Combining next-generation sequencing and mtDNA data to uncover cryptic lineages of Mexican highland frogs

Eugenia Zarza[1], Elizabeth M. Connors[1], James M. Maley[1], Whitney L.E. Tsai[1], Peter Heimes[2], Moises Kaplan[3] & John E. McCormack[1,*]

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

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

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

EZ email: eugenia.zarza@gmail.com

EMC email: beth.connors.94@gmail.com

JMM email: jmaley@oxy.edu

WLET email: wtsai@oxy.edu

PH email: heimesp@hotmail.com

MK email: moises.kaplan@gmail.com

*Corresponding author: mccormack@oxy.edu

Abstract

Recently, molecular studies have uncovered significant cryptic diversity in the Mexican Highlands, leading to the description of many new endemic species. DNA approaches to this kind of species discovery have included both mitochondrial DNA (mtDNA) sequencing and multilocus genomic methods. While these marker types have often been pitted against one another, there are benefits to deploying them together, as linked mtDNA data can provide the bridge between uncovering lineages through rigorous multilocus genomic analysis and identifying lineages through comparison to existing mtDNA databases. Here, we apply one class of multilocus genomic marker, ultraconserved elements (UCEs), and linked mtDNA data to a species complex of frogs (*Sarcohyla bistincta*) found in the Mexican Highlands. We generated data from 1,891 UCEs, which contained 1,742 informative SNPs for *S. bistincta* and closely related species and captured mitochondrial genomes for most samples. Genetic analyses based on both whole loci and SNPs agree there are numerous distinct and divergent lineages within *S. bistincta*. The SNP-based species tree provides the most conservative estimate of 8 well-supported lineages in three major clades. Having linked mtDNA data allowed us to tap into the large number of mtDNA sequences available on GenBank and identify one of these lineages as an already-described species, *S. pentheter*. One identified clade (containing 2 of the 8 lineages) was 10% divergent in mtDNA and paraphyletic with other *S. bistincta*, making this clade a clear candidate for species status. Phylogenies from UCEs and mtDNA mostly agreed in their topologies, but differed in that mtDNA suggested a more complex evolutionary history perhaps influenced by gene flow between some neighboring lineages. Our study demonstrates that the Mexican Highlands still hold substantial undescribed diversity. Combining multilocus genomic data with linked mtDNA data is a useful approach for identifying potential new species and associating them with already described taxa, which is especially important in groups with undescribed subadult phenotypes, where geographic ranges are unclear, or where phenotypes are conserved.

Keywords: ultraconserved elements, genomics, population genetics, phylogeography, phylogenetics, systematics, species limits, Hylinae

Introduction

The Mexican Highlands are a global biodiversity hotspot (Myers et al. 2000). Recent molecular studies have uncovered significant cryptic diversity in the Mexican Highlands, leading to the description of new endemic species or the elevation of former subspecies to species rank (Bryson et al. 2010; Bryson et al. 2012; Bryson et al. 2011; McCormack et al. 2008; Rovito et al. 2015). At the same time, habitat loss threatens many of these species before they have even been described (Ponce-Reyes et al. 2012). Amphibians are particularly sensitive to habitat alterations, and many are threatened by habitat loss and invasive diseases (Stuart et al. 2008). Because most amphibians have reduced dispersal, they also show patterns of microendemism (Parra-Olea et al. 2014), meaning that important pockets of diversity and new species are still being uncovered in the Mexican Highlands (Campbell et al. 2009; Meik et al. 2006).

We investigate potential cryptic diversity within the *Sarcohyla bistincta* species complex in the Mexican Highlands. The genus *Sarcohyla*, which is considered distinct from *Plectrohyla* by some authors to reflect those species west of the Isthmus of Tehuantepec (Duellman et al. 2016), contains 24 species of stream-dwelling frogs, many of them critically endangered and many that have never been seen after their original discovery (references compiled in Stuart et al. 2008). Some species are thought to be in serious decline or extinct (Lips et al. 2004). The actual number of species, their relationships, and geographic ranges are not well known (Duellman 2001; Duellman et al. 2016; Faivovich et al. 2009). *S. bistincta* is the most broadly distributed member of the genus. It occurs in several mountain ranges separated by lowland barriers and might therefore be comprised of multiple species.

We sought to assess potential cryptic diversity in *S. bistincta* using multilocus genomic markers collected via next-generation sequencing (NGS) as well as mitochondrial DNA (mtDNA) data. Often, mtDNA and multilocus nuclear markers have been pitted against one another in biodiversity discovery (Edwards & Bensch 2009; Moritz & Cicero 2004; Zink & Barrowclough 2008). Mitochondrial DNA offers many benefits, including economy, efficiency, and wide comparative potential across species and studies. Consequently, there is now a vast

trove of mtDNA sequences on GenBank. A drawback of using only mtDNA is that a single marker will often fail to accurately depict the speciation process (Edwards & Bensch 2009). In response, multilocus methods have multiplied (Edwards 2009; Fujita et al. 2012). While multiple loci help model a more realistic speciation process, multilocus studies suffer from the lack of a standardized marker set, which limits the ability to link uncovered lineages with species already identified and described in prior studies and in public databases. This is especially important in groups with multiple subadult phenotypes (e.g., insects and frogs) and where adult phenotypes are conserved across species. While many studies have used both types of markers, few have explicitly explored the benefits of linked mtDNA and NGS data at the level of the individual for lineage discovery and identification.

For our multilocus genomic markers, we use ultraconserved elements (UCEs). UCEs are an appealing multilocus marker set because the same loci are found across major branches of the tree of life, where they act as anchors for variable DNA in flanking regions (Faircloth et al. 2012). For instance, it is possible to capture the same set of 1000 or more UCEs across all mammals (McCormack et al. 2012), all reptiles (Crawford et al. 2012), or hundreds of UCEs across arachnid lineages separated by hundreds of millions of years (Starrett et al. 2016). While the power of UCEs for deep-level systematics is clear, their utility at shallower scales around the species level is still coming into focus (McCormack et al. 2016; Smith et al. 2014; Zarza et al. 2016). An added benefit of the UCE enrichment process (and all so-called "sequence capture" methods) is that whole mtDNA genomes are often captured as off-target "bycatch" (do Amaral et al. 2015), meaning each individual has associated nuclear and mtDNA data (e.g., Zarza et al. 2016). Our specific goals were to determine if UCEs are useful for addressing potential cryptic diversity within a species complex of frogs. Then, we assess whether linked mtDNA data, by providing a bridge to Genbank data, allows for more refined conclusions about the identification of any discovered lineages.

Methods

Sampling and Ingroup Determination

MK and PH collected tadpoles from January to June 2004 across most of the range where *Sarcohyla bistincta* are known to exist (Duellman 2001) in the Transvolcanic Belt of Michoacán, Morelos, and the state of México, the Sierra Madre del Sur of Guerrero, and the highlands of Oaxaca stretching into Puebla and Veracruz (Fig. 1; Table 1). Unsampled parts of the *S. bistincta* range include the far west Transvolcanic Belt in Michoacán and Jalisco, the far northwest in the Sierra Madre Occidental (Nayarit, Durango, and Sinaloa), and the far northeast in Hidalgo (see Fig. S1 for sampled and unsampled locations and known ranges of all *Sarcohyla* species).

Tadpoles were targeted to improve sampling efficiency, which allowed for a larger sampling range and sample density. After collection from a sampling location with a dip net, tadpoles were, to the extent possible, separated by species based on morphology and reared to subadults in the laboratory prior to vouchering. Species identification was based on the most recent diagnosis of *S. bistincta* and other closely related species (Duellman 2001). One tadpole was chosen for the tissue voucher, while the other individuals became physical vouchers with museum catalog numbers. Thus, we provide both field numbers and catalog numbers in Table 1 to provide a link to both the exact genetic material (field number) and the associated phenotype voucher representing that genotype (museum catalog number). Before limiting our taxonomic sampling to 38 individuals in a clade thought to represent *S. bistincta* (as we discuss later, some turned out to be already-described species nested within *S. bistincta*), we ran preliminary phylogenetic analyses including broader sampling of 45 *Sarcohyla* individuals and an outgroup genus *Exerodonta* to ensure we had correctly identified the ingroup (Table S1; Fig. S2).

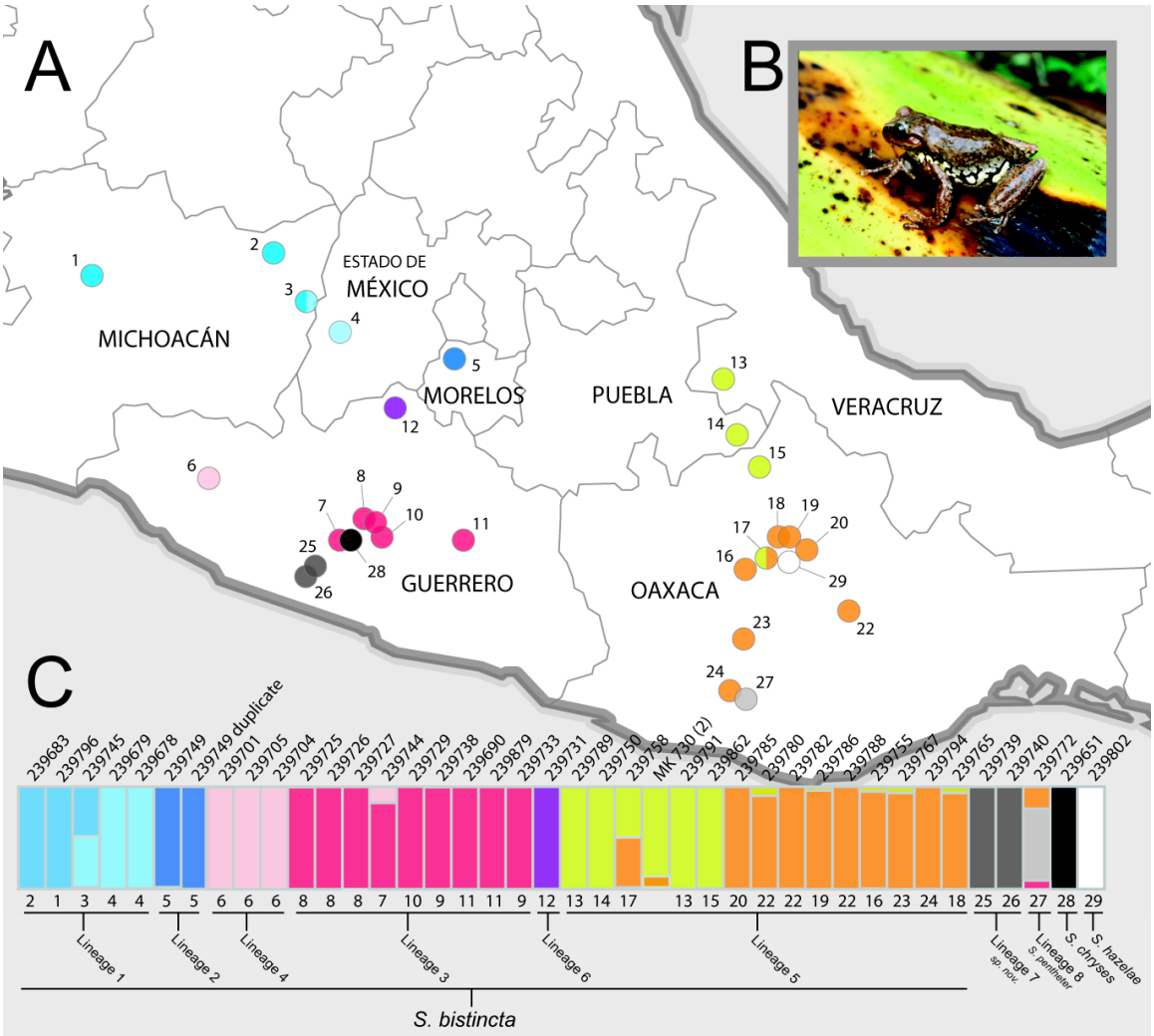


Figure 1. (A) Map of central Mexico showing sampling sites for *S. bistincta* and close outgroups, with numbers matching localities listed in Table 1 and colors matching Structure results below. Unsampled parts of the distribution of *S. bistincta* are shown in Fig. S1; (B) *S. bistincta* individual from near site 1; (C) Composite results of repeated Structure runs at *K* = 2 showing the finest detectable structure in the genetic data. Each vertical line represents an individual labeled with its UMMZ catalog number above and, in descending order below, the site number and the lineage number based on the SNAPP analysis. One tissue voucher MK 730 (2) could not be linked to a specimen voucher and thus its geographic locality is unknown.

136

137 *Sequence capture and next-generation sequencing*

138 We extracted genomic DNA from tissue using a Qiagen (Valencia, CA) DNAeasy Blood and
139 Tissue extraction kit. We visualized extractions on an agarose gel to ensure fragments were
140 larger than 200 base pairs (bp) and quantified the resulting double-stranded DNA using a Qubit
141 2.0 Fluorometer (Carlsbad, CA). For each sample, we sheared 100 μl of 20 ng/μl concentration
142 DNA to a size distribution with its peak between 400 and 600 bp using a Bioruptor
143 ultrasonicator (Diagenode). We prepared libraries for each sheared sample with a KAPA
144 (Boston, MA) LTP library preparation kit for the Illumina platform, attaching custom indexing
145 tags (Faircloth & Glenn 2012) to each sample to allow sample pooling.

146 We enriched pools of eight samples using a set of synthetic RNA probes that target 5,060
147 tetrapod UCEs (MYbaits_Tetrapods-UCE-5K kit, Mycroarray) following the standard UCE
148 enrichment protocol (Faircloth et al. 2012) with one modification. Amphibians have large and
149 variable genome sizes with a high percentage of repetitive DNA (Olmo 1991). While we do not
150 have information about the genome size and composition of *Sarcohyla* specifically, we wanted to
151 decrease the potential risk of the probes hybridizing to repetitive elements (McCartney-Melstad
152 et al. 2016). We thus increased by 6X the amount of the Cot-1 blocker, a synthetic DNA derived
153 from chicken that binds to repetitive regions. After enrichment and recovery PCR, we verified
154 the library size range with an Agilent 2100 Bioanalyzer (Palo Alto, CA). We quantified the
155 enriched pools using qPCR and combined them in equimolar ratios before sequencing on an
156 Illumina HiSeq 2000 lane (100-bp paired-end cycle) at the University of California Santa Cruz
157 Genome Technology Center.

158

159 *Bioinformatics of next-generation sequencing data*

160 We demultiplexed the Illumina raw reads and converted them to FASTQ format with the
161 program bcl2fastq v.1.8.4 (Illumina, Inc.). To eliminate adapter contamination and low quality
162 bases, we trimmed the FASTQC output with illumiprocessor (Faircloth 2013). We trimmed and

163 assembled these reads into contigs with Trinity (Haas et al. 2013) and ABySS (Simpson et al.
164 2009), both of which are built into the PHYLUCE pipeline (Faircloth 2015). PHYLUCE uses
165 LASTZ (Harris 2007) to align all assembled contigs to UCE probes in order to isolate only UCE
166 contigs and to identify and eliminate paralogs.
167
168 *Phylogenetic trees from concatenated UCE data*
169 We extracted UCE contigs into a single FASTA file and aligned the output for each locus using
170 MAFFT (Katoh et al. 2005). We required that 75% of the taxa needed to have data for a given
171 locus to be included in the final concatenated matrix, which led to dropping loci that did not
172 meet this threshold. We then constructed a maximum-likelihood (ML) tree in RAxML v8.0.19
173 (Stamatakis 2014) under the GTRGAMMA model of evolution with 100 bootstrap searches,
174 followed by a search for the tree with the highest likelihood.
175
176 *Mitochondrial DNA assembly and analysis*
177 We identified and assembled mtDNA genomes from off-target, trimmed Illumina reads using
178 the reference genome of a closely related species, *Hyla annectans* (Genebank accession number
179 KM271781; Ye et al. 2016). We used MITObim 1.7 (Hahn et al. 2013), a Perl wrapper for MIRA
180 4.0.2 (Chevreux et al. 1999), that takes an iterative mapping approach for assembly. We
181 conducted *de novo* annotation of the assembled mtDNA regions with the MITOchondrial
182 genome annotation Server, MITOS (Bernt et al. 2013). We selected for analysis only those
183 individual genomes with MIRA quality score greater than 30. We aligned each protein-coding
184 region separately in Geneious vR8 (Kearse et al. 2012) using MUSCLE (Edgar 2004) and
185 corrected the alignments manually when necessary and constructed a concatenated mtDNA
186 matrix, which we also ran in RAxML v8.0.19.

187 We melded this mtDNA data with existing *Sarcohyla* and *Plectrohyla* mtDNA data on
188 Genbank to determine whether any of the lineages we uncovered in *S. bistincta* relate to already-

described species. We determined that *cytochrome b* is the best-represented mtDNA gene on Genbank for this group. We downloaded all existing *cytochrome b* sequences from *Sarcohyla* and *Plectrohyla* taxa. We combined these sequences with those from a subset of our *S. bistincta* samples, choosing the individual with the most raw reads from each major genetic lineage in the UCE tree. We aligned the trimmed, filtered reads for each individual to a *Sarcohyla cytochrome b* reference sequence and formed a consensus sequence for each individual from the mapped reads. We then created an alignment and generated a phylogeny using BEAST v2.4.2 (Bouckaert et al. 2014). We repeated this process with 16S data because we suspected *S. pentheter* was closely related to our ingroup, but *S. pentheter* had no *cyt b* sequence on Genbank.

Calling SNPs from UCE loci

We called SNPs from UCE loci so that we could run genetic clustering tests and infer a species tree using SNAPP (Bryant et al. 2012), which uses SNPs as input data. Calling SNPs requires a reference sequence, and we chose the sample with the most UCE contigs recovered within the ingroup (UMMZ 239727). We then used BWA (Li & Durbin 2009) to map the reads of each sample to this reference. We used SAMtools (Li et al. 2009) to sort the reads, and Picard (http://broadinstitute.github.io/picard) to identify and remove PCR duplicates. We realigned the mapped reads to minimize mismatched bases due to indels, and we removed indels using the Genome Analysis Toolkit 3.2 (GATK; McKenna et al. 2010) and a custom script (indelrealigner.sh), as suggested by the Best Practices workflow (DePristo et al. 2011; van der Auwera et al. 2013).

There is no SNP database available for treefrogs, so we followed best practices for base recalibration for non-model organisms suggested by GATK (McKenna et al. 2010). This consists of (1) doing an initial round of calling SNPs on the original, uncalibrated data, (2) selecting the SNPs with the highest confidence (a minimum emission and call quality of 40 or more), and (3) using these SNPs as the reference database of known SNPs. We executed four rounds of base recalibration on the original data to filter out systematic error using a custom script. We called

genotypes on the last recalibrated BAM file. We used vcf-tools (Danecek et al. 2011) to select one SNP per UCE and produce two data sets, one allowing 25% missing data for STRUCTURE v 2.3.4 (Pritchard et al. 2000), and one with no missing data – a requirement for SNAPP as implemented in BEAST v2.2.1 (Bouckaert et al. 2014).

STRUCTURE analyses

We used STRUCTURE v2.3.4 (Pritchard et al. 2000) as an unbiased way to assess the limits of fine-scale genetic structure in our data. Our intent with using STRUCTURE was not to determine the single most likely number of genetic clusters; thus, we did not use a method for identifying the "true K" (Evanno et al. 2005), which can underestimate fine-scale population structure (Janes et al. 2017). Rather, our goal was to determine the maximum possible number of genetic clusters in our data (e.g., Brown et al. 2007; Gowen et al. 2014). We began by analyzing all individuals of *S. bistincta* plus two outgroup species *S. chryses* and *S. hazelae* under $K = 4$, reasoning that this would likely split out the two outgroups as well as reveal at most one division within *S. bistincta*. After this, each identified genetic cluster within *S. bistincta* was further analyzed at $K = 2$ until no coherent geographically-based structure was evident in the plots. All runs were completed twice and each used an admixture model and 10M generations with 1M generations as burn-in, which led to convergence for all analyses.

SNAPP tree and species delimitation

We generated a species tree from the SNP matrix using SNAPP 1.1.10 (Bryant et al. 2012). This analysis included putative *S. bistincta* samples (again, later shown to include other species nested within) and one outgroup, *S. chryses.* For this run, we made no *a priori* assumptions about how individuals grouped into species and allowed each individual to be considered its own "species" (i.e., terminal tip). We ran two instances of SNAPP for seven million generations using default priors. We combined tree and parameter files from both runs with LogCombiner

242 2.1.3 and displayed the full set of likely species trees with Densitree v2.2.1 (Bouckaert et al.
243 2014).
244
245 **Results**
246 *NGS summary statistics*
247 Detailed summary statistics for each of the 38 ingroup samples and two outgroups are
248 described in Table 1. ABySS produced longer contigs than Trinity, and a higher number of UCE
249 loci, so we used ABYSS contigs in all downstream analyses. Reads per sample ranged from
250 17,052 to 3,423,330 with an average of 1,185,165 reads. The number of identified UCEs ranged
251 from 381 to 2,444 with an average of 1,976 UCEs. The mean length of individual UCE loci per
252 individual ranged from 222 to 717 bp with an average of 522 bp. On average, 18% of the
253 assembled contigs corresponded to unique UCE loci.
254 For SNP calling, across 40 samples of *S. bistincta* and outgroups, 9% of the trimmed
255 reads mapped to our designated reference individual. SNP read depth ranged from 2.4 to 35.0
256 with an average depth of 21.2. The recalibration and quality control steps resulted in an initial
257 matrix of 16,578 SNPs. After removing non-biallelic loci, selecting one SNP for every UCE, and
258 allowing 25% missing data, there were 1,742 SNPs left in the STRUCTURE data set, while the
259 100% complete data matrix for SNAPP contained 399 SNPs.
260
261 *UCE phylogeny from concatenated data*
262 Our more taxonomically inclusive data set with all available *Sarcohyla* and outgroup *Exerodonta*
263 *xera* (Table S1) contained 1,866 UCE loci and 1,030,450 bp for a concatenated analysis. The
264 resulting ML tree (Fig. S2) showed strong support for monophyly of *Sarcohyla*, and identified *S.*
265 *arborescandens* and *S. cyclada* as sister species that together form a clade sister to the rest of the
266 *Sarcohyla* included in the study. We limited further analyses to a smaller data set of 40 samples
267 with *S. hazelae* as the outgroup (Table 1). This focal data set contained 1,891 UCE loci and
268 1,038,600 bp.

The ML tree of these 40 samples found strong support for many clades within the species currently described as *S. bistincta*, conforming to distinct geographic areas (Fig. 2a).

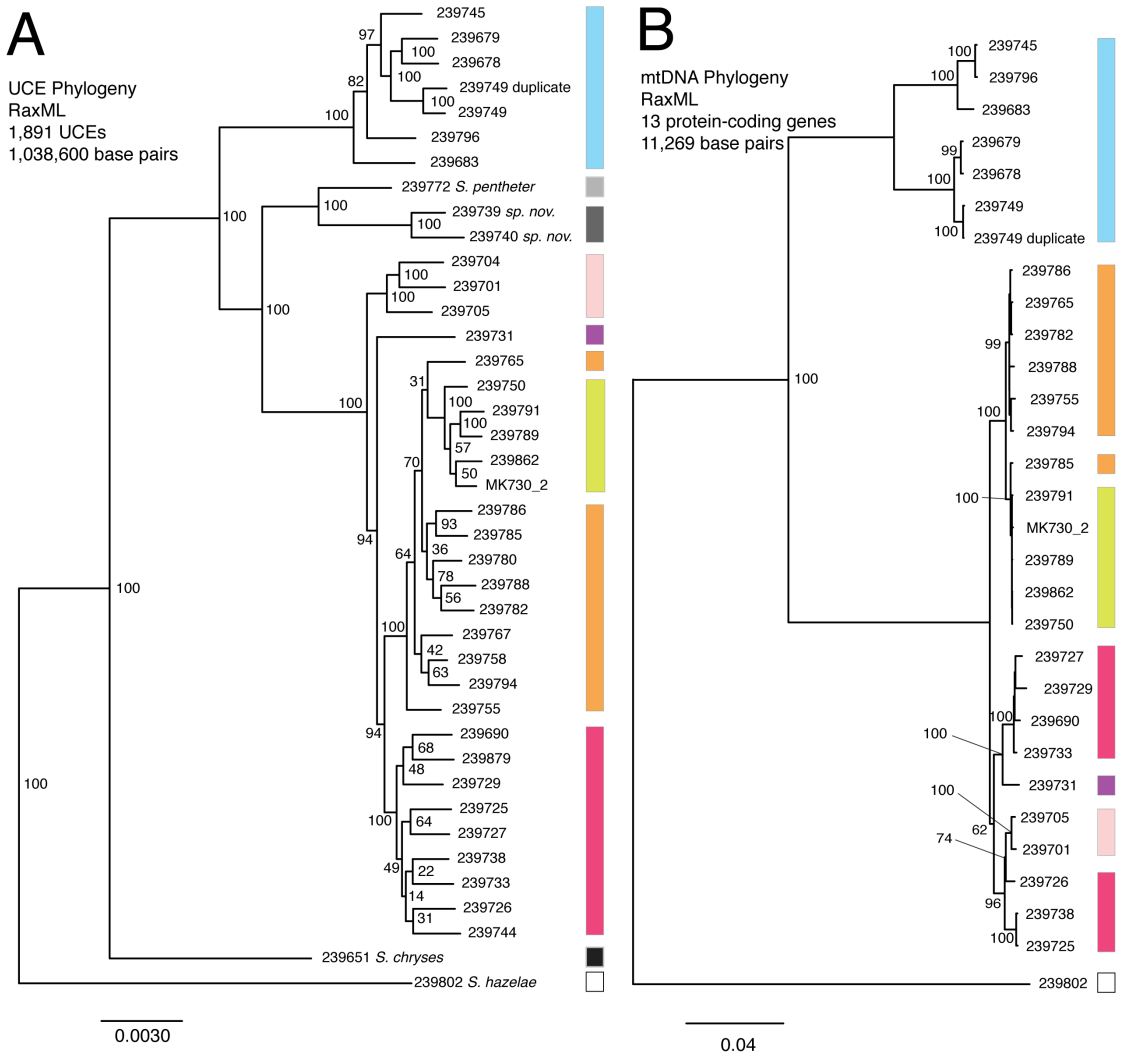


Figure 2. (A) UCE tree; (B) mtDNA tree. Colors match Structure groups identified in Fig. 1. Tips are labeled with their UMMZ catalog number.

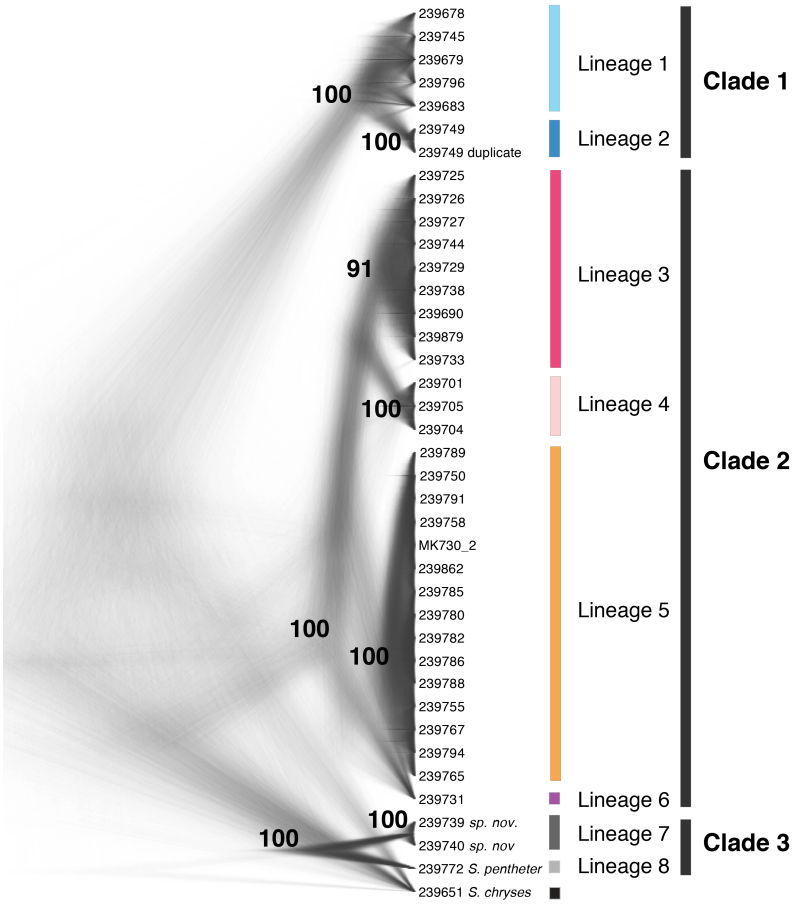
In brief, there were three clades on relatively long branches: (1) a clade distributed across the Transvolcanic Belt (shaded blue in figures); (2) a clade inhabiting two disjunct areas along the coastal slopes of the Sierra Madre del Sur in Guerrero and Oaxaca (shaded gray); and (3) a clade broadly distributed in the Sierra Madre del Sur (shaded red and pink), the Oaxaca Highlands

280 (shaded yellow and orange), and one individual in the southern portion of the Transvolcanic
281 Belt (shaded purple). One individual that nested within *S. bistincta* was labeled as a different
282 species, *S. mykter*, from Guerrero. We suspect that this sample was mislabeled and is actually a
283 duplicate of an *S. bistincta* sample already included in the study because their field numbers are
284 similar (last two digits transposed) and the two samples grouped together in all analyses. We
285 have left this sample in all analyses, but have labeled it as a duplicate of *S. bistincta* UMMZ
286 239749. Another tissue voucher, MK 730 (2), could not be linked definitively to a physical
287 voucher, and thus its geographic location is unknown. Its tip label has been left as the field
288 number. Connecting UCE with mtDNA data, discussed later, revealed that one specimen in the
289 gray clade (UMMZ 239772) is an already-described species, *S. pentheter*.
290
291 *mtDNA tree*
292 Our final concatenated mtDNA matrix (individual *cyt b* and 16S tree are discussed later) was
293 11,269 base pairs including gaps, as coverage of the mtDNA genome varied from sample to
294 sample in accordance with the non-targeted nature of the DNA collection (Table 1).
295 Relationships in the ML tree (Fig. 2b) among the 29 individuals with high quality scores were
296 similar to the concatenated UCE tree with two key differences, both within the broadly
297 distributed clade in Guerrero and Oaxaca: (1) in the mtDNA tree, individuals from eastern and
298 western Guerrero (shaded pink and red in figures) formed a clade, whereas they were more
299 distantly related in the UCE tree; (2) in the mtDNA tree, individual UMMZ 239731 (shaded
300 purple) was nested within the Guerrero clade instead of being sister to a much more inclusive
301 clade, as in the UCE tree.
302
303 *Structure analysis*
304 As expected, the first run of STRUCTURE at K = 4 split the two outgroup species into distinct
305 clusters and split the remaining individuals into two clusters. Further analysis of each cluster at
306 K = 2 revealed ten genetic clusters in all (Fig. 1c), which are largely concordant with clades

observed in the UCE and mtDNA phylogeny. Most individuals did not share assignment among clusters, although a few individuals showed mixed assignment among clusters.

SNAPP species tree

The SNAPP tree and its cloudogram of posterior species trees (Fig. 3) revealed eight well-supported lineages consistent with many of the genetic clusters in the Structure analysis and with relationships in the UCE and mtDNA whole-genome phylogenies. With respect to the discrepancies between UCE and mtDNA trees detailed above, the SNAPP tree agreed with the mtDNA tree that eastern and western Guerrero individuals form a clade, but agreed with the UCE tree in the placement of individual UMMZ 239731.



319 **Figure 3. Cloudogram of the posterior distribution of SNAPP trees from 399 high-quality**
320 **SNPs mined from UCE loci. Tip labels are UMMZ catalog numbers. Lineages and clades are**
321 **discussed in text. Colors match genetic clusters from Figure 1.**
322
323 *mtDNA phylogeny combining new data with Genbank sequences*
324 Using 16S sequences, we determined that one lineage from Fig. 3 matched an *S. pentheter*
325 sequence on Genbank. This individual (UMMZ 239772) had one of the lowest read counts of
326 any samples and very few mtDNA reads. However, five Illumina reads mapped to 16S covering
327 421 bp of the 681 bp reference sequence (Genbank *S. pentheter* accession number DQ055825).
328 Over this stretch, UMMZ 239772 was identical to the *S. pentheter* reference. As a point of
329 comparison, UMMZ 239679 (a member of the blue *S. bistincta* Lineage 1 in the Transvolcanic
330 Belt) had 70 differences across the 681 bp (10.3% divergence). This DNA identification of
331 UMMZ 239772 as *S. pentheter* was later confirmed by re-examining the subadult specimen.
332 After confirming UMMZ 239679 as *S. pentheter*, we generated a Bayesian tree of
333 *cytochrome b* combining the samples from this study with Genbank sequences (Fig. 4). This tree
334 revealed not only that *S. pentheter* was nested within the current *S. bistincta*, but so was another
335 species not included in our sampling, *S. calthula*. The tree also helped clarify relationships
336 outside of *S. bistincta* by supporting *S. chryses* + *S. mykter* to be sister to the *S. bistincta* + *S.*
337 *pentheter* + *S. calthula* clade.
338

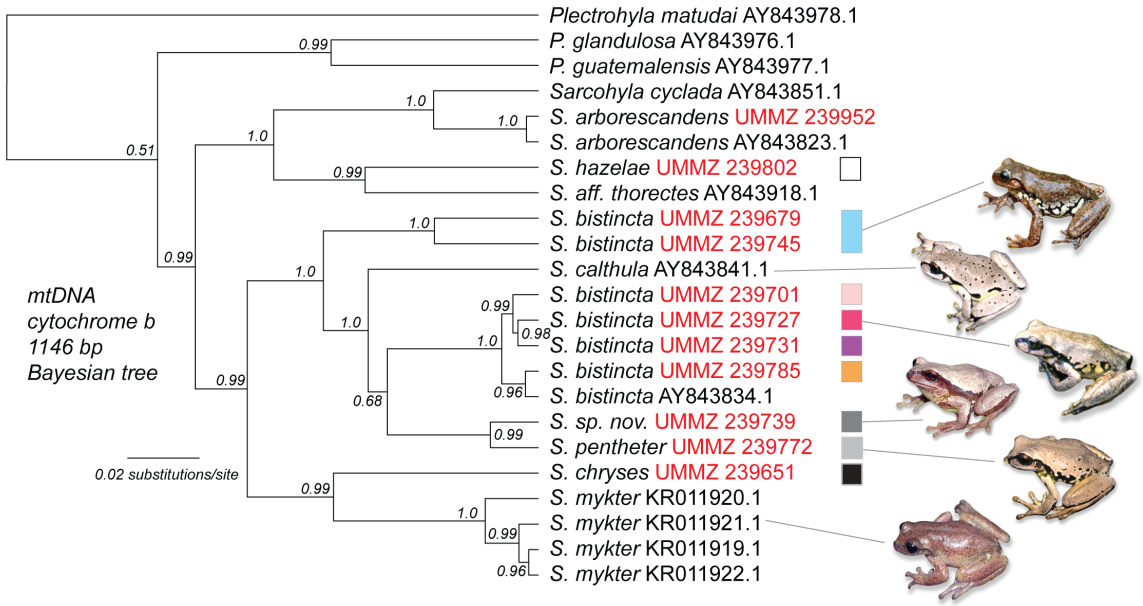


Figure 4. Phylogeny of mtDNA for a subset of the individuals from this study (labeled with red UMMZ catalog numbers) combined with existing sequences from Genbank (labeled with accession numbers). Colored boxes relate to genetic lineages in prior figures. The tree was rooted with *Exerodonta xera*. Photo credit: Peter Heimes.

Discussion

Bridging multilocus and mtDNA data for cryptic lineage discovery

Multilocus NGS data identified numerous divergent lineages within the *Sarcohyla bistincta* complex, supporting similar patterns observed in a recent study of this complex based on mtDNA and a few nuclear genes (Caviedes-Solis & Nieto-Montes 2017). We discuss these uncovered lineages in detail below, but note here that linking NGS with mtDNA data allowed us to query our lineages – first uncovered with multilocus genomic data – against GenBank to see if any of them corresponded to already-described species. Doing so revealed that *S. bistincta* is paraphyletic, with two already-described species nested within its current taxonomic limits (Fig. 4). The approach of discovering lineages with multilocus data and identifying them with the help of linked mtDNA data should be especially useful in understudied groups where basic natural history information is lacking.

UCEs as a universal genomic marker set for species discovery?

Although it requires further study, UCEs could answer the call for an multilocus DNA marker set that satisfies criteria for both ease-of-use, universality, and genomic coverage (Coissac et al. 2016). UCE probe sets are now available for many taxonomic groups (Faircloth et al. 2015; Faircloth et al. 2013; Starrett et al. 2016). They capture a discrete and replicable portion of the genome, in this case a set of around 2,000 loci in frogs (from a larger set of ~5,000 vertebrate loci) that query approximately 1,000,000 base pairs, or 0.02% of the frog genome. The replicable nature of UCEs sets them apart from other types of genomic markers, like RAD loci, which can vary from experiment to experiment (DaCosta & Sorenson 2014) and find fewer orthologs with increasing phylogenetic distance (Cruaud et al. 2014). Other alternatives to UCEs exist, like exons, but in mammals exons had fewer loci conserved over broad taxonomic scales than UCEs, making them less able to be universally applied (McCormack et al. 2012). A recent study successfully applied UCEs in conjunction with species delimitation methods to two frog genera, *Melanophryniscus* and *Brachycephalus* (Pie et al. 2017). As more studies are conducted, one future research avenue would be determine how much locus-sharing occurs among studies, and whether objective benchmarks for species-level divergence can be identified.

Implications for species limits within Sarcohyla bistincta

The monophyletic group containing *Sarcohyla bistincta* is comprised of many distinct lineages as well as two already-described species (*S. pentheter* and *S. calthula*). These lineages are grouped into three fairly divergent clades, which agree with clades uncovered in a recent study based on fewer markers (Caviedes-Solis & Nieto-Montes 2017). In discussing these clades and lineages below, we use the most conservative estimate of 8 lineages, supported by the species tree, as a framework. Within that framework, we discuss the potential for further genetic structuring, as suggested by genetic clustering results and individual UCE and mtDNA phylogenies. We do not attempt to estimate the "true" number of species through species delimitation or other

methods, because we feel species delimitation is best carried out through integrative taxonomy (Will et al. 2005), including at minimum analysis of the phenotype.

Clade 1 – Transvolcanic Belt of central Mexico. This clade is sister to the rest of *S. bistincta* plus two other species, *S. pentheter* and *S. calthula*, and is 10% divergent in mtDNA from other *S. bistincta*, making it an obvious candidate for species status. This clade might itself contain multiple species in the form of the lineages below. Unsampled populations of *S. bistincta* in the Sierra Madre Occidental (Fig. S1) are very likely related to Lineage 1 of this clade, but should be included in future studies, as they might be distinctive.

Lineage 1 (light blue in Fig. 1) – Michoacán to western Mexico state. The Structure results show fine-scale genetic structure across this range, and the presence of a geographic and genetic intermediate hints at continuity of gene flow along the distribution from sites 1 to 4 in Figure 1. In addition to unsampled populations mentioned above, some populations in far western Michoacán (Fig. S1) are also unsampled and could reveal further genetic structure.

Lineage 2 (dark blue in Fig. 1) – Morelos. Denser sampling between sites 4 and 5 could help determine whether the genetic distinctness of this individual in the mtDNA and UCE trees is a true discontinuity or the result of a sampling gap.

Clade 2 – Guerrero to Puebla and Veracruz and south through Oaxaca. This clade was also found to be distinct in a recent study (Caviedes-Solis & Nieto-Montes 2017) and forms the core *S. bistincta* (*sensu stricto*).

Lineage 3 (red in Fig. 1) – Central and eastern Guerrero. Members of this lineage are distinct from Lineage 4 and are monophyletic in the UCE-based trees (though not in the mtDNA tree). Further sampling in between site 6 and site 7 would clarify whether the genetic discontinuity between Lineages 3 and 4 results from a sampling gap.

411 Lineage 4 (pink in Fig. 1) – Western Guerrero. This lineage is monophyletic in all trees,
412 although only a few individuals were sampled from a single locality.
413 Lineage 5 (orange and yellow in Fig. 1) – Puebla, Veracruz, and Oaxaca. This lineage
414 contains the type locality for *S. bistincta*. Central and southern Oaxaca individuals
415 (orange) are genetically distinct from individuals to the north (yellow). One genetic
416 intermediate in central Oaxaca suggests genetic continuity across this range. An
417 unsampled northern population in Hidalgo is most likely related to this lineage,
418 and should be included in future studies.
419 Lineage 6 (purple in Fig. 1) – far northern Guerrero. This lineage is distinct but
420 represented by only a single individual. However, in the mtDNA tree, this
421 individual is nested within Lineages 4 and 5 above. Sampling more individuals is
422 needed to determine how distinct this lineage might be.
423 **Clade 3** – Pacific slope of Guerrero and Oaxaca. This clade contains two species: one already
424 described (*S. pentheter*) and one currently being described (*sp. nov.*; Kaplan et al. in
425 prep). The relationship of *S. calthula* to this clade is unclear in our results, as
426 bootstrap support was low in the combined mtDNA tree (Fig. 4). Caviedes-Solis et
427 al. (2017) placed *S. calthula* sister to *S. pentheter* based on five genes. Thus *S. calthula*
428 is likely also a member of Clade 3.
429 Lineage 7 (dark gray in Fig. 1) – Pacific slope of Guerrero. This lineage was being
430 described as a new species (we call it *sp. nov.*) on the basis of phenotypic differences
431 before this genetic study was begun (Kaplan et al. in prep). Thus, our study lends
432 support to species status for this lineage.
433 Lineage 8 (light gray in Fig. 1) – *S. pentheter*. Pacific slope of Oaxaca.
434
435 The three clades and nearly all of the lineages were distinct in the mtDNA tree as well as the
436 UCE tree (Fig. 2). The mtDNA tree, however, supports a more complicated history for Lineages
437 3, 4, and 6 in Guerrero. It is unclear why UCE and mtDNA results differed in this regard, but

438 some reticulate processes might have influenced the mtDNA genomes of these lineages,
439 perhaps, given their close geographic proximity, mtDNA capture of one lineage by another
440 through gene flow (e.g., Bryson Jr et al. 2010).

441 Additional insights into broader *Sarcohyla* relationships offered by the mtDNA tree
442 include support for a previously hypothesized close relationship between *S. hazelae* and *S.*
443 *thorectes* (Caviedes-Solis & Nieto-Montes 2017; Faivovich et al. 2009), and a sister relationship
444 between *S. mykter* and *S. chryses* (also found by Caviedes-Solis et al. (2017)). As *Sarcohyla* is very
445 poorly represented by voucher specimens and DNA sequences (Fig. S1), a complete
446 understanding of the history of this genus must await more complete taxonomic and genomic
447 sampling. Unfortunately, there appear to be some microendemic *Sarcohyla* that might have
448 already gone extinct (Lips et al. 2004), especially in the Oaxacan highlands, although recent
449 resurveys provide hope for rediscoveries (Delia et al. 2013).

450 This and other recent studies show there is still substantial diversity remaining to be
451 described in the Mexican Highlands. Genetic studies to uncover this diversity might use
452 different approaches and marker types, but these efforts need not be in opposition. As our
453 study shows, NGS and mtDNA data work well together, when lineages uncovered via
454 multilocus methods (but with linked mtDNA data) can be checked against and combined with
455 existing mtDNA databases like Genbank. Identification of vouchered material might not be a
456 problem with some well-studied groups, like birds, but it is invaluable for groups with
457 undescribed larval stages, unclear range boundaries, and difficult or highly conserved
458 phenotypes. As the destruction of native habitats continues apace, it is important that we
459 identify distinctive lineages and geographic centers of diversity before they are lost.

460

461 **List of abbreviations**

462 **UCEs:** ultraconserved elements **SNPs:** single nucleotide polymorphisms **mtDNA:**
463 mitochondrial DNA **bp:** base pairs **ML:** maximum-likelihood **UMMZ:** University of Michigan
464 Museum of Zoology **RAD loci:** restriction digest-associated loci

Declarations

Ethics approval and consent to participate

This study complied with standard ethical guidelines for the rearing and collection of tadpoles.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed in the current study will be made available on Genbank (BioProject ID PRJNA393258) and Dryad upon publication.

Competing Interests

The authors declare that they have no competing interests.

Funding

The project was partially funded by National Science Foundation grant DEB-1258205 to JM.

Authors' contributions

MK and PH collected samples in the field; MK and JEM designed the study; EZ and WLET produced the genomic data; EZ, EMC, and JMM analyzed the data; JEM and EZ wrote the manuscript; JEM created the figures; All authors read and approved the final manuscript.

Acknowledgments

We thank Greg Schneider and Ron Nussbaum at UMMZ for providing access to tissue samples and vouchers, and Oscar Flores for help with permits.

492

493 **Literature Cited**

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

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

500 Brown DM, Brenneman RA, Koepfli K-P, Pollinger JP, Milá B, Georgiadis NJ, Louis EE, Grether
501 GF, Jacobs DK, and Wayne RK. 2007. Extensive population genetic structure in the
502 giraffe. *BMC biology* 5:57.

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

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

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

512 Bryson RW, García-Vázquez UO, and Riddle BR. 2012. Diversification in the Mexican horned
513 lizard *Phrynosoma orbiculare* across a dynamic landscape. *Molecular Phylogenetics and*
514 *Evolution* 62:87-96.

515 Bryson RW, Murphy RW, Lathrop A, and Lazcano-Villareal D. 2011. Evolutionary drivers of
516 phylogeographical diversity in the highlands of Mexico: a case study of the *Crotalus*
517 *triseriatus* species group of montane rattlesnakes. *Journal of Biogeography* 38:697-710.

518 Campbell JA, Blancas-Hernández JC, and Smith EN. 2009. A new species of stream-breeding
519 treefrog of the genus Charadrahyla (Hylidae) from the Sierra Madre del Sur of Guerrero,
520 Mexico. *Copeia* 2009:287-295.

521 Caviedes-Solis I, and Nieto-Montes dOA. 2017. A multilocus phylogeny of the genus *Sarcohyla*
522 (Anura: Hylidae), and an investigation of species boundaries using statistical species
523 delimitation. *Molecular Phylogenetics and Evolution* 118:184-193.

524 Chevreux B, Wetter T, and Suhai S. 1999. Genome sequence assembly using trace signals and
525 additional sequence information. *Computer Science and Biology: Proceedings of the German*
526 *Conference on Bioinformatics* 99:45-56.
527 Coissac E, Hollingsworth PM, Lavergne S, and Taberlet P. 2016. From barcodes to genomes:
528 extending the concept of DNA barcoding. *Molecular ecology* 25:1423-1428.
529 Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, and Glenn TC. 2012.
530 More than 1000 ultraconserved elements provide evidence that turtles are the sister
531 group of archosaurs. *Biology Letters* 8:783-786. 10.1098/rsbl.2012.0331
532 Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, Genson G, Dubois E, Nidelet S, Deuve T,
533 and Rasplus J-Y. 2014. Empirical assessment of RAD sequencing for interspecific
534 phylogeny. *Molecular Biology and Evolution* 31:1272-1274.
535 DaCosta JM, and Sorenson MD. 2014. Amplification biases and consistent recovery of loci in a
536 double-digest RAD-seq protocol. *PloS one* 9:e106713.
537 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
538 Marth GT, and Sherry ST. 2011. The variant call format and VCFtools. *Bioinformatics*
539 27:2156-2158.
540 Delia JR, Whitney JL, and Burkhardt T. 2013. Rediscovery of 'lost' treefrogs from the Oaxacan
541 highlands of Mexico. *Biodiversity and conservation* 22:1405-1414.
542 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel
543 G, Rivas MA, and Hanna M. 2011. A framework for variation discovery and genotyping
544 using next-generation DNA sequencing data. *Nature genetics* 43:491-498.
545 do Amaral FR, Neves LG, Resende Jr MF, Mobili F, Miyaki CY, Pellegrino KC, and Biondo C.
546 2015. Ultraconserved elements sequencing as a low-cost source of complete
547 mitochondrial genomes and microsatellite markers in non-model amniotes. *PloS one*
548 10:e0138446.
549 Duellman WE. 2001. *Hylid frogs of Middle America*. Ithaca, New York: Society for the Study of
550 Amphibians and Reptiles.
551 Duellman WE, Marion AB, and Hedges SB. 2016. Phylogenetics, classification, and
552 biogeography of the treefrogs (Amphibia: Anura: Arboranae). *Zootaxa* 4104:1-109.
553 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
554 throughput. *Nucleic acids research* 32:1792-1797.
555 Edwards S, and Bensch S. 2009. Looking forwards or looking backwards in avian
556 phylogeography? A comment on. *Molecular ecology* 18:2930-2933.
557 Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution*
558 63:1-19.

559 Evanno G, Regnaut S, and Goudet J. 2005. Detecting the number of clusters of individuals using
560 the software STRUCTURE: a simulation study. *Molecular ecology* 14:2611-2620.
561 Faircloth BC. 2013. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality
562 trimming.https://github.com/faircloth-lab/illumiprocessor/.
563 Faircloth BC. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci.
564 *Bioinformatics* 32:786-788.
565 Faircloth BC, Branstetter MG, White ND, and Brady SG. 2015. Target enrichment of
566 ultraconserved elements from arthropods provides a genomic perspective on
567 relationships among Hymenoptera. *Molecular Ecology Resources* 15:489-501.
568 Faircloth BC, and Glenn TC. 2012. Not all sequence tags are created equal: designing and
569 validating sequence identification tags robust to indels. *PloS one* 7:e42543.
570 Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, and Glenn TC. 2012.
571 Ultraconserved elements anchor thousands of genetic markers spanning multiple
572 evolutionary timescales. *Systematic Biology* 61:717-726. 10.1093/sysbio/sys004
573 Faircloth BC, Sorenson L, Santini F, and Alfaro ME. 2013. A phylogenomic perspective on the
574 radiation of ray-finned fishes based upon targeted sequencing of ultraconserved
575 elements (UCEs). *PloS one* 8:e65923.
576 Faivovich J, Haddad CF, Garcia PC, Frost DR, Campbell JA, and Wheeler WC. 2009. Systematic
577 review of the frog family Hylidae, with special reference to Hylinae: phylogenetic
578 analysis and taxonomic revision. *Bulletin of the American Museum of Natural History* 294:1-
579 240.
580 Fujita MK, Leaché AD, Burbrink FT, McGuire JA, and Moritz C. 2012. Coalescent-based species
581 delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* 27:480-488.
582 Gowen FC, Maley JM, Cicero C, Peterson AT, Faircloth BC, Warr TC, and McCormack JE. 2014.
583 Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary
584 contact. *BMC evolutionary biology* 14:135.
585 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
586 Li B, and Lieber M. 2013. De novo transcript sequence reconstruction from RNA-seq
587 using the Trinity platform for reference generation and analysis. *Nature protocols* 8:1494-
588 1512.
589 Hahn C, Bachmann L, and Chevreux B. 2013. Reconstructing mitochondrial genomes directly
590 from genomic next-generation sequencing reads—a baiting and iterative mapping
591 approach. *Nucleic acids research* 41:e129.
592 Harris RS. 2007. *Improved pairwise alignment of genomic DNA*. PhD Thesis: Pennsylvania State
593 University.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, and Andrew RL. 2017. The K= 2 conundrum. *Molecular ecology* 14:3594–3602.

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

Li H, and 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, and 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, and Mulcahy DG. 2004. Amphibian population declines in montane southern Mexico: resurveys of historical localities. *Biological Conservation* 119:555-564.

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

McCormack J, Tsai WLE, and 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, and 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. 10.1101/gr.125864.111

McCormack JE, Peterson A, Bonaccorso E, and Smith TB. 2008. Speciation in the highlands of Mexico: genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). *Molecular ecology* 17:2505-2521.

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

Meik JM, Smith EN, Canseco-Márquez L, and Campbell JA. 2006. New species of the Plectrohyla bistincta group (Hylidae: Hylinae: Hylini) from Oaxaca, Mexico. *Journal of Herpetology* 40:304-309.

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

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, and Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853-858.

Olmo E. 1991. Genome variations in the transition from amphibians to reptiles. *J Mol Evol* 33:68-75.

630 Parra-Olea G, Flores-Villela O, and Mendoza-Almeralla C. 2014. Biodiversidad de anfibios en
631 México. *Revista mexicana de biodiversidad* 85:460-466.
632 Pie MR, Bornschein MR, Ribiero LF, Faircloth BC, and McCormack JE. 2017. Phylogenomic
633 species delimitation in microendemic frogs Of the Brazilian Atlantic forest. *bioRxiv*:doi:
634 https://doi.org/10.1101/143735.
635 Ponce-Reyes R, Reynoso-Rosales V-H, Watson JE, VanDerWal J, Fuller RA, Pressey RL, and
636 Possingham HP. 2012. Vulnerability of cloud forest reserves in Mexico to climate
637 change. *Nature Climate Change* 2:448.
638 Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using
639 multilocus genotype data. *Genetics* 155:945-959.
640 Rovito SM, Parra-Olea G, Recuero E, and Wake DB. 2015. Diversification and biogeographical
641 history of Neotropical plethodontid salamanders. *Zoological Journal of the Linnean Society*
642 175:167-188.
643 Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, and Birol I. 2009. ABySS: a parallel
644 assembler for short read sequence data. *Genome Research* 19:1117-1123.
645 Smith BT, Harvey MG, Faircloth BC, Glenn TC, and Brumfield RT. 2014. Target capture and
646 massively parallel sequencing of ultraconserved elements for comparative studies at
647 shallow evolutionary time scales. *Systematic Biology* 63:83-95.
648 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
649 phylogenies. *Bioinformatics* 30:1312-1313.
650 Starrett J, Derkarabetian S, Hedin M, Bryson RW, McCormack JE, and Faircloth BC. 2016. High
651 phylogenetic utility of an ultraconserved element probe set designed for Arachnida.
652 *Molecular Ecology Resources* 17:812-823.
653 Stuart SN, Hoffmann M, Chanson J, Cox N, Berridge R, Ramani P, and Young B. 2008.
654 *Threatened amphibians of the world*. Barcelona, Spain: Lynx Edicions.
655 van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan
656 T, Shakir K, Roazen D, and Thibault J. 2013. From FastQ data to high-confidence variant
657 calls: the genome analysis toolkit best practices pipeline. *Current protocols in*
658 *bioinformatics* 11: 11.10:11.10. 11-11.10. 33.
659 Will KW, Mishler BD, and Wheeler QD. 2005. The perils of DNA barcoding and the need for
660 integrative taxonomy. *Systematic Biology* 54:844-851.
661 Ye L, Zhu C, Yu D, Zhang Y, and Zhang J. 2016. The complete mitochondrial genome of *Hyla*
662 *annectans* (Anura: Hylidae). *Mitochondrial DNA Part A* 27:1593-1594.

663 Zarza E, Faircloth BC, Tsai WL, Bryson RW, Klicka J, and McCormack JE. 2016. Hidden histories
664 of gene flow in highland birds revealed with genomic markers. *Molecular ecology*
665 25:5144-5157.
666 Zink RM, and Barrowclough GF. 2008. Mitochondrial DNA under siege in avian
667 phylogeography. *Molecular ecology* 17:2107-2121.

668

669

670 **Tables**

671

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

673 species.

674

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
	MK 730 (2)	?	*P. bistincta*	?	?	?	2,967,630	2,404	716.9	31.1	3,581	21.7	84
25	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
26	MK 689 (2)	UMMZ 239740	*S. sp. nov.*	GUERRERO	17.3000	-100.2792	67,313	502	221.6	2.6	32	3.1	6
27	MK 727 (2)	UMMZ 239772	*S. pentheter*	OAXACA	16.1916	-97.0958	17,052	381	247.1	2.4	10	3.0	2
28	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
29	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. One specimen could not be linked to a catalog number.

675

676

677

678

679

680

681

682

Table S1. Information and summary statistics on all 45 samples used to determine the ingroup 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	68	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	1,083,058	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*	?	?			2,967,630	2404	716.9	3581	21.7	84		
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	25	
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	26	
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	27	
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	28	
MK 770	UMMZ 239802	*Plectrohyla hazelae*	OAXACA	EL PUNTO SIERRA JUAREZ	17.22156	-96.58386	538,500	2066	526.2	557	5.9	44	29	
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		

Additional Files

Fig. S1. Sampled and unsampled parts of *S. bistincta* range in relation to known distributions (or localities, where distributional information is lacking) of other *Sarcohyla* species.

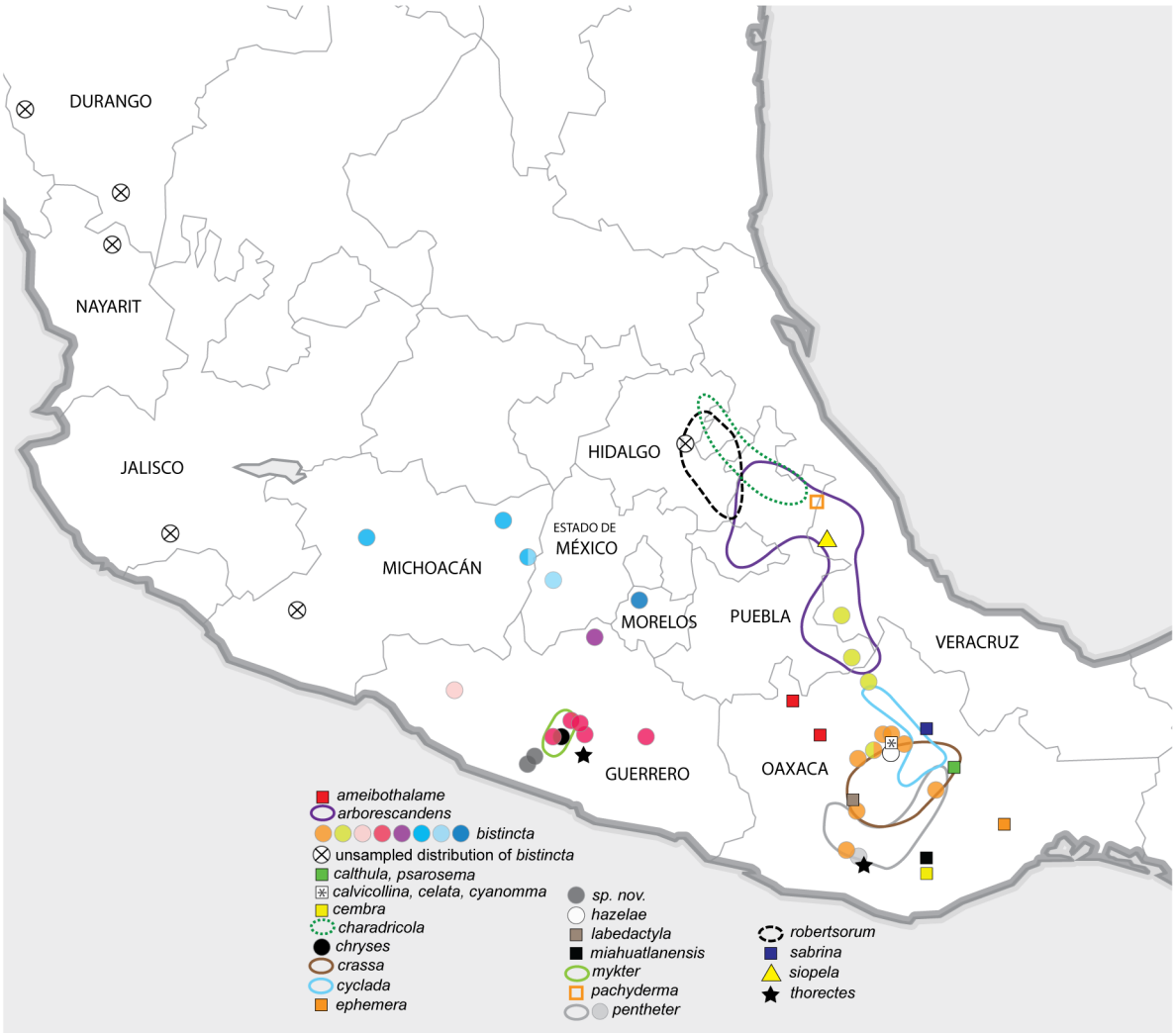
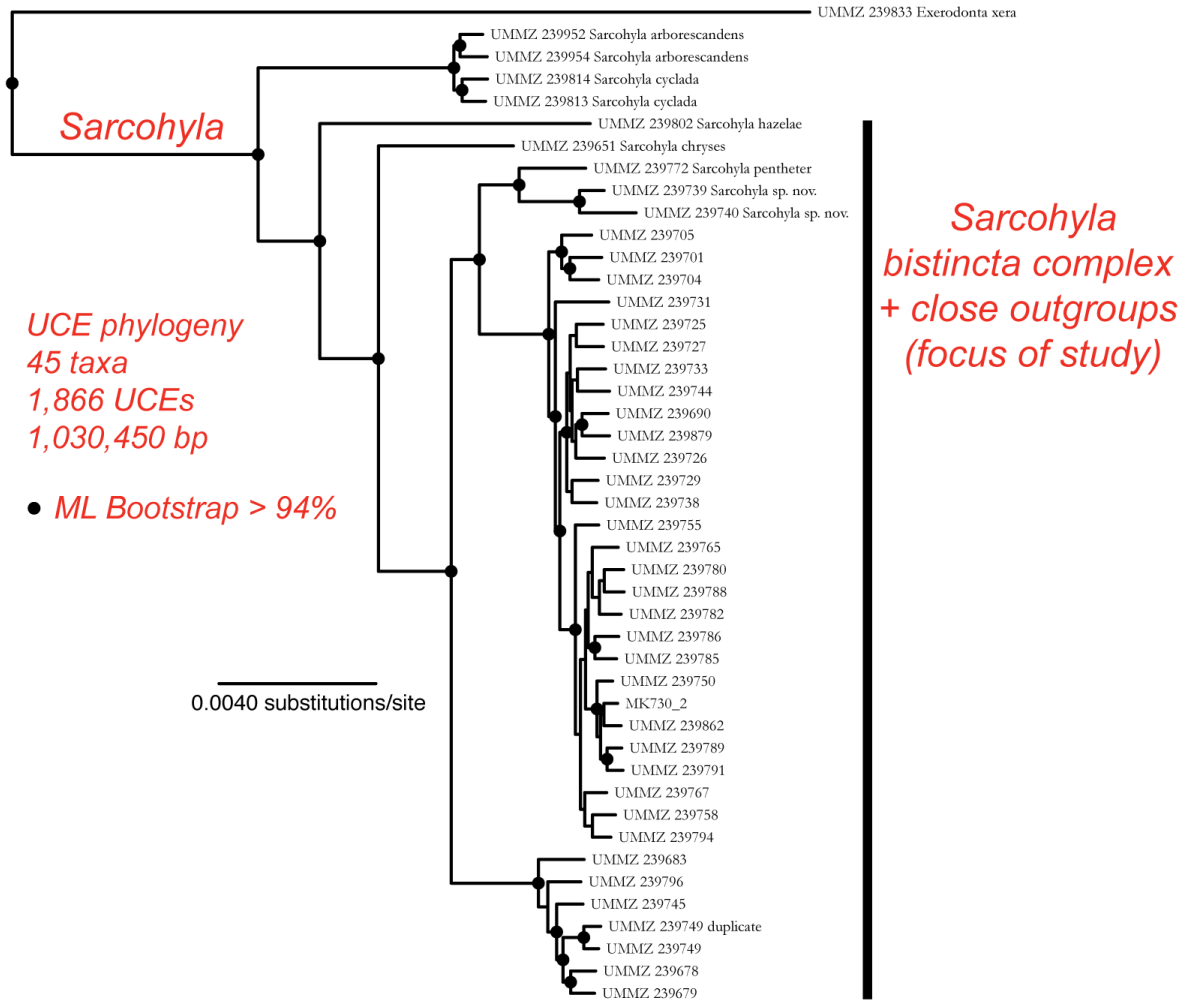



Fig. S2. UCE tree of 45 samples of *Sarcohyla* and outgroup *Exerodonta xera* used to determine the ingroup.