshank 2014; Coissac et al. 2016). Here, we show that UCEs not only are a good

328   candidate for this standardized "genomic barcode" for multilocus species delimitation, but also

329   that they are especially powerful when they can be linked to mtDNA data, which provides a

330   bridge to well-developed mtDNA databases for species identification. Having linked mtDNA

331   data allowed us to query our lineages, delimited with multilocus genomic data, against

332   GenBank to see if any of them corresponded to already-described species and to place them in

333   as broad a phylogenetic context as possible. Doing so revealed that *S. bistincta* is paraphyletic,

334   with two already-described species and one species in the process of description nested within

335   its current species limits (Fig. 4). Our approach of delimiting species with multilocus data and

336   identifying them with the help of mtDNA data will be especially powerful in groups where

337   geographic ranges and subadult phenotypes are not well described or where species are

338   phenotypically cryptic.

339   It was not our intention with this study to compare species delimitation methods, as this

340   has been examined elsewhere (Camargo et al. 2012; Carstens et al. 2013; Miralles and Vences

341   2013). We employed one, rather new delimitation method (Leaché et al. 2014), but there are

342   others that would surely suggest somewhat higher or lower numbers of delimited species. We

343   are not advocating that species be delimited and described solely based on DNA. There are

344   cases where both DNA barcoding and multilocus species delimitation will under-split species

345   (Hime et al. 2016) and cases where coalescent methods will confound population structure with

346   speciation, thereby recommending over-splitting species (Sukumaran and Knowles 2017). We

347   discuss the results of our delimitation analysis below, but by calling the putative species

348   "lineages" we acknowledge that new species descriptions should await integrative taxonomy

15

349 including, at minimum, careful study of phenotypic differences in museum specimens. Finally,

350 although we discuss discrepancies between the mtDNA and nuclear DNA phylogenies below,

351 we did not set out explicitly to compare multilocus species delimitation to DNA barcoding. As

352 mentioned, we feel our results demonstrate how mtDNA and nuclear data are most powerful

353 when used together. Despite a few notable differences, discussed below, our UCE and mtDNA

354 phylogenies are broadly congruent, suggesting that, if we carried out such a comparison, our

355 results would be in line with prior work showing that multilocus species delimitation and DNA

356 barcoding produce similar results (Collins and Cruickshank 2014; Dowton et al. 2014).

357 Rather, in addition to showing the power of linking nuclear genomic data with mtDNA

358 data, another goal of our study was to demonstrate that UCEs adhere to the concept of a

359 "extended genomic barcode" (sensu Coissac et al. 2016) and produce data sets with enough

360 resolving power for the necessarily recent divergences involved in species delimitation.

361 Another recent study applied UCEs to species delimitation in two other frog genera,

362 *Melanophryniscus* and *Brachycephalus* (Pie et al. 2017). This study used a more restrictive cut-off

363 for including loci and produced 800 loci (in a 100% complete data matrix) with sufficient

364 resolving power for species delimitation, compared to the roughly 1,900 loci (in a 75% complete

365 data matrix) in our study. UCE probe sets are now available for many taxonomic groups

366 (Faircloth et al. 2013; Faircloth et al. 2015; Starrett et al. 2016). They capture a discrete and

367 replicable portion of the genome, in this case a set of around 2,000 loci in frogs (from a larger set

368 of ~5,000 vertebrate loci) that query approximately 1,000,000 base pairs, or 0.02% of the frog

369 genome. The replicable nature of UCEs sets them apart from other types of genomic markers,

370 like RAD loci, which can vary from experiment to experiment (DaCosta and Sorenson 2014) and

371 find fewer orthologs with increasing phylogenetic distance (Cruaud et al. 2014).

372 Bayes factor species delimitation identified nine lineages within the current *S. bistincta*

373 and a close outgroup, *S. chryses*. Apart from *S. chryses*, which was correctly delimited, the other

374 eight lineages fell into three clades (Fig. 3). One lineage turned out to be the already-described

16

375   *S. pentheter* of Oaxaca, which we discovered by querying its 16S mtDNA data against existing

376   Genbank sequences and reexamining the subadult specimen. Thus, both *S. pentheter* and *S.*

377   *calthula* are nested within the currently-described *S. bistincta*. Below, we describe the three

378   clades containing the eight lineages. Where warranted, formal species descriptions will follow

379   in later publications.

380   **Clade 1** – Members of this clade are found in the Transvolcanic Belt of central Mexico. They are

381   sister to the rest of the current *S. bistincta* + *S. pentheter* + *S. calthula*. This clade is a

382   clear candidate for species recognition and might itself contain multiple species, as

383   delimited below. Based on geography, unsampled populations of *S. bistincta* in the

384   Sierra Madre Occidental (Fig. S1) are most likely related to this clade, but should be

385   included in future studies, as they might comprise their own lineage.

386   Lineage 1 (light blue in Fig. 1) – Michoacán to western Mexico state. There is fine-scale

387   genetic structure across this range, and the presence of a geographic and genetic

388   intermediate hints at continuity of gene flow along the distribution from sites 1 to 4

389   in Figure 1. Some populations in far western Michoacán (Fig. S1) are as yet

390   unsampled and could reveal further genetic structure.

391   Lineage 2 (dark blue in Fig. 1) – Morelos. Further sampling between sites 4 and 5 could

392   help determine whether the genetic distinctness of this individual is a true genetic

393   discontinuity or the result of a sampling gap.

394   **Clade 2** – Members of this clade form the core *S. bistincta* and occur from Guerrero to Puebla

395   and Veracruz and south through Oaxaca.

396   Lineage 3 (red in Fig. 1) – Central and eastern Guerrero. Members of this lineage are

397   distinct from the pink lineage below and are monophyletic in the UCE trees (though

398   not in the mtDNA tree). Further sampling in between site 6 and site 7 would clarify

17

399   whether the genetic discontinuity between Lineages 3 and 4 results from a sampling

400   gap.

401   Lineage 4 (pink in Fig. 1) – Western Guerrero. This lineage is distinct and monophyletic

402   in all trees, although only a few individuals were sampled only from one locality.

403   Lineage 5 (orange and yellow in Fig. 1) – Puebla, Veracruz, and Oaxaca. The geographic

404   range of this group contains the type locality for *S. bistincta*. Central and southern

405   Oaxaca individuals (orange) are genetically distinct from individuals to the north

406   (yellow). One genetic intermediate in central Oaxaca suggests genetic continuity

407   across this range. One individual from southern Oaxaca (site 25 in Fig. 1) has an

408   unexpectedly northern genetic profile given its geographic location, suggesting a

409   more complex geographic structuring within this group than a simple cline from

410   north to south. An unsampled northern population of *S. bistincta* in Hidalgo is most

411   likely related to this lineage, and should be included in future studies.

412   Lineage 6 (purple in Fig. 1) – far northern Guerrero. Represented by a single individual,

413   this lineage is distinct and on a relatively long branch. However, in the mtDNA

414   tree, this individual is nested within Lineages 4 and 5 above. Sampling more

415   individuals is needed to determine how distinct this lineage might be.

416   **Clade 3** – Pacific slope of Guerrero and Oaxaca. Sister to Clade 2, members of this clade

417   comprise two species, one already described and one currently being described.

418   Lineage 7 (dark gray in Fig. 1) – Pacific slope of Guerrero. This lineage was being

419   described as a new species (we call it *sp. nov.*) on the basis of phenotypic differences

420   before this genetic study was begun. Thus, our analysis lends support to species

421   status for this lineage.

18

422   Lineage 8 (light gray in Fig. 1) – Pacific slope of Oaxaca. Genbank 16S mtDNA data and

423   the phenotype of the voucher reveal this lineage to be the already-described species,

424   *S. pentheter*, of southern and central Oaxaca.

425   The three clades and nearly all of the lineages were distinct in the mtDNA tree as well as

426 the UCE tree (Fig. 2). The mtDNA tree, however, supports a more complicated history for

427 Lineages 3, 4, and 6 in Clade 1 from Guerrero. It is unclear why UCE and mtDNA results

428 differed in this regard, but some reticulate processes might have influenced the mtDNA

429 genomes of these lineages, perhaps, given their close geographic proximity, mtDNA capture of

430 one lineage by another through ancient hybridization (e.g., Bryson Jr et al. 2010).

431   The phylogeny combining our mtDNA data with Genbank data (Fig. 4) is the most

432 taxonomically complete phylogeny of *Sarcohyla* and its sister genus *Plectrohyla* to date and

433 demonstrates the power of linking multilocus species delimitation with mtDNA data by

434 allowing delimited taxa to be placed in a broader context. This revealed that two already-

435 described species are nested within the current *S. bistincta*, lending support to species status of

436 the divergent Clade 1 from the Transvolcanic Belt (which was ~10% divergent in mtDNA from

437 *S. pentheter*). Additional insights afforded by the combined data mtDNA tree include support

438 for a previously hypothesized close relationship between *S. hazelae* and *S. thorectes* (Faivovich et

439 al. 2009), and a sister relationship between *S. mykter* and *S. chryses*. As the genus *Sarcohyla* is

440 very poorly represented by voucher specimens and DNA sequences (Fig. S1), a complete

441 understanding of the history of this genus must await more complete taxonomic and genomic

442 sampling. Unfortunately, there appear to be some microendemic *Sarcohyla* that might have

443 already gone extinct (Lips et al. 2004), especially in the Oaxacan highlands, although recent

444 resurveys give some cause for hope of rediscoveries (Delia et al. 2013).

445   By introducing UCEs as extended DNA barcodes, we are not advocating abandoning

446 traditional mtDNA barcoding and moving solely toward multilocus species delimitation. Even

447 though next-generation sequencing techniques are becoming more standardized and

448 widespread, they still pose a significant financial barrier and require more technical expertise

449 and equipment compared to mtDNA sequencing. However, there is no reason that nuclear

450 genomic DNA barcodes like UCEs cannot develop their own user community and databases in

451 parallel with mtDNA barcoding. As our study shows, mtDNA barcoding and multilocus

452 species delimitation are powerful when deployed together. Where they agree in this study is

453 that there is significant frog diversity in the Mexican Highlands that is currently unrecognized.

454 To the extent that these distinct lineages can be rigorously delimited and compared against

455 existing species, they can be added to biodiversity lists, allowing for their recognition and

456 protection.

Acknowledgments

Literature Cited

462 Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M,
463   Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome
464   annotation. *Molecular Phylogenetics and Evolution*, **69**, 313-319.

465 Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic
466   species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**,
467   148-155.

468 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,
469   Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis.
470   *PLoS Computational Biology*, **10**, e1003537.

471 Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species
472   trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent
473   analysis. *Molecular Biology and Evolution*, **29**, 1917-1932.

Bryson Jr RW, De Oca ANM, Jaeger JR, Riddle BR (2010) Elucidation of cryptic diversity in a widespread Nearctic treefrog reveals episodes of mitochondrial gene capture as frogs diversified across a dynamic landscape. *Evolution*, **64**, 2315-2330.

Camargo A, Morando M, Avila LJ, Sites JW (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834-2849.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics*, **99**, 45-56.

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology*, **25**, 1423-1428.

Collins RA, Cruickshank RH (2014) Known knowns, known unknowns, unknown unknowns and unknown knowns in DNA barcoding: a comment on Dowton et al. *Systematic Biology*, **63**, 1005-1009.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*, **8**, 783-786.

Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, Genson G, Dubois E, Nidelet S, Deuve T, Rasplus J-Y (2014) Empirical assessment of RAD sequencing for interspecific phylogeny. *Molecular Biology and Evolution*, **31**, 1272-1274.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One*, **9**, e106713.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879-886.

Delia JR, Whitney JL, Burkhardt T (2013) Rediscovery of 'lost' treefrogs from the Oaxacan highlands of Mexico. *Biodiversity and Conservation*, **22**, 1405-1414.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491-498.

508 DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**,
509   702-712.
510 DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and
511   DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**,
512   1905-1916.
513 do Amaral FR, Neves LG, Resende Jr MF, Mobili F, Miyaki CY, Pellegrino KC, Biondo C (2015)
514   Ultraconserved elements sequencing as a low-cost source of complete mitochondrial
515   genomes and microsatellite markers in non-model amniotes. *PLoS One*, **10**, e0138446.
516 Dowton M, Meiklejohn K, Cameron SL, Wallman J (2014) A preliminary framework for DNA
517   barcoding, incorporating the multispecies coalescent. *Systematic Biology*, **63**, 639-644.
518 Duellman WE (2001) Hylid frogs of Middle America. Society for the Study of Amphibians and
519   Reptiles, Ithaca, New York.
520 Duellman WE, Marion AB, Hedges SB (2016) Phylogenetics, classification, and biogeography of
521   the treefrogs (Amphibia: Anura: Arboranae). *Zootaxa*, **4104**, 1-109.
522 Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. *Nature*, **434**, 697-
523   697.
524 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
525   throughput. *Nucleic Acids Research*, **32**, 1792-1797.
526 Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography?
527   A comment on. *Molecular Ecology*, **18**, 2930-2933.
528 Faircloth BC (2015) PHYLUCE is a software package for the analysis of conserved genomic loci.
529   *Bioinformatics*, **32**, 786-788.
530 Faircloth BC, Branstetter MG, White ND, Brady SG (2015) Target enrichment of ultraconserved
531   elements from arthropods provides a genomic perspective on relationships among
532   Hymenoptera. *Molecular Ecology Resources*, **15**, 489-501.
533 Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: designing and validating
534   sequence identification tags robust to indels. *PLoS One*, **7**, e42543.
535 Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012)
536   Ultraconserved elements anchor thousands of genetic markers spanning multiple
537   evolutionary timescales. *Systematic Biology*, **61**, 717-726.
538 Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013) A phylogenomic perspective on the
539   radiation of ray-finned fishes based upon targeted sequencing of ultraconserved
540   elements (UCEs). *PLoS One*, **8**, e65923.
541 Faivovich J, Haddad CF, Garcia PC, Frost DR, Campbell JA, Wheeler WC (2009) Systematic
542   review of the frog family Hylidae, with special reference to Hylinae: phylogenetic

543   analysis and taxonomic revision. *Bulletin of the American Museum of Natural History*, **294**, 1-240.

545   Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 480-488.

547   Galtier N, Nabholz B, Glémin S, Hurst G (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541-4550.

549   Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119-133.

552   Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494-1512.

555   Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129.

558   Harris RS (2007) Improved pairwise alignment of genomic DNA. Pennsylvania State University, PhD Thesis.

560   Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313-321.

562   Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.

566   Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, **18**, 597-603.

569   Hime PM, Hotaling S, Grewelle RE, O'Neill EM, Voss SR, Shaffer HB, Weisrock DW (2016) The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, **25**, 5959-5974.

572   Isaac NJ, Mallet J, Mace GM (2004) Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, **19**, 464-469.

574   Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-1649.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887-895.

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, **105**, 2923-2928.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534-542.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Lips KR, Mendelson Iii JR, Munoz-Alonso A, Canseco-Márquez L, Mulcahy DG (2004) Amphibian population declines in montane southern Mexico: resurveys of historical localities. *Biological Conservation*, **119**, 555-564.

Mace GM (2004) The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**, 711-719.

McCartney-Melstad E, Mount GG, Shaffer HB (2016) Exon capture optimization in amphibians with large genomes. *Molecular Ecology Resources*, **16**, 1084-1094.

McCormack J, Tsai WLE, Faircloth BC (2016) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189-1203.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research*, **22**, 746-754.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.

Miralles A, Vences M (2013) New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in Madascincus lizards. *PLoS One*, **8**, e68242.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**, e354.

O'Meara BC (2009) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59-73.

Olmo E (1991) Genome variations in the transition from amphibians to reptiles. *Journal of Molecular Evolution*, **33**, 68-75.

614 Pie MR, Bornschein MR, Ribiero LF, Faircloth BC, McCormack JE (2017) Phylogenomic species
615   delimitation in microendemic frogs Of the Brazilian Atlantic forest. *BioRxiv*, doi:
616   https://doi.org/10.1101/143735.

617 Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin
618   WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of
619   undescribed insects. *Systematic Biology*, **55**, 595-609.

620 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
621   genotype data. *Genetics*, **155**, 945-959.

622 Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler
623   for short read sequence data. *Genome Research*, **19**, 1117-1123.

624 Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and
625   massively parallel sequencing of ultraconserved elements for comparative studies at
626   shallow evolutionary time scales. *Systematic Biology*, **63**, 83-95.

627 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
628   large phylogenies. *Bioinformatics*, **30**, 1312-1313.

629 Starrett J, Derkarabetian S, Hedin M, Bryson RW, McCormack JE, Faircloth BC (2016) High
630   phylogenetic utility of an ultraconserved element probe set designed for Arachnida.
631   *Molecular Ecology Resources*, DOI: 10.1111/1755-0998.12621.

632 Stuart SN, Hoffmann M, Chanson J, Cox N, Berridge R, Ramani P, Young B (2008) Threatened
633   amphibians of the world. Lynx Edicions, Barcelona, Spain.

634 Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species.
635   *Proceedings of the National Academy of Sciences*, **114**, 1607-1612.

636 Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G,
637   Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast trnL (UAA)
638   intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, e14-e14.

639 Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy.
640   *Trends in Ecology & Evolution*, **18**, 70-74.

641 van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan
642   T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant
643   calls: the genome analysis toolkit best practices pipeline. *Current Protocols in*
644   *Bioinformatics*, **11: 11.10**, 11.10. 11-11.10. 33.

645 Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005) DNA barcoding Australia's fish
646   species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **360**,
647   1847-1857.

648   Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic*
649   *Biology*, **56**, 875-878.
650   Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for
651   integrative taxonomy. *Systematic Biology*, **54**, 844-851.
652   Witt JD, Threloff DL, Hebert PD (2006) DNA barcoding reveals extraordinary cryptic diversity
653   in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*,
654   **15**, 3073-3082.
655   Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data.
656   *Proceedings of the National Academy of Sciences*, **107**, 9264-9269.
657   Ye L, Zhu C, Yu D, Zhang Y, Zhang J (2016) The complete mitochondrial genome of *Hyla*
658   *annectans* (Anura: Hylidae). *Mitochondrial DNA Part A*, **27**, 1593-1594.
659   Zarza E, Faircloth BC, Tsai WL, Bryson RW, Klicka J, McCormack JE (2016) Hidden histories of
660   gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, **25**, 5144-
661   5157.

662

663

664 **Tables**

665 Table 1. Specimen information and summary statistics for *Sarcohyla bistincta* and closely related

666 species.

Map Number[1]	Field Number[2]	Catalog Number[3]	Current Taxonomy	State	Latitude	Longitude	Trimmed Paired Reads	UCEs	Average UCE Length	SNP Read Depth	mtDNA Reads	mtDNA Average Coverage	mtDNA Average Quality
1	MK 618	UMMZ 239796	*S. bistincta*	MICHOACAN	19.7911	-100.6605	1,889,670	2,323	529.7	31.8	3,381	20.8	86
2	MK 627-31	UMMZ 239683	*S. bistincta*	MICHOACAN	19.4266	-102.0736	1,933,019	2,254	370.9	21.6	3,415	20.8	85
3	MK 666	UMMZ 239745	*S. bistincta*	MICHOACAN	19.3452	-100.3128	2,186,871	2,350	621.3	33.9	3,900	23.4	83
4	MK 600	UMMZ 239679	*S. bistincta*	MEXICO	19.1501	-100.1469	932,448	2,174	500.8	24.4	2,377	15.7	83
4	MK 600 (1)	UMMZ 239678	*S. bistincta*	MEXICO	19.1501	-100.1469	1,217,987	2,267	532.8	21.9	2,570	16.4	84
5	MK 645	UMMZ 239749	*S. bistincta*	MORELOS	18.9224	-99.2442	2,217,054	2,406	532.5	34.2	3,494	21.8	86
5	MK 645 dup	UMMZ 239749	*S. bistincta*	MORELOS	18.9224	-99.2442	548,545	1,964	513.0	17.7	841	7.4	60
6	MK 759	UMMZ 239701	*S. bistincta*	GUERRERO	18.0013	-101.1716	1,224,835	2,199	557.6	25.0	2,522	16.2	84
6	MK 760	UMMZ 239705	*S. bistincta*	GUERRERO	18.0013	-101.1716	1,074,315	2,203	559.4	24.3	743	6.9	58
6	MK 760 (2)	UMMZ 239704	*S. bistincta*	GUERRERO	18.0013	-101.1716	927,774	2,246	579.2	26.3	906	9.4	6
7	MK 691 (5)	UMMZ 239744	*S. bistincta*	GUERRERO	17.5324	-99.8994	2,084,203	2,353	672.5	27.2	1,678	11.9	80
8	MK 650 (1)	UMMZ 239725	*S. bistincta*	GUERRERO	17.6843	-99.8034	941,101	2,124	556.6	22.6	1,278	9.7	74
8	MK 650 (2)	UMMZ 239726	*S. bistincta*	GUERRERO	17.6843	-99.8034	2,224,898	2,394	506.4	28.4	1,737	12.2	79
8	MK 652	UMMZ 239727	*S. bistincta*	GUERRERO	17.6843	-99.8034	3,423,330	2,444	526.0	29.8	1,396	10.4	74
9	MK 671 (4)	UMMZ 239733	*S. bistincta*	GUERRERO	17.6407	-99.6797	1,012,300	2,107	557.0	22.5	842	7.4	62
9	MK 672	UMMZ 239738	*S. bistincta*	GUERRERO	17.6407	-99.6797	297,782	1,667	439.8	11.9	400	5.1	41
10	MK 656 (1)	UMMZ 239729	*S. bistincta*	GUERRERO	17.5526	-99.6626	580,194	1,950	497.6	18.4	547	5.9	42
11	MK 674 (1)	UMMZ 239690	*S. bistincta*	GUERRERO	17.5087	-99.1258	659,418	1,941	565.4	18.8	1,086	8.7	68
11	MK 675 (2)	UMMZ 239879	*S. bistincta*	GUERRERO	17.5087	-99.1258	336,474	1,565	519.1	10.4	227	4.2	24
12	MK 662	UMMZ 239731	*S. bistincta*	GUERRERO	18.6359	-99.6491	1,137,742	2,130	548.7	23.5	796	7.2	56
13	MK 697 (3)	UMMZ 239789	*S. bistincta*	VERACRUZ	18.6585	-97.1574	1,524,182	2,214	625.6	24.3	2,479	16.0	82
13	MK 699 (1)	UMMZ 239791	*S. bistincta*	VERACRUZ	18.6477	-97.1574	1,342,010	2,206	610.9	24.4	1,769	12.3	79
14	MK 700 (2)	UMMZ 239750	*S. bistincta*	PUEBLA	18.3220	-97.0285	2,203,360	2,298	697.8	27.6	4,155	24.8	86
15	MK 705 (1)	UMMZ 239862	*S. bistincta*	OAXACA	18.1576	-96.8684	2,529,703	2,417	644.6	30.3	3,164	19.6	85
16	MK 715	UMMZ 239755	*S. bistincta*	OAXACA	17.2390	-97.0032	347,152	1,712	437.7	12.1	254	4.3	31
17	MK 716 (1)	UMMZ 239758	*S. bistincta*	OAXACA	17.3036	-96.7930	52,461	594	267.4	2.7	8	3.0	1
18	MK 718 (2)	UMMZ 239765	*S. bistincta*	OAXACA	17.4211	-96.6876	606,978	2,014	568.6	19.7	366	4.9	37
19	MK 755 (1)	UMMZ 239786	*S. bistincta*	OAXACA	17.4153	-96.5671	871,034	2,062	563.2	22.0	1,022	8.4	69
20	MK 751	UMMZ 239785	*S. bistincta*	OAXACA	17.3160	-96.4435	1,577,190	2,342	621.5	28.5	2,376	15.4	84
22	MK 748 (2)	UMMZ 239780	*S. bistincta*	OAXACA	16.9791	-96.1364	91,912	1,009	302.5	4.2	23	3.1	3
22	MK 748 (4)	UMMZ 239782	*S. bistincta*	OAXACA	16.9791	-96.1364	1,058,426	2,170	553.8	23.7	499	5.7	42
22	MK 767	UMMZ 239788	*S. bistincta*	OAXACA	16.9859	-96.1358	1,083,058	2,206	524.7	24.3	617	6.3	52
23	MK 721	UMMZ 239767	*S. bistincta*	OAXACA	16.7377	-97.0384	146,471	1,215	376.0	6.1	62	3.3	10
24	MK 766	UMMZ 239794	*S. bistincta*	OAXACA	16.2522	-97.1536	994,521	2,166	577.4	23.3	467	5.5	43
25	MK 730 (2)		*P. bistincta*	OAXACA	16.0186	-96.5301	2,967,630	2,404	716.9	31.1	3,581	21.7	84
26	MK 685 (2)	UMMZ 239739	*S. sp. nov.*	GUERRERO	17.3812	-100.2009	269,729	1,645	450.6	10.8	184	4.0	20
27	MK 689 (2)	UMMZ 239740	*S. sp. nov.*	GUERRERO	17.3000	-100.2792	67,313	502	221.6	2.6	32	3.1	6
28	MK 727 (2)	UMMZ 239772	*S. pentheter*	OAXACA	16.1916	-97.0958	17,052	381	247.1	2.4	10	3.0	2
29	MK 691 (3)	UMMZ 239651	*S. chryses*	GUERRERO	17.5324	-99.8994	2,267,946	2,347	663.6	35.0	4,919	33.4	83
30	MK 770	UMMZ 239802	*S. hazelae*	OAXACA	17.2216	-96.5839	538,500	2,066	526.2	19.8	557	5.9	44

[1] Map number in Figure 1

[2] The first three-digit number corresponds to a sampling location. If there is a second number in parentheses, this corresponds to different aquaria where tadpoles were sorted by species before

[3] All specimens are from the University of Michigan Museum of Zoology

667

668 Table 2. Bayes factor species delimitation results.

| Model	Species	MLE	Rank	*ln*(BF)
runA	9	-2832.33	1	10.52
runB	10	-2837.59	2	132.52
runC	7	-2903.85	3	69.36
runD	6	-2938.53	4	3822
runE	2	-4849.53	5	-

669

670

27

671 **Figures**

672 Figure 1. (A) Map of central Mexico showing sampling sites for *S. bistincta* and close outgroups,

673 with numbers matching localities listed in Table 1 and colors matching Structure results below.

674 Unsampled parts of the distribution of *S. bistincta* are shown in Fig. S1; (B) *S. bistincta* individual

675 from near site 1; (C) Composite results of repeated Structure runs at $K = 2$ showing the finest

676 detectable structure in the genetic data. Each vertical line represents an individual labeled with

677 its UMMZ catalog number above and, in descending order below, the site number and the

678 lineages indicated by the species delimitation analysis.
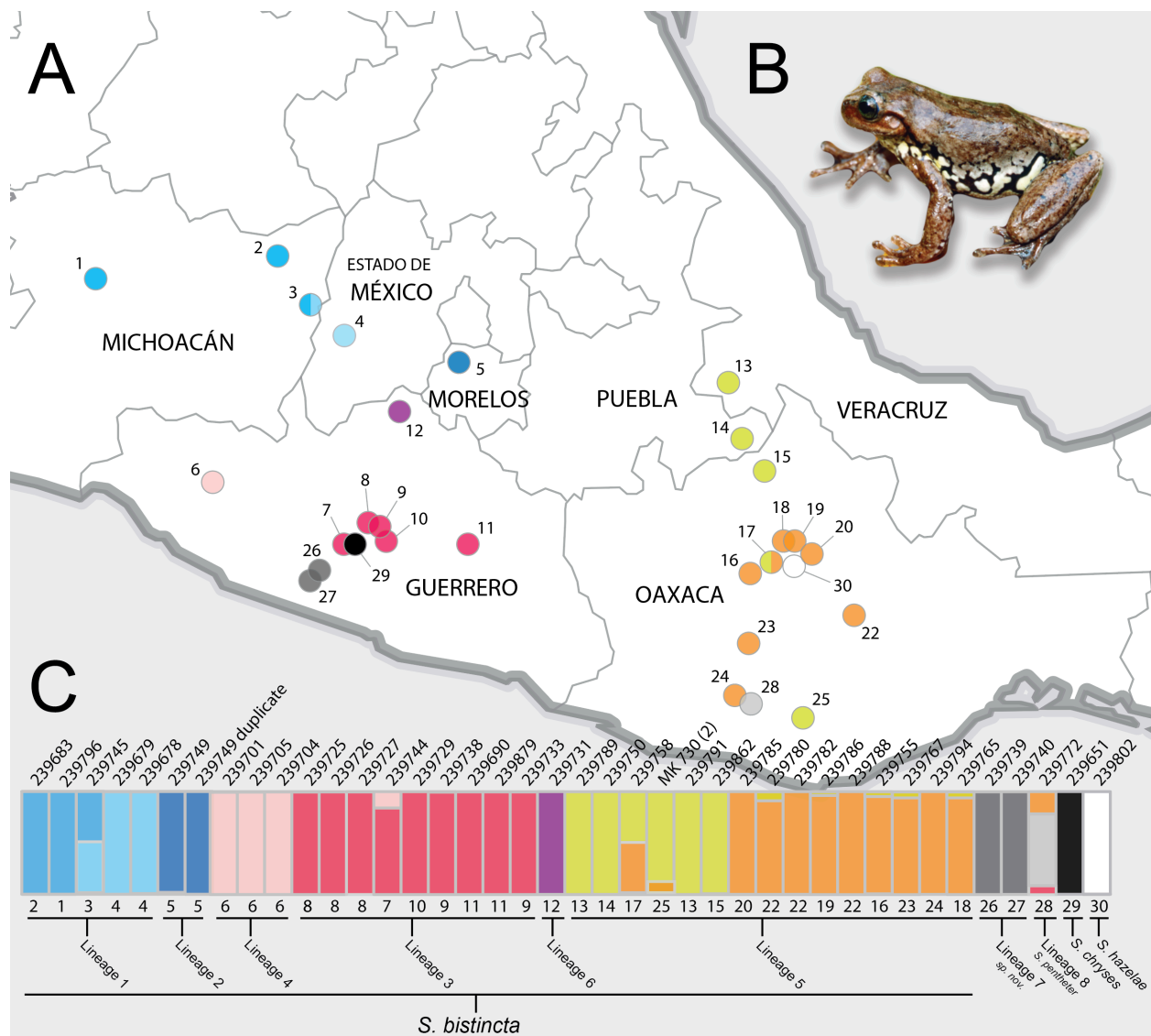
679

680   Figure 2. (A) UCE tree; (B) mtDNA tree. Colors match Structure groups identified in Fig. 1. Tips

681   are labeled with their UMMZ catalog number.

682


29
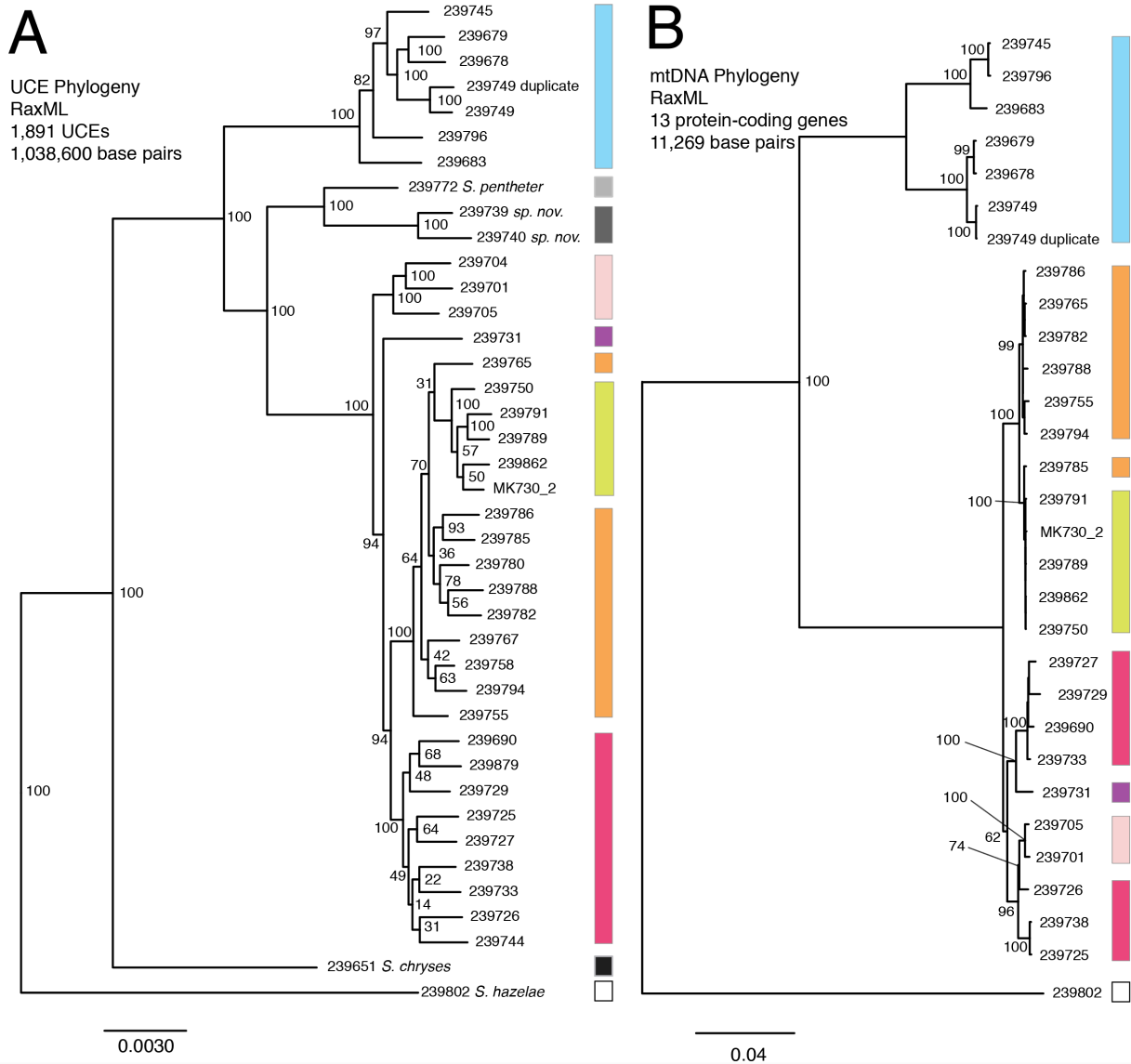
683   Figure 3. Cloudogram of the posterior distribution of SNAPP trees from 399 high-quality SNPs

684   mined from UCE loci. Tip labels are UMMZ catalog numbers. Colored bars show the eight

685   lineages identified by the species delimitation analysis in addition to the outgroup *S. chryses*.

686   Clades are discussed in text. Colors match genetic clusters from Figure 1.
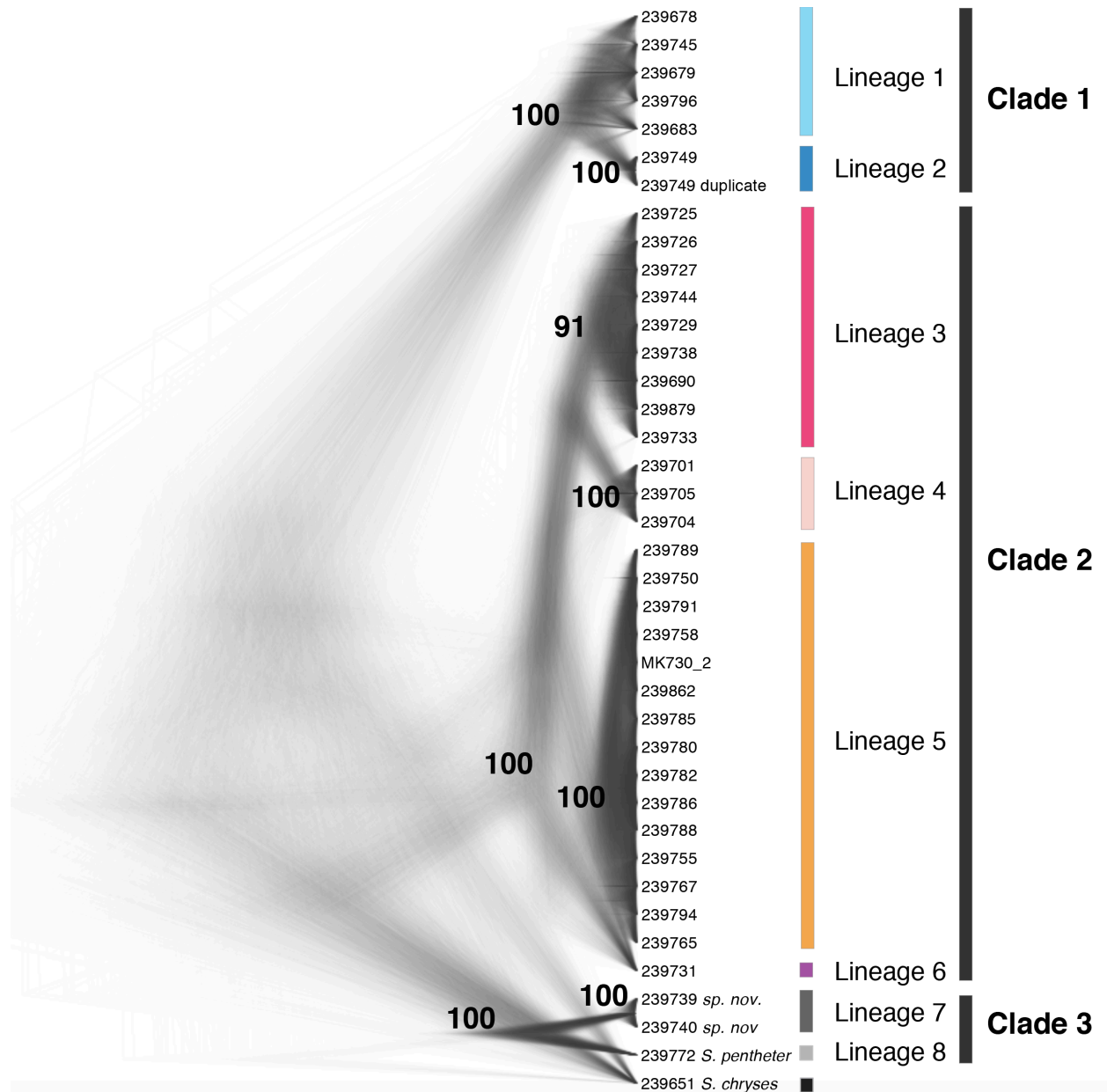
687

688   Figure 4. Phylogeny of mtDNA for a subset of the individuals from this study (labeled with red

689   UMMZ catalog numbers) combined with existing sequences from Genbank (labeled with

690   accession numbers). Colored boxes relate to genetic lineages in prior figures. The tree was

691   rooted with *Exerodonta xera*. Photo credit: Peter Heimes.
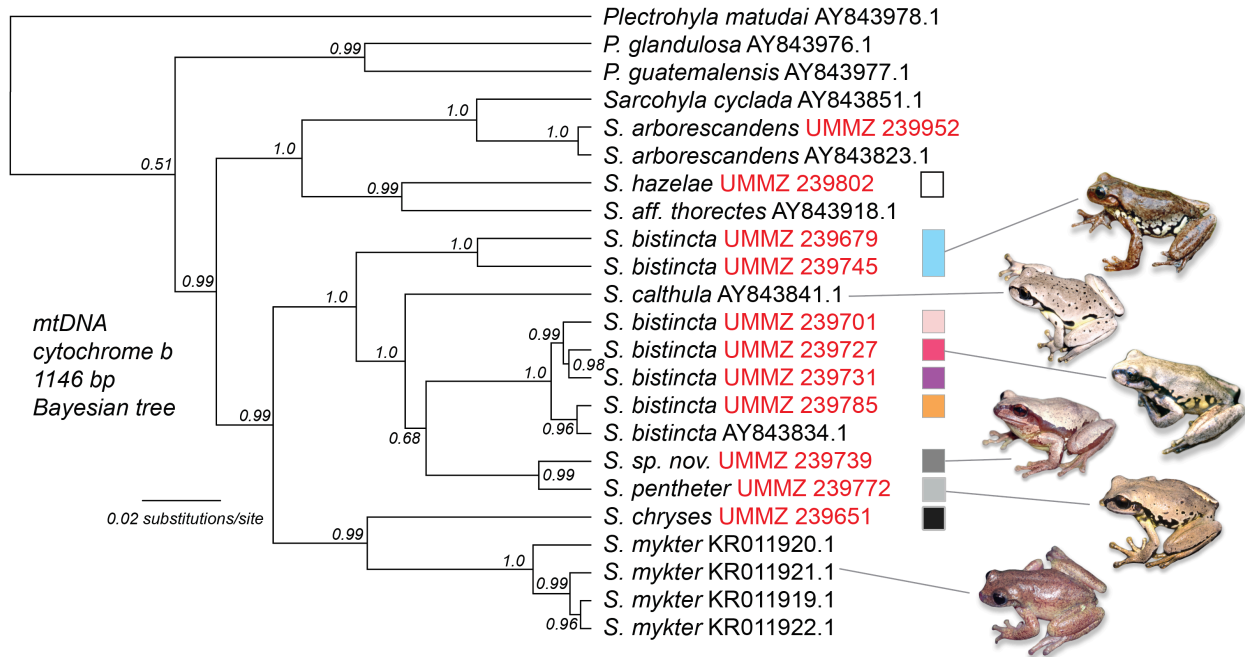

692

693

694 Table S1. Information and summary statistics on all 45 samples used to determine the ingroup

695 for this study.

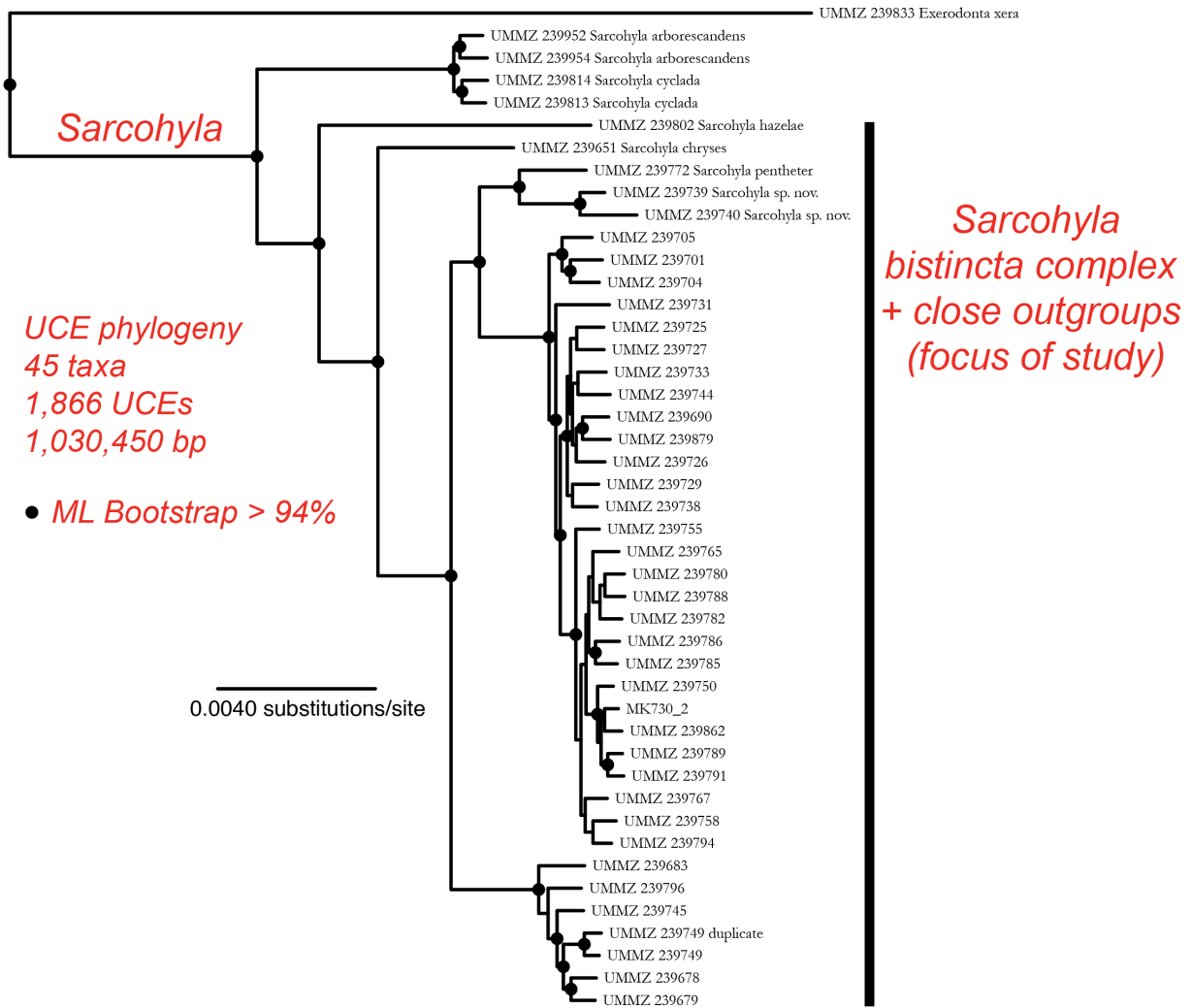| Field Number	UMMZ Number	Current Taxonomy	State	Locality	Latitude	Longitude	Trimmed Paired Reads	UCEs	Average UCE length	mtDNA reads	mtDNA average coverage	mtDNA average quality	Fig. 1 Map Number	Accession number
MK 618	UMMZ 239796	*Sarcohyla bistincta*	MICHOACAN	LOS AZUFRES / SAN PEDRO ROAD	19.791051	-100.660542	1,889,670	2323	529.7	3381	20.8	86	1	
MK 627-31	UMMZ 239683	*Sarcohyla bistincta*	MICHOACAN	PARQUE BARRANCA DEL CUPATITZIO URUAPAN	19.426646	-102.073574	1,933,019	2254	370.9	3415	20.8	85	2	
MK 666	UMMZ 239745	*Sarcohyla bistincta*	MICHOACAN	13 KMS ROAD ZITACUARO MACHERO (LITTLE TOWN WERE THERE IS A MONARCH BUTTERFLY SANCTUARY)	19.3452	-100.3128	2,186,871	2350	621.3	3900	23.4	83	3	
MK 600	UMMZ 239679	*Sarcohyla bistincta*	MEXICO	CARRETERA VALLE DE BRAVO-SAN PEDRO TENEYAC ARROYO 1 (6860 FEET) (POR ALBARRADA) BOSQUE BIEN CONSERVADO Y HUMEDO	19.150083	-100.1469	932,448	2174	500.8	2377	15.7	83	4	
MK 600 (1)	UMMZ 239678	*Sarcohyla bistincta*	MEXICO	CARRETERA VALLE DE BRAVO-SAN PEDRO TENEYAC ARROYO 1 (6860 FEET) (POR ALBARRADA) BOSQUE BIEN CONSERVADO Y HUMEDO	19.150083	-100.1469	1,217,987	2267	532.8	2570	16.4	84	4	
MK 645	UMMZ 239749	*Sarcohyla bistincta*	MORELOS	2 KMS OUT OF CUERNAVACA ON THE CUERNAVACA CHALMA RD. (PASSING CONJUNTO CERRADO) OF LA BARRANCA WHERE THE SALTO SAN ANTONIO IS	18.922402	-99.244151	2,217,054	2406	532.5	3494	21.8	86	5	
MK 645 dupe	UMMZ 239749	*Sarcohyla bistincta*	MORELOS	2 KMS OUT OF CUERNAVACA ON THE CUERNAVACA CHALMA RD. (PASSING CONJUNTO CERRADO) OF LA BARRANCA WHERE THE SALTO SAN ANTONIO IS	18.922402	-99.244151	548,585	1964	513.0	841	7.4	60	5	
MK 759	UMMZ 239701	*Sarcohyla bistincta*	GUERRERO	CA 100 DEG KMS FROM CIUDAD ALTAMIRANO VIA IXTAPA ZIHUATANEJO	18.0013	-101.1716	1,224,835	2199	557.6	2522	16.2	84	6	
MK 760	UMMZ 239705	*Sarcohyla bistincta*	GUERRERO	CA 100 DEG KMS FROM CIUDAD ALTAMIRANO VIA IXTAPA ZIHUATANEJO	18.0013	-101.1716	1,074,315	2203	559.4	743	6.9	58	6	
MK 760 (2)	UMMZ 239704	*Sarcohyla bistincta*	GUERRERO	CA 100 DEG KMS FROM CIUDAD ALTAMIRANO VIA IXTAPA ZIHUATANEJO	18.0013	-101.1716	927,774	2246	579.2	906	9.4	6	6	
MK 691 (5)	UMMZ 239744	*Sarcohyla bistincta*	GUERRERO	2-3 KMS IN THE ROAD TO JALEACA FROM POINT WHERE ROAD TRIFURCATE TO PTO. DEL GALLO / YEXTLA / ANDJALEACA.COMMING FROM CARRIZAL DE BRAVO / IN RIVER UNDER THE BRIDGE	17.5324	-99.8994	2,084,203	2353	672.5	1678	11.9	80	7	
MK 650 (1)	UMMZ 239725	*Sarcohyla bistincta*	GUERRERO	MOUNTAINS W. OF CHILPANCINGO / TOWN LOS MORROS	17.6843	-99.80339	941,101	2124	556.6	1278	9.7	74	8	
MK 650 (2)	UMMZ 239726	*Sarcohyla bistincta*	GUERRERO	MOUNTAINS W. OF CHILPANCINGO / TOWN LOS MORROS	17.6843	-99.80339	2,224,898	2394	506.4	1737	12.2	79	8	
MK 652	UMMZ 239727	*Sarcohyla bistincta*	GUERRERO	MOUNTAINS W. OF CHILPANCINGO / TOWN LOS MORROS	17.6843	-99.80339	3,423,330	2444	526.0	1396	10.4	74	8	
MK 671 (4)	UMMZ 239733	*Sarcohyla bistincta*	GUERRERO	BEHIND CHICHIHUALCO / ON ROAD TO CARRIZAL DE BRAVO / 2 KMS FROM ENTRONQUE	17.6407	-99.6797	1,012,300	2107	557.0	842	7.4	62	9	
MK 672	UMMZ 239738	*Sarcohyla bistincta*	GUERRERO	BEHIND CHICHIHUALCO / ON ROAD TO CARRIZAL DE BRAVO / 2 KMS FROM ENTRONQUE	17.6407	-99.6797	297,782	1667	439.8	400	5.1	41	9	
MK 656 (1)	UMMZ 239729	*Sarcohyla bistincta*	GUERRERO	3 KMS FROM THE TOWN OF OMILTEMI IN THE ROAD OMILTEMI CHILPANCINGO	17.552603	-99.662569	580,194	1950	497.6	547	5.9	42	10	
MK 674 (1)	UMMZ 239690	*Sarcohyla bistincta*	GUERRERO	ON ATZACUALOYA HUEYCATENANGO RD.	17.5087	-99.1258	659,418	1941	565.4	1086	8.7	44	11	
MK 675 (2)	UMMZ 239879	*Sarcohyla bistincta*	GUERRERO	ON ATZACUALOYA HUEYCATENANGO RD.	17.5087	-99.1258	336,474	1565	519.1	227	4.2	24	11	
MK 662	UMMZ 239731	*Sarcohyla bistincta*	GUERRERO	1.6 KMS FROM THE TOWN OF TETIPAC ON THE TETIPAC TAXCO ROAD (MAYBE ARROYO LAS DAMAS)	18.635895	-99.6491	1,137,742	2130	548.7	796	7.2	56	12	
MK 697 (3)	UMMZ 239789	*Sarcohyla bistincta*	VERACRUZ	ON ATZOMPA XOXOCOTLA RD. 1.5 KMS FROM XOXOCOTLA	18.6585	-97.1574	1,524,182	2214	625.6	2479	16.0	82	13	
MK 699 (1)	UMMZ 239791	*Sarcohyla bistincta*	VERACRUZ	ON STREAM CROSSING THE TOWN OF XOXOCOTLA	18.6477	-97.1574	1,342,010	2206	610.9	1769	12.3	79	13	
MK 700 (2)	UMMZ 239750	*Sarcohyla bistincta*	PUEBLA	IN THE STREAM LOCATED AFTER THE TOWN OF ZOQUITLAN TURNING DOWN AT THE CENTRO DE SALUD	18.322	-97.0285	2,203,360	2298	697.8	4155	24.8	86	14	
MK 705 (1)	UMMZ 239862	*Sarcohyla bistincta*	OAXACA	56 KMS FROM TEOTITLAN VIA HUAUTLA	18.1576	-96.8684	2,529,703	2417	644.6	3164	19.6	85	15	
MK 715	UMMZ 239755	*Sarcohyla bistincta*	OAXACA	NEAR "EL TEJOCOTE" (FRENTE AL KINDER)	17.239	-97.0032	347,152	1712	437.7	254	4.3	31	16	
MK 716 (1)	UMMZ 239758	*Sarcohyla bistincta*	OAXACA	ON ROAD BETWEEN THE TOWNS OF SAN JUAN DEL ESTADO AND SAN MIGUEL ALOAPAN	17.3036	-96.793	52,461	594	267.4	8	3.0	1	17	
MK 718 (2)	UMMZ 239765	*Sarcohyla bistincta*	OAXACA	PASSING SAN MIGUEL ALOAPAN	17.4211	-96.6876	606,978	2014	568.6	366	4.9	37	18	
MK 755 (1)	UMMZ 239786	*Sarcohyla bistincta*	OAXACA	ON ROAD SAN JUAN ATEPEC SAN MIGUEL ABEJONES	17.4153	-96.5671	871,034	2062	563.2	1022	8.4	69	19	
MK 751	UMMZ 239785	*Sarcohyla bistincta*	OAXACA	3.8 KMS PASSING "RANCHO TEXAS" ON THE ROAD FROM THE TOWN OF IXTLAN DE JUAREZ	17.316	-96.4435	1,577,190	2342	621.5	2376	15.4	84	20	
MK 748 (2)	UMMZ 239780	*Sarcohyla bistincta*	OAXACA	CA 37 KMS FROM MITLA ON THE ROAD MITLA AYUTLA	16.9791	-96.1364	91,912	1009	302.5	23	3.1	3	22	
MK 748 (4)	UMMZ 239782	*Sarcohyla bistincta*	OAXACA	CA 37 KMS FROM MITLA ON THE ROAD MITLA AYUTLA	16.9791	-96.1364	1,058,426	2170	553.8	499	5.7	42	22	
MK 767	UMMZ 239788	*Sarcohyla bistincta*	OAXACA	9.2 KMS E STA MARIA ALBARRADAS / SIERRA MIXE	16.985888	-96.135816	1,058,426	2206	524.7	617	6.3	52	22	
MK 721	UMMZ 239767	*Sarcohyla bistincta*	OAXACA	2 KMS N. THE TOWN OF STA. MARIA LAXICHIO VIA THE TOWN OF SAN SEBASTIAN RIO DULCE	16.7377	-97.0384	146,471	1215	376.0	62	3.3	10	23	
MK 766	UMMZ 239794	*Sarcohyla bistincta*	OAXACA	CERRO DE VIDRIO VIA A PUERTO ESCONDIDO	16.25216	-97.15359	994,521	2166	577.4	467	5.5	43	24	
MK 730 (2)		*Sarcohyla bistincta*	OAXACA	SLIGHTLY NORTH OF THE TOWN JALATENGO ON HWY 175	16.0186	-96.5301	2,967,630	2404	716.9	3581	21.7	84	25	
MK 685 (2)	UMMZ 239739	*Sarcohyla sp. nov.*	GUERRERO	ON ATOYAC PTO DE GALLO RD / BETWEEN 10 - 20 KMS NORTH OF THE TOWN "EL PARAISO"	17.3812	-100.2009	269,729	1645	450.6	184	4.0	20	26	
MK 689 (2)	UMMZ 239740	*Sarcohyla sp. nov.*	GUERRERO	ON ATOYAC PTO DEL GALLO RD. / 500 M. NORTH OF THE TOWN OF SAN VICENTE	17.3	-100.2792	67,313	502	221.6	32	3.1	6	27	
MK 727 (2)	UMMZ 239772	*Sarcohyla pentheter*	OAXACA	RIO "EL SALADO" / 8 KMS N SAN JUAN LACHAO ON HWY 135	16.1916	-97.0958	17,052	381	247.1	10	3.0	2	28	
MK 691 (3)	UMMZ 239651	*Plectrohyla chryses*	GUERRERO	2-3 KMS IN THE ROAD TO JALEACA FROM POINT WHERE ROAD TRIFURCATE TO PTO. DEL GALLO / YEXTLA / ANDJALEACA.COMMING FROM CARRIZAL DE BRAVO / IN RIVER UNDER THE BRIDGE	17.5324	-99.8994	2,267,946	2347	663.6	4919	33.4	83	29	
MK 770	UMMZ 239802	*Plectrohyla hazelae*	OAXACA	EL PUNTO SIERRA JUAREZ	17.22156	-96.58386	538,500	2066	526.2	557	5.9	44	30	
MK 667	UMMZ 239952	*Sarcohyla arborescandens*	VERACRUZ	PUERTO DEL AIRE (ARRIBA DE ALCUTZINGO)	18.6787	-97.3485	923,330	2116	511.8	1154	9.0	69		
MK 700 (1)	UMMZ 239813	*Sarcohyla cyclada*	PUEBLA	IN THE STREAM LOCATED AFTER THE TOWN OF ZOQUITLAN TURNING DOWN AT THE CENTRO DE SALUD	18.322	-97.0285	2,490,616	2385	661.4	2961	18.4	82		
MK 701	UMMZ 239814	*Sarcohyla cyclada*	OAXACA	24 KMS FROM THE TOWN OF TEOTITLAN DE FLORES MAGON VIA HUAUTLA DE JIMENEZ	18.1781	-97.0054	1,166,574	2129	565.5	1061	8.7	69		
MK 742 (1)	UMMZ 239954	*Sarcohyla arborescandens*	VERACRUZ	LEFT ROAD BIFURCATING FROM ROAD TO THE TOWN OF "LAS MINAS"	19.6758	-97.1751	1,074,225	2240	553.8	1750	12.0	79		
MK 768	UMMZ 239833	*Exerodonta xera*	PUEBLA	5 KM SW ZAPOTITLAN DE SALINAS	18.311958	-97.51266	1,257,938	2288	581.5	3506	21.4	84		

696

697 Fig. S1. Sampled and unsampled parts of *S. bistincta* range in relation to known distributions (or

698 localities, where distributional information is lacking) of other *Sarcohyla* species.

699
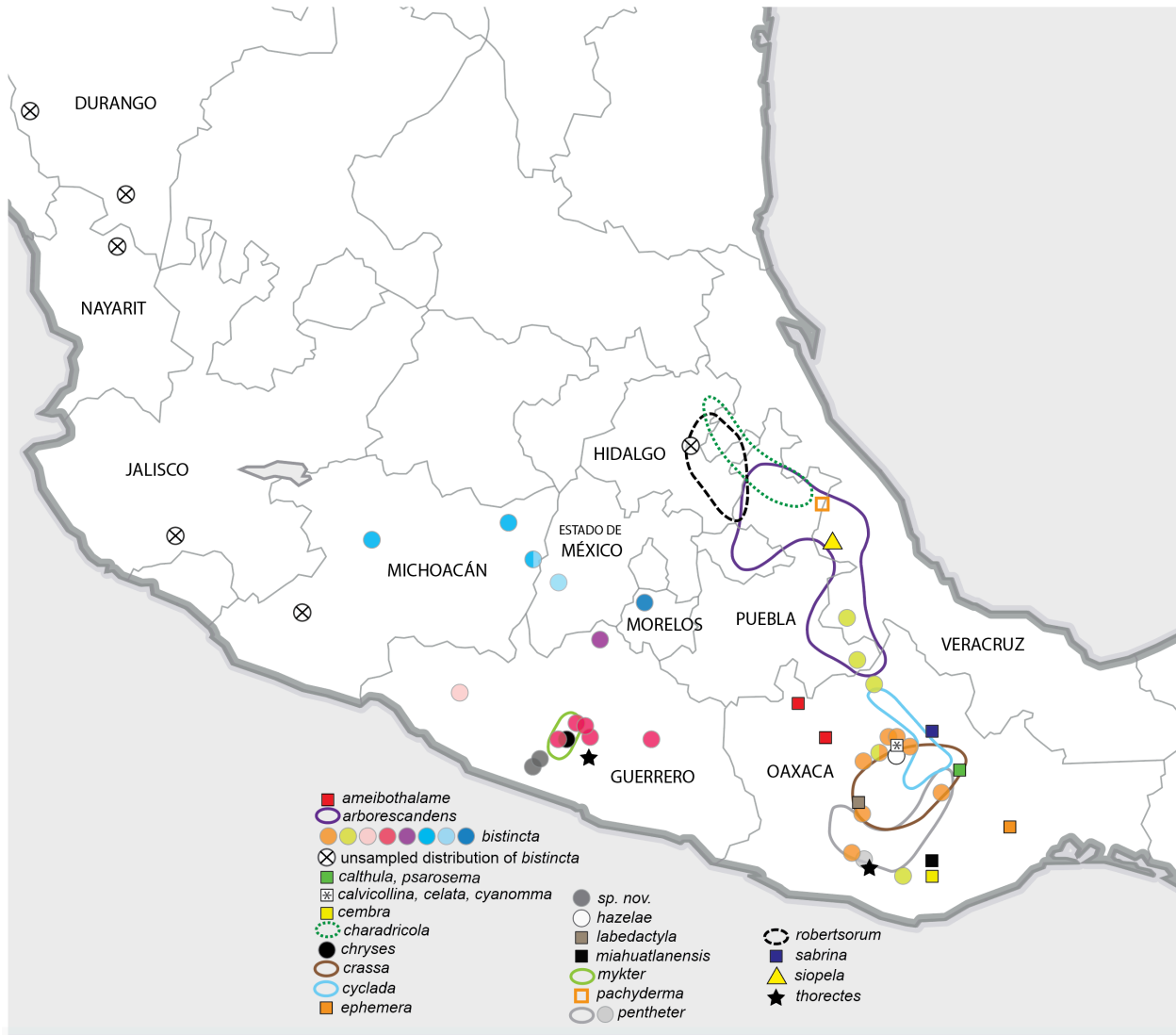
700   Fig. S2. UCE tree of 45 samples of *Sarcohyla* and outgroup *Exerodonta xera* used to determine the

701   ingroup.

702