

11-2017

Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: Phrynosoma)

Christopher Blair

Robert W. Bryson Jr.

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/ny_pubs

 Part of the [Computational Biology Commons](#), [Evolution Commons](#), [Terrestrial and Aquatic Ecology Commons](#), and the [Zoology Commons](#)

Recommended Citation

Blair, Christopher, and Robert W. Bryson Jr. "Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: Phrynosoma)." *Molecular ecology resources* 17, no. 6 (2017): 1168-1182.

This Article is brought to you for free and open access by the New York City College of Technology at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

1 Cryptic diversity and discordance in single-locus species delimitation methods within horned
2 lizards (Phrynosomatidae: *Phrynosoma*)

3
4
5 CHRISTOPHER BLAIR^{1,2}, ROBERT W. BRYSON JR.^{3,4}

6
7
8 ¹*Department of Biological Sciences, New York City College of Technology, The City University*
9 *of New York, 300 Jay Street, Brooklyn, NY 11201, USA*

10
11 ²*Biology PhD Program, CUNY Graduate Center, 365 5th Avenue, New York, NY, 10016*

12
13 ³*Department of Biology and Burke Museum of Natural History and Culture, University of*
14 *Washington, 4331 Memorial Way Northeast, Seattle, WA 98195, USA*

15
16 ⁴*Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041, USA*

17
18
19
20
21 *Correspondence: Dr. Christopher Blair; cblair@citytech.cuny.edu; cblair@gc.cuny.edu; 718-*
22 *260-5342 (ph)*

23 **Abstract**

24 Biodiversity reduction and loss continues to progress at an alarming rate, and thus there is
25 widespread interest in utilizing rapid and efficient methods for quantifying and delimiting
26 taxonomic diversity. Single-locus species-delimitation methods have become popular, in part due
27 to the adoption of the DNA barcoding paradigm. These techniques can be broadly classified into
28 tree-based and distance-based methods depending on whether species are delimited based on a
29 constructed genealogy. Although the relative performance of these methods has been tested
30 repeatedly with simulations, additional studies are needed to assess congruence with empirical
31 data. We compiled a large data set of mitochondrial ND4 sequences from horned lizards
32 (*Phrynosoma*) to elucidate congruence using four tree-based (single-threshold GMYC, multiple-
33 threshold GMYC, bPTP, mPTP) and one distance-based (ABGD) species delimitation models.
34 We were particularly interested in cases with highly uneven sampling and/or large differences in
35 intraspecific diversity. Results showed a high degree of discordance among methods, with
36 multiple-threshold GMYC and bPTP suggesting an unrealistically high number of species (29
37 and 26 species within the *P. douglasii* complex alone). The single-threshold GMYC model was
38 the most conservative, likely a result of difficulty in locating the inflection point in the
39 genealogies. mPTP and ABGD appeared to be the most stable across sampling regimes and
40 suggested the presence of additional cryptic species that warrant further investigation. These
41 results suggest that the mPTP model may be preferable in empirical data sets with highly uneven
42 sampling or large differences in effective population sizes of species.

43

44 *Key Words:* ABGD, GMYC, ND4, *Phrynosoma*, PTP, speciation

45 **Introduction**

46

47 Proper identification and delimitation of species is of utmost importance for most fields of
48 biology (de Queiroz 2007). Many research programs address fundamental questions through a
49 comparative framework, necessitating the use of both a robust phylogeny and accurate species
50 assignments for hypothesis testing. Molecular-based species-delimitation methods can generally
51 be classified into single- or multilocus, and discovery and validation-based techniques (Carstens
52 *et al.* 2013). DNA barcoding threshold methods (Hebert *et al.* 2003, 2004; Hebert & Gregory
53 2005; Edgar 2010; Puillandre *et al.* 2012a) comprise one common example of a single-locus
54 technique, where threshold or cut-off values are used to differentiate inter- from intraspecific
55 divergences. The Refined Single Linkage (RESL) method, for example, is a popular clustering
56 algorithm implemented within the Barcode of Life Data Systems (Ratnasingham & Hebert 2013)
57 to delineate operational taxonomic units (OTUs) based on animal COI barcode data. Although
58 these threshold-type methods continue to be a quick and effective way to document and describe
59 diversity, they do not take into account tree structure and often rely on arbitrarily defined
60 thresholds (e.g., 2–3% pairwise sequence divergence) to delimit species (Blaxter 2004; Hebert &
61 Gregory 2005; Hamilton *et al.* 2011).

62 More recently, coalescent-based methods of species delimitation have become common,
63 in part due to the continual ease in which researchers can generate vast quantities of molecular
64 data (Leaché & Fujita 2010; Fujita *et al.* 2012). Bayesian multilocus coalescent methods, for
65 example, can explicitly account for gene tree/species tree incongruence when delimiting species
66 and estimating a species tree (Yang & Rannala 2014; Jones 2014). Although an attractive
67 alternative to threshold-type methods, the utility of many Bayesian multilocus coalescent

68 methods for large data sets remains uncertain due to the relatively large computational demand
69 of the algorithms (Yang & Rannala 2010, 2014; Satler *et al.* 2013; Leaché *et al.* 2014).

70 Single-locus, coalescent-based methods like the General Mixed Yule Coalescent model
71 (GMYC; Pons *et al.* 2006; Fujisawa & Barraclough 2013) have become a popular tree-based
72 species-delimitation technique often applied to barcoding data (e.g., animal mitochondrial DNA).
73 The GMYC model uses maximum likelihood and an ultrametric gene tree to model the transition
74 between inter- and intraspecific branching patterns. Branching patterns older than the inferred
75 threshold represent speciation events (Yule process), whereas younger branching indicates
76 neutral coalescence within species. GMYC has been used in numerous empirical studies (e.g.,
77 Monaghan *et al.* 2009; e.g. Barraclough *et al.* 2009; Hamilton *et al.* 2011; Gebiola *et al.* 2012;
78 Esselstyn *et al.* 2012; Blair *et al.* 2015), and recent simulations and empirical data suggest that
79 the method is fairly robust to different assumptions (Reid & Carstens 2012; Esselstyn *et al.* 2012;
80 Fujisawa & Barraclough 2013; Talavera *et al.* 2013; Tang *et al.* 2014). The Poisson Tree
81 Processes (PTP/bPTP) model is similar in that it seeks to model the transition in branch lengths
82 between versus within species (Zhang *et al.* 2013). However, PTP estimates branching processes
83 using the expected number of substitutions (versus time in GMYC), and thus utilizes a non-
84 ultrametric phylogenetic tree as input. One limitation of the original PTP model is that it assumes
85 only two independent distributions to model branch lengths (one exponential distribution for
86 speciation and one exponential distribution for coalescence). This generally ignores the
87 stochastic variation among species due to different population sizes and demographic histories.
88 Conversely, the recently-developed multi-rate Poisson Tree Processes model (mPTP) fits
89 multiple independent exponential distributions to each delimited species to explicitly account for
90 differences in sampling intensity and/or population history (Kapli *et al.* 2016). Although the

91 mPTP model may potentially lead to more accurate delimitations versus other single-locus
92 methods, testing and comparison using empirical data characterized by highly heterogeneous
93 sampling intensity and/or large differences in genetic diversity among species is lacking.

94 Horned lizards (*Phrynosoma*) are a genus of phrynosomatid lizards consisting of 17–21
95 species distributed from Canada to Guatemala (Leaché & Linkem 2015; Montanucci 2015). The
96 unique morphology and behavior of these lizards, including ocular blood squirting (Sherbrooke
97 2003), has made them the target of numerous systematic investigations. Early studies based on
98 mitochondrial DNA (mtDNA) sequences yielded conflicting phylogenetic relationships with
99 both nuclear and morphological data (e.g. Hodges and Zamudio, 2004; Leaché and McGuire,
100 2006; Reeder and Montanucci, 2001), presumably because of mtDNA introgression (Leaché &
101 McGuire 2006; Mulcahy *et al.* 2006). More recently, next-generation sequencing has been used
102 to estimate a robust phylogeny for the genus, with results suggesting that cladogenesis initiated
103 in the Miocene around 22 million years ago (Leaché & Linkem 2015). Phylogeographic studies
104 have also been conducted on several species including *P. douglasii* (Zamudio *et al.* 1997), *P.*
105 *mcallii* (Mulcahy *et al.* 2006), *P. platyrhinos* (Jezkova *et al.* 2016), *P. coronatum* (Leaché *et al.*
106 2009), and *P. orbiculare* (Bryson *et al.* 2012). Many of these studies have indicated the presence
107 of undocumented cryptic diversity, but whether any of this diversity may warrant species status
108 has not been evaluated in detail for most groups (but see Leaché *et al.* 2009). These studies have
109 also revealed substantially different levels of intraspecific diversity, ranging from deep lineages
110 within *P. orbiculare* to low levels of genetic diversity in *P. mcallii*. More recent taxonomic work
111 on the genus has suggested that levels of diversity may be underestimated at a clade-wide level
112 (Nieto-Montes de Oca *et al.* 2014; Montanucci 2015).

113 Given the need for additional empirical studies to compare and contrast single-locus
114 “discovery-based” species delimitation methods, particularly in cases with large differences in
115 sampling intensity and levels of intraspecific diversity, in this study we assess congruence among
116 four tree-based and one distance-based methods of delimiting horned lizard species. We
117 particularly focus on the utility of one method (mPTP) with the ability to accommodate highly
118 heterogeneous data sets comprised of species with dramatically different levels of molecular
119 diversity. Based on our results, we provide further quantitative evidence for undescribed species
120 in the genus.

121

122 **Materials and Methods**

123

124 *Data collection*

125 We obtained from GenBank a total of 368 orthologs of the mitochondrial ND4 gene from
126 multiple representatives of *Phrynosoma*. Although many molecular phylogenetic studies have
127 been performed on the genus, there have been relatively few phylogeographic investigations of
128 species using the same marker. We extracted from GenBank multiple sequences from *P. mcallii*
129 (Mulcahy *et al.* 2006), *P. platyrhinos* (Jezkova *et al.* 2016), *P. orbiculare* (Bryson *et al.* 2012),
130 and the *P. douglasii* complex (Zamudio *et al.* 1997). We included singletons of several other
131 species, including *P. cornutum*, *P. coronatum*, *P. solare*, *P. asio*, and *P. taurus*. Although the
132 COI gene is the generally accepted standard for most animal barcoding studies (Hebert *et al.*
133 2003; Ratnasingham & Hebert 2013), ND4 likely serves as a good proxy due to the linked
134 inheritance of mtDNA. Full sampling information can be found in Supplementary Table S1. All
135 gene trees and alignments used in this study can be found on Dryad (doi:10.5061/dryad.r7989).

136

137 *Phylogenetic analysis*

138 The primary goal of our study was to test multiple methods of single-locus species delimitation
139 of horned lizards based on ND4 GenBank data. We were particularly interested in testing for
140 similarities and differences in those methods requiring an ultrametric tree (GMYC-type methods)
141 versus those that do not rely on temporal calibration (PTP-type methods). We were also
142 interested in the performance of both types of methods under scenarios with divergent rates of
143 coalescence across species. We tested a total of four tree-based methods on the data, including
144 the single-threshold GMYC (sGMYC; Pons *et al.* 2006; Fujisawa & Barraclough 2013),
145 multiple-threshold GMYC (mGMYC; Monaghan *et al.* 2009), bPTP (Zhang *et al.* 2013), and
146 mPTP (Kapli *et al.* 2016). All data were aligned using MUSCLE v.3.8.31 (Edgar 2004)
147 implemented in AliView v.1.17.1 (Larsson 2014). The total alignment consisted of 871 bp,
148 although sequence lengths for the *P. douglasii* complex were shorter (alignments available on
149 Dryad). Prior to species-delimitation analyses, we used RAxML v.8.0.0 (Stamatakis 2014) to
150 remove duplicate haplotypes from a matrix of 368 sequences. This left a total of 220 haplotypes
151 for species delimitation. Although identical sequences should generally be removed prior to tree-
152 based methods of species delimitation (J. Zhang, pers. comm.), we also performed a series of
153 duplicate analyses using all 368 sequences for comparison. We used MEGA7 (Kumar *et al.*
154 2016) to calculate average within-species genetic distances using the Tamura-Nei model with
155 gamma distributed rate heterogeneity to account for multiple substitutions. In addition, we used
156 the R-packages APE (Paradis *et al.* 2004) and PEGAS (Paradis 2010) to calculate Watterson's
157 estimator of theta ($\theta = 4N_e\mu$) as an indicator of effective population sizes.

158 We used BEAST v.2.4.3 (Bouckaert *et al.* 2014) to generate ultrametric gene trees under
159 a strict clock and constant-size coalescent tree prior following the relative performance of clock
160 models in previous studies (Monaghan *et al.* 2009; Satler *et al.* 2013; Talavera *et al.* 2013). A
161 GTR+I+ Γ model of substitution was used as estimated using BIC in jModelTest2 (Darriba *et al.*
162 2012). We calibrated the rate of mtDNA substitution by specifying a normal prior with a mean of
163 0.00805 substitutions/site/million years and sigma of 0.001 (Bryson *et al.* 2012), a rate initially
164 estimated using vicariant scenarios in geckos (Macey *et al.* 1999). Similar substitution rates have
165 been applied in numerous studies to estimate divergence times in several vertebrate groups (see
166 Macey *et al.* 1999; Bryson *et al.* 2012 for examples), and our specified prior distribution
167 accommodated uncertainty in the estimate. Analyses were run for 20 million generations,
168 sampling every 2,000. Convergence and mixing were monitored in Tracer v.1.6 (Rambaut *et al.*
169 2014), and ESS values >200 indicated adequate sampling of the posterior. TreeAnnotator v.2.4.3
170 (Bouckaert *et al.* 2014) was used to create a maximum clade credibility (MCC) tree using mean
171 heights for node annotation. Maximum-likelihood (ML) phylogenetic analyses were
172 implemented in RAxML under a GTRGAMMA model by first implementing a rapid bootstrap
173 search (Stamatakis *et al.* 2008) with autoMRE bootstopping followed by a full ML search (-f a
174 option). Trees were rooted using *P. asio* (Leaché & Linkem 2015).

175

176 *Species delimitation analyses*

177 Among the multiple methods of single-locus species delimitation currently available, the most
178 popular is the GMYC model (Pons *et al.* 2006; Fujisawa & Barraclough 2013). Previous studies
179 indicate that the GMYC model is fairly robust (Fujisawa & Barraclough 2013), especially if
180 applied to an ultrametric tree constructed using BEAST (Tang *et al.* 2014), and that the choice of

181 clock model and tree prior has a relatively low impact on the results (Talavera *et al.* 2013). We
182 used the R package SPLITS (Ezard *et al.* 2009) to fit both the single- and multiple-threshold
183 models to the data. We initially included haplotypes only for species containing multiple
184 sequences (*P. mcallii*, *P. platyrhinos*, *P. orbiculare*, *P. hernandesi*, *P. douglasii* complex).
185 However, sGMYC results were not significantly different from the null model of coalescence.
186 Thus, we added singletons of *P. cornutum*, *P. coronatum*, *P. solare*, *P. asio*, and *P. taurus* to
187 increase the Yule portion of the tree and better fit the model to the data (Talavera *et al.* 2013).

188 bPTP analyses were performed using the online server (<http://species.h-its.org/>) and the
189 ML trees from RAxML. We ran the analyses for 500,000 generations with a thinning of 500 and
190 burnin of 0.1. Convergence was assessed by visualizing plots of MCMC iteration versus log-
191 likelihood. We ran analyses both with and without the outgroup taxon (*P. asio*). As results were
192 qualitatively similar (not shown), all subsequent comparisons were made with the outgroup to
193 negate taxonomic discrepancy among analyses. We compared the results from bPTP to the
194 recently developed mPTP model that accommodates different rates of coalescence within clades
195 (Kapli *et al.* 2016). Discordant coalescent patterns could be due to uneven sampling intensity
196 among species or varying degrees of genetic structure arising from differences in evolutionary
197 processes and effective population sizes (N_e). We performed both ML and MCMC analyses on
198 RAxML ML trees using the standalone mPTP software (v.0.1.1). MCMC analyses were run for
199 100 million generations, sampling every 10,000. The first 2 million generations were discarded
200 as burnin and analyses started from the ML species delimitation estimate (identical results were
201 obtained when starting from both random and null delimitations). Convergence was again
202 assessed by monitoring the plot of generation versus log-likelihood. Both ML and MCMC
203 analyses utilized the *--multi* option to incorporate differences in rates of coalescence among

204 species and used a minimum branch length of 0.0001. We compared results among multiple
205 MCMC runs (10) to assess congruence.

206 Because our taxonomic sampling for the four target taxa was highly uneven (i.e., *P.*
207 *platyrhinos* was represented by ~3x the number of haplotypes), we re-ran all of the above
208 analyses after pruning haplotypes from *P. platyrhinos* to determine the potential influence of
209 highly heterogeneous sampling intensity on species delimitation. Haplotypes were selected (40
210 of 111 total) based on the previous phylogenetic analyses to maximize diversity within the
211 species. A new model of substitution was then calculated (HKY+I+ Γ) and all phylogenetic and
212 species delimitation analyses were repeated as described above. However, only unique
213 haplotypes (149) were included in this set of analyses.

214 To compare the results from the tree-based methods above, we ran Automatic Barcode
215 Gap Discovery (ABGD; Puillandre *et al.* 2012a) on both the full and pruned data sets. ABGD is
216 a computationally efficient distance-based method of species delimitation that has been shown to
217 perform well when compared to tree-based coalescent methods (Puillandre *et al.* 2012b;
218 Kekkonen & Hebert 2014; Kapli *et al.* 2016) and other threshold techniques (Ratnasingham &
219 Hebert 2013). The method seeks to quantify the location of the barcode gap that separates intra-
220 from interspecific distances. As the presence of singletons may bias the analysis (Puillandre *et al.*
221 2012a), ABGD analyses were restricted to *P. mcallii*, *P. orbiculare*, *P. platyrhinos*, and the *P.*
222 *douglasii* complex. Default settings were used for the prior range for maximum intraspecific
223 divergence (0.001, 0.1). Results were compared using both JC69 and K80 corrected distances
224 and minimum slope increase (X) of 1.5 (default) and 1.0.

225 For all analyses, we reported the number of delimited species inferred by each method
226 along with the corresponding confidence intervals. In addition, we used current horned lizard

227 taxonomy (Leaché & Linkem 2015; Montanucci 2015), to compare the proportion of delimited
228 species matching taxonomic species, the proportion of taxonomic species lumped into a
229 delimited species, and the number of taxonomic species splits. We note that large values for
230 species lumps and splits may not necessarily indicate poor performance of methods, but they do
231 provide evidence that horned lizard taxonomy might be in need of revision. Next, we reported
232 the match ratio (following Ahrens *et al.* 2016) using the following formula:

$$233 \quad \text{match ratio} = 2 * \frac{N_{\text{match}}}{(N_{\text{delimited}} + N_{\text{morph}})}$$

234 where N_{match} is the number of delimited species exactly matching taxonomic species, $N_{\text{delimited}}$ is
235 the total number of delimited species, and N_{morph} is the number of taxonomic, morphologically
236 defined species. Finally, we quantified performance of methods using the recently developed
237 Relative Taxonomic Resolving Power Index (R_{tax}) and the Taxonomic Index of Congruence
238 (C_{tax}) following Miralles & Vences (2013). The R_{tax} index quantifies the relative power of a
239 method to infer all estimated speciation events present in a data set (large R_{tax} = small Type-II
240 error), but does not necessarily imply correct delimitations (i.e. can lead to oversplitting). R_{tax}
241 metrics were calculated as follows:

$$242 \quad R_{\text{tax}}(A) = \frac{nA}{n(A \cup B \cup C \cup D \cup E)}$$

243 where A,B,C,D,E represent the five species delimitation methods tested, the numerator (nA)
244 represents the number of speciation events inferred by method A, and the denominator represents
245 the cumulative number of speciation events inferred by all methods. An R_{tax} value of 1 would
246 indicate that the method recovered all speciation events present across methods.
247 The C_{tax} index is a measure of congruence in species assignments among two methods, with a
248 value of 1 indicating complete congruence. C_{tax} metrics were calculated as follows:

249
$$C_{tax}(AB) = \frac{n(A \cap B)}{n(A \cup B)}$$

250 where $A \cap B$ represents the number of speciation events shared by methods A and B , and $A \cup B$
251 represents the total number of speciation events inferred by method A and/or B . We refer the
252 reader to the original publication for additional descriptions of these metrics (Miralles & Vences
253 2013).

254

255 **Results and Discussion**

256

257 *Uneven sampling*

258 Inferred genealogies were congruent with previous studies based on mtDNA (Reeder &
259 Montanucci 2001; Hodges & Zamudio 2004; Leaché & McGuire 2006) and discordant from
260 nuclear phylogenies (Leaché & Linkem 2015), presumably due to historical introgression
261 (Leaché & McGuire 2006; Leaché & Linkem 2015). Our inferred divergence times based on an
262 assumed mtDNA substitution rate of 0.00805 substitutions per site per million years were
263 congruent with times previously inferred using secondary calibration information for the crown
264 age of phrynosomatids (Leaché & Linkem 2015). Average pairwise sequence divergence varied
265 considerably among species, from relatively low levels in *P. mcallii* (0.94%) to high divergence
266 within the *P. douglasii* complex (7.7%; Table 1). Similarly, effective population sizes varied by
267 an order of magnitude.

268 All MCMC species delimitation analyses indicated adequate convergence based on
269 visualizing plots of generation versus likelihood score. Because no differences were detected
270 when using a single run versus 10 independent MCMC runs, we present results from single runs

271 only. The number of horned lizard species inferred by each delimitation method was
272 substantially different between the full (220 haplotypes) and reduced (149 haplotypes) data sets.

273 For the full data set, sGMYC was the most conservative method, inferring a total of 10
274 species with wide confidence intervals (Table 1; Fig. 1). The comparatively small number of
275 species inferred with sGMYC was likely a result of the method having difficulty in locating the
276 threshold point in the data (Supplementary Fig. S1A). sGMYC is expected to work well when
277 there is a clear demarcation in branching rates between versus within species (Pons *et al.* 2006;
278 Reid & Carstens 2012; Esselstyn *et al.* 2012; Fujisawa & Barraclough 2013), which was not the
279 case in the horned lizard data as the threshold was placed relatively deep in the genealogy.
280 Conversely, mGMYC suggested an unrealistically large number of species (81), a high
281 proportion of splits (0.95), and low match ratio (0.13), and appeared to have an equally difficult
282 time placing transition points between inter- and intraspecific branching processes
283 (Supplementary Fig. S1B). mGMYC suggested up to 31 species within *P. platyrhinos* alone (Fig.
284 1), which is unlikely given the low levels of divergence within the species (Table 1). bPTP also
285 suggested an unrealistically high number of horned lizard species (52) with wide confidence
286 intervals from MCMC analyses. However, unlike mGMYC, bPTP inferred only four species
287 within *P. platyrhinos* (Fig. 1). Both mGMYC and bPTP inferred a large number of species
288 within the *P. douglasii* complex (29 and 26 species, respectively). mPTP analyses suggested an
289 intermediate number of total species (18) that was the most congruent with the relative levels of
290 structure in the inferred genealogies (Table 1; Fig. 1). These results were consistent with
291 previous findings that the original PTP model tends to oversplit, whereas mPTP is more
292 conservative and likely to represent true species clusters (Kapli *et al.* 2016).

293 ABGD analysis suggested a total of 12 species based on initial partitioning over a range
294 of prior values for maximum intraspecific divergence (Table 1; Supplementary Fig. S2). Results
295 based on JC69 and K80 corrected distances were identical. The number of species decreased to
296 five with a maximum intraspecific divergence prior value (P) of 0.021554, and to one species
297 with a value of 0.035938. Although there is still a lack of consensus of how to interpret
298 discordant ABGD results (Kekkonen & Hebert 2014), previous studies advocate using a P-value
299 of ~ 0.01 (Puillandre *et al.* 2012a), which in our data would result in the recognition of 12 species
300 of horned lizards (excluding the singletons). The relatively low value for the proportion of
301 species matches (0.17) and match ratio (0.20) was likely a result of both excluding singletons
302 from the analysis and the likely presence of multiple undescribed species in the data. Applying
303 the recursive algorithm resulted in a maximum of 20 species when $X = 1.5$ (Supplementary Fig.
304 S2A). This value increased to a maximum of 36 species when $X = 1.0$ (Supplementary Fig. S2B).
305 To be conservative, we focus on the results from the initial partitioning as this scheme has been
306 shown to be more stable across parameter settings and congruent with other species delimitation
307 methods (Puillandre *et al.* 2012a; Puillandre *et al.* 2012b; Kekkonen & Hebert 2014), including
308 those examined in this study.

309 R_{tax} values for the full 220 unique haplotype data set ranged from 0.05 for sGMYC to
310 0.96 for mGMYC (Table 2), further illustrating the tendency of mGMYC to delimit a large
311 number of (likely erroneous) species. Congruence among methods, based on C_{tax} values, was
312 highest between mPTP and ABGD (0.64) and lowest between sGMYC and mGMYC (0.05).
313 mPTP and ABGD also showed the largest mean C_{tax} among methods (Table 2).

314 Results of the full 368 sequence data sets (including identical sequences) revealed
315 varying degrees of sensitivity of methods to the presence of duplicates, with most algorithms

316 suggesting additional species (Supplementary Table S2). mGMYC was by far the most sensitive
317 of the methods compared, inferring a total of 164 species versus 81 species in the unique 220
318 haplotype data set. In general, the performance of all methods (except bPTP) was reduced based
319 on match ratios. Interestingly, mPTP failed to distinguish between *P. cornutum* and *P. solare*
320 even though these species are distantly related (Leaché & Linkem 2015). The performance of
321 ABGD also seemed to be impacted by the inclusion of identical sequences, concordant with
322 other recent findings (Ahrens *et al.* 2016). Thus, although previous studies suggest that the
323 GMYC model may be robust to the inclusion of identical sequences (Talavera *et al.* 2013), our
324 results suggest that additional bias may be introduced when duplicate haplotypes are not
325 removed.

326

327 *Even sampling*

328 Results from our analyses using a relatively even sampling scheme (after pruning the number of
329 *P. platyrhinos* haplotypes to 40) revealed varying levels of sensitivity among the species
330 delimitation methods to uneven sampling among species. sGMYC results were virtually identical
331 to the full analysis due to the same position of the inflection point (Supplementary Fig. S3A), but
332 the confidence interval was substantially reduced. In contrast, mGMYC was extremely sensitive
333 to sampling regime (Figs. 1,2; Table 3), and placed the four thresholds at different times in the
334 evenly sampled genealogy (Supplementary Fig. S3B). The proportion of species splits inferred
335 by mGMYC was substantially reduced in the pruned data set (0.71 vs. 0.95) and the match ratio
336 was increased from 0.13 to 0.39. For example, only a single species within *P. platyrhinos* was
337 inferred by mGMYC for the reduced data set (Fig. 2), compared to 31 species for the full data
338 (Fig. 1). These results were surprising since we pruned *P. platyrhinos* haplotypes evenly

339 throughout the original genealogy. mGMYC also inferred different numbers of species within *P.*
340 *orbiculare* and the *P. douglasii* complex. Thus, in contrast to sGMYC, mGMYC appears quite
341 sensitive to sampling conditions, which may further limit the utility of the method (Esselstyn *et*
342 *al.* 2012; Fujisawa & Barraclough 2013; Talavera *et al.* 2013). Conversely, bPTP and mPTP
343 appear less sensitive to sampling issues as the number of inferred species, proportion of matches,
344 lumps and splits and match ratios were similar between both sets of analyses (Tables 1,3). mPTP
345 was the most consistent, with 18 species inferred from the full data set and 17 in the reduced data
346 set (Figs. 1,2). The discrepancy in the single species arose from a slightly different gene tree for
347 the reduced data set within the *P. platyrhinos* clade. Zhang *et al.* (2013) also tested for the
348 influence of sampling evenness on species delimitation results and found that PTP outperformed
349 GMYC with even sampling, whereas GMYC was slightly more accurate with uneven sampling.
350 Due to the extreme sensitivity of mGMYC to sampling conditions, the largest R_{tax} value was
351 obtained from bPTP (1.00 vs. 0.59 for mGMYC) in the pruned data set (Table 4). Similar to the
352 full 220 haplotype analysis, pairwise C_{tax} was highest between mPTP and ABGD (0.69) and in
353 this case lowest between sGMYC and bPTP (0.11; Table 4).

354 Collectively, these results suggest mGMYC and bPTP were more sensitive to sampling
355 regime, and that the large difference in the inferred number of species between bPTP and mPTP
356 is likely due to the latter fitting multiple exponential branch length distributions to species to
357 account for different rates of coalescence in heterogeneous data sets containing species with
358 different N_e and demographic histories. Thus, our results provide further empirical evidence that
359 mPTP may be a good choice for single-locus species delimitation based on accuracy,
360 consistency, and speed (Kapli *et al.* 2016).

361 ABGD analyses on the evenly sampled (pruned) data set also indicated 12 species for
362 most values of P using the initial partition (Table 3; Supplementary Fig. S4A,B). Once again, the
363 recursive partition suggested a substantially higher number of species, particularly when $P <$
364 0.0129 . There were also slight differences in the recursive partition between K80 and JC69
365 corrected distances (Supplementary Fig. S4A,B). However, K80 distances indicated the same
366 number of groups (12) among the initial and recursive partition when $P = 0.0129$.

367

368 *Performance of methods*

369 Many studies have examined the GMYC model in detail using both simulated and empirical data
370 (Reid & Carstens 2012; Esselstyn *et al.* 2012; Fujisawa & Barraclough 2013; Talavera *et al.*
371 2013; Tang *et al.* 2014; Ahrens *et al.* 2016). Tang *et al.* (2014) quantified the influence of gene
372 tree reconstruction method and rate smoothing technique on the performance of both GMYC and
373 PTP and found that the former was generally more sensitive to the selected model. They found
374 that most of the sensitivity was likely due to errors during the smoothing step and subsequently
375 advocated the use of BEAST to generate ultrametric gene trees. Talavera *et al.* (2013) used a
376 large butterfly data set to test the performance of GMYC and suggested that the model is highly
377 stable under a variety of conditions including tree reconstruction method, the number of
378 singletons included, the number of species included in the gene tree, and sampling coverage.
379 They provided a summary table and chart with recommendations for running GMYC on
380 empirical data sets. Interestingly, their analysis suggested that sGMYC often overestimates the
381 number of species, in contrast to our analysis where the model may be underestimating true
382 diversity. However, our results were concordant in the sense that species delimitations did not
383 change significantly with different sample coverage, although it was slightly impacted by the

384 inclusion of identical sequences. Studies have also indicated that GMYC may be sensitive to
385 general phylogenetic history, sampling intensity, DNA sequence length, speciation rate, N_e , and
386 differences in N_e among species (Reid & Carstens 2012; Esselstyn *et al.* 2012), and may
387 sometimes underestimate (sGMYC) or overestimate (mGMYC) the true number of species
388 (Esselstyn *et al.* 2012). This is a likely explanation for our sGMYC results that estimated only 10
389 species of horned lizards, as there was no abrupt change in branching rates between versus
390 within species. More recently, Ahrens *et al.* (2016) used both simulated and empirical data to
391 better understand potential biases in GMYC due to sampling and population genetic artifacts.
392 Their results suggested that the majority of bias is introduced by variation in N_e among species,
393 which can be exacerbated by uneven species abundance/sampling. In these cases, sGMYC tends
394 to lump species and return wide confidence intervals, which is consistent with our results for
395 horned lizards. To help overcome these issues, they suggest increasing the number of clades
396 examined to balance out the large skew in N_e among species. Although this solution may
397 alleviate some of the issues with sGMYC, mPTP may be more reliable in such cases as the
398 method can explicitly account for differences in N_e and rates of coalescence among species.

399 To our knowledge, few studies have examined the potential influence of methods for
400 summarizing node height information in BEAST analyses for subsequent species delimitation
401 using GMYC. To provide preliminary data on this issue, we performed a suite of additional
402 sGMYC and mGMYC analyses on both the full (368 sequences) and unique haplotype (220
403 sequences) data sets. Three methods of summarizing node heights were compared: mean heights,
404 median heights, common ancestor heights. Different results were detected depending on whether
405 the single- or multiple-threshold model was used (Table 5). Using common ancestor node
406 heights with sGMYC dramatically increased the number of delimited species, due to the

407 threshold point be pushed closer to the present (Supplementary Fig. S5). This effect was
408 negligible with mGMYC, which appeared to be more affected by the inclusion of identical
409 sequences (Supplementary Fig. S6) resulting in a doubling of the number of inferred species
410 (Table 5). Thus, there appears to be additional nuances of GMYC that should be considered
411 when utilizing these methods on empirical data. Based on relative concordance with our mPTP
412 and ABGD analyses, using sGMYC with mean or median node heights may be a good approach.
413 Additional simulation studies will be needed to test this prediction further.

414 Given the high levels of discordance observed among the methods tested, how should
415 researchers use these algorithms to discover and delimit diversity? As detailed above, mPTP has
416 numerous advantages over other methods. The consistency of mPTP to delimit putative species
417 in our study despite varied sampling depths and effective population sizes provides additional
418 evidence suggesting that the model may be appropriate for a wide variety of empirical data sets
419 (Kapli *et al.* 2016). Further, mPTP may be a good choice in data sets such as ours where sGMYC
420 has difficulty in placing the threshold point due to a more gradual slope in branching times,
421 possibly as a result of sampling a low species-to-individual ratio or due to large differences in N_e
422 among species (Ahrens *et al.* 2016). However, we agree with previous authors in that taxonomic
423 changes should not be made solely on the results of these methods (Lohse 2009; Puillandre *et al.*
424 2012a; Esselstyn *et al.* 2012; Zhang *et al.* 2013; Talavera *et al.* 2013), although concordance
425 using multiple analyses does tend to increase reliability (Puillandre *et al.* 2012b; Carstens *et al.*
426 2013; Satler *et al.* 2013). Rather, robust single-locus approaches should be used to form primary
427 taxonomic hypotheses that are subsequently tested with other types of data as part of an
428 integrative taxonomic framework (Fujita *et al.* 2012). Although ABGD has the potential to offer
429 a rapid and robust framework for assessing concordance (Puillandre *et al.* 2012a; Puillandre *et*

430 *al.* 2012b; Ahrens *et al.* 2016), additional work is needed to determine optimal parameter settings
431 and whether the recursive partition tends to oversplit. Moreover, additional studies are needed to
432 compare the performance of RESL through BOLD against some of the newer tree-based
433 methods (e.g., bPTP and mPTP).

434 The adoption of single-locus species delimitation methods to biodiversity research seems
435 particularly relevant to large metabarcoding studies (including microbial 16S rRNA sequencing)
436 as a rapid and cost-effective means to target groups for additional investigation. High-
437 throughput, multiplex amplicon sequencing using next-generation sequencing platforms allows
438 for the rapid generation of single-locus data from a large number of samples for primary species
439 delineation. Due to its speed and accuracy, mPTP seems to be an ideal analytical tool for these
440 large heterogeneous data sets consisting of species with different coalescent histories. We
441 anticipate that empiricists will continue to explore the utility of single-locus methods as
442 sequencing technologies improve and new analytical tools are developed.

443

444 *Taxonomy of horned lizards*

445 Based on the performance and consistency of single-locus species delimitation methods tested in
446 this study, species-level diversity within *Phrynosoma* might be underestimated. Analyses of the
447 mitochondrial gene ND4 provide evidence to suggest at least 11 species could be present: *P.*
448 *mcallii* (1 spp.), *P. platyrhinos* (2 spp.), *P. orbiculare* (5 spp.) and *P. douglasii* complex (3 spp.).
449 We note that this interpretation is conservative and based on congruence between mPTP and
450 ABGD analyses. Within *P. orbiculare*, both methods suggest the presence of one species in the
451 northern Sierra Madre Occidental of Chihuahua, one species in the southern Sierra Madre
452 Occidental, one in the northern Sierra Madre Oriental and Central Mexican Plateau, one in the

453 southern Sierra Madre Oriental and adjacent Trans-Mexican Volcanic Belt (Veracruz and
454 Puebla), and one in the central Trans-Mexican Volcanic Belt (Estado de México and Distrito
455 Federal). mPTP also suggested an additional species in the southern Sierra Madre Oriental in
456 Hidalgo (SMOr-H; Fig. 3).

457 Species delimitation scenarios for the *P. douglasii* complex varied among the methods. In
458 our RAxML gene trees, a strongly supported *P. douglasii* clade (bootstrap=93%) was nested
459 deep within *P. hernandesi* sensu lato (s.l.), albeit with weak support (<50% bootstrap; Fig. 3).
460 However, this relationship resulted in non-monophyly of *P. hernandesi* in RAxML gene trees. In
461 contrast, BEAST analyses reconstructed two major clades, one consisting of haplotypes from *P.*
462 *hernandesi* s.l. and one containing haplotypes from *P. douglasii*. This is one reason why the tree-
463 based species delimitation methods failed to differentiate the two species. In all phylogenetic
464 analyses, however, a single divergent haplotype was found that may represent a third species
465 within this clade (PDU71587 from Arizona), a finding that warrants further evaluation in future
466 studies. ABGD further split *P. hernandesi* into two putative species that correspond to the “SER”
467 and “GB/CP” lineages of Zamudio et al. (1997). There was weak evidence to support the
468 recently proposed taxonomy of Montanucci (2015), who resurrected the names *P. brevisrostris*
469 and *P. ornatissimum* and described two new taxa (*P. bauri* and *P. diminutum*) based on
470 morphological comparisons. *Phrynosoma bauri* and *P. ornatissimum* were represented in our
471 data by one haplotype each, which were both strongly nested within *P. hernandesi* sensu stricto
472 (s.s.) in all phylogenetic analyses. In addition, nearly all species delimitation analyses (including
473 tree-based and distance-based) suggested that these taxa are not distinct from *P. hernandesi*.
474 Only bPTP and mGMYC analyses suggested that these species may be valid, but these analyses
475 generally tended to split *P. hernandesi* s.l. into a large number of singletons that are not

476 consistent with the recovered genealogies, and thus we view these results as unlikely.
477 *Phrynosoma brevirostris* was recovered as a clade deeply nested within *P. hernandesi* s.s. in all
478 the genealogies. Once again, only bPTP and mGMYC suggested that *P. brevirostris* is a distinct
479 species. Interestingly, mGMYC on the full 220 haplotype data set split *P. brevirostris* into
480 multiple species, whereas mGMYC on the evenly sampled data set lumped all haplotypes into a
481 single entity. Much more taxonomic work remains for the *P. douglasii* complex, and it is likely
482 that nuclear data will be required to resolve the discrepancies between morphology and mtDNA
483 data.

484 Within *P. platyrhinos*, our results suggested between two (ABGD) and four (mPTP)
485 species, with strong support from mPTP. Concordance among results suggests two samples from
486 Yuma Proving Ground in La Paz County, Arizona (LVT9951 and LVT9952) may represent a
487 new species, consistent with previous suggestions (Mulcahy *et al.* 2006; Jezkova *et al.* 2016).
488 mPTP results also suggested that sample LVT818 was a separate taxon as was a clade containing
489 samples DGM478 and DGM481, although the latter clade was weakly placed in alternative
490 positions in all of our ML and Bayesian analyses and unlikely to represent a distinct species.
491 Finally, the majority of analyses suggested that *P. mcallii* consists of a single species, concordant
492 with the low level of genetic diversity in the species (Mulcahy *et al.* 2006). However, mGMYC
493 on the full 368 sequence data set suggested a highly unrealistic estimate of 18 species within *P.*
494 *mcallii* (Supplementary Table S2), further illustrating the propensity of this method to oversplit.

495 We note that although many of the tree-based methods we tested are robust under a
496 variety of scenarios, poorly supported nodes and/or non-monophyly in gene trees may render
497 results unreliable until further data are collected—a broad limitation of the single-locus approach
498 (Esselstyn *et al.* 2012; Fujisawa & Barraclough 2013; Kapli *et al.* 2016). Thus, when delimiting

499 putative species based on single-locus data researchers should consider using both tree- and
500 threshold-based methods like ABGD to account for the shortcomings of each type of method
501 (Hamilton *et al.* 2011; Puillandre *et al.* 2012b). Prior to formal taxonomic changes, results should
502 be subsequently tested with additional types of data and analyses, such as morphological and/or
503 ecological data and multilocus coalescent-based methods (Puillandre *et al.* 2012b; Satler *et al.*
504 2013). This multiple-lines-of-evidence approach can therefore account for gene tree/species tree
505 discordance before formal taxonomic changes are implemented (Fujita *et al.* 2012; Carstens *et*
506 *al.* 2013). Nevertheless, single-locus methods such as mPTP have the potential to quickly and
507 accurately target lineages that warrant additional investigation.

508 **Acknowledgements**

509 We would like to thank A. Stamatakis, P. Kapli, J. Zhang and members of the PTP Discussion
510 Board for their assistance running all bPTP and mPTP analyses; N. Puillandre and S. Brouillet
511 for suggestions on ABGD analyses; and J. Jones for help with creating the *P. orbiculare* image.
512 We would also like to thank four reviewers who provided valuable comments to improve the
513 manuscript.

514 **References**

- 515 Ahrens D, Fujisawa T, Krammer HJ, Eberle J, Fabrizi S, Vogler AP (2016) Rarity and
516 incomplete sampling in DNA-based species delimitation. *Systematic Biology*, **65**, 478–
517 494.
- 518 Barraclough TG, Hughes M, Ashford-Hodges N, Fujisawa T (2009) Inferring evolutionarily
519 significant units of bacterial diversity from broad environmental surveys of single-locus
520 data. *Biology Letters*, **5**, 425–428.
- 521 Blair C, Méndez de la Cruz FR, Law C, Murphy RW (2015) Molecular phylogenetics and
522 species delimitation of leaf-toed geckos (Phyllodactylidae: *Phyllodactylus*) throughout
523 the Mexican tropical dry forest. *Molecular Phylogenetics and Evolution*, **84**, 254–265.
- 524 Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal*
525 *Society of London. Series B, Biological Sciences*, **359**, 669–679.
- 526 Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A Software Platform for Bayesian
527 Evolutionary Analysis. *PLOS Computational Biology*, **10**, e1003537.
- 528 Bryson RW, García-Vázquez UO, Riddle BR (2012) Diversification in the Mexican horned
529 lizard *Phrynosoma orbiculare* across a dynamic landscape. *Molecular Phylogenetics and*
530 *Evolution*, **62**, 87–96.
- 531 Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation.
532 *Molecular Ecology*, **22**, 4369–4383.
- 533 Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics
534 and parallel computing. *Nature Methods*, **9**, 772–772.
- 535 de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–
536 886.

537 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
538 throughput. *Nucleic Acids Research*, **32**, 1792–1797.

539 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*,
540 **26**, 2460–2461.

541 Esselstyn JA, Evans BJ, Sedlock JL, Anwarali Khan FA, Heaney LR (2012) Single-locus species
542 delimitation: a test of the mixed Yule-coalescent model, with an empirical application to
543 Philippine round-leaf bats. *Proceedings of the Royal Society of London B: Biological
544 Sciences*, **279**, 3678–3686.

545 Ezard T, Fujisawa T, Barraclough TG (2009) SPLITS: species’ limits by threshold statistics. *R
546 package version*, **1**.

547 Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the
548 Generalized Mixed Yule Coalescent approach: a revised method and evaluation on
549 simulated data sets. *Systematic Biology*, **62**, 707–724.

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

552 Gebiola M, Gómez-Zurita J, Monti MM, Navone P, Bernardo U (2012) Integration of molecular,
553 ecological, morphological and endosymbiont data for species delimitation within the
554 *Pnigalio soemius* complex (Hymenoptera: Eulophidae). *Molecular Ecology*, **21**, 1190–
555 1208.

556 Hamilton CA, Formanowicz DR, Bond JE (2011) Species delimitation and phylogeography of
557 *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): cryptic diversity in
558 North American tarantulas. *PLOS ONE*, **6**, e26207.

559 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA
560 barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–
561 321.

562 Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Systematic
563 Biology*, **54**, 852–859.

564 Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA
565 barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*.
566 *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.

567 Hodges WL, Zamudio KR (2004) Horned lizard (*Phrynosoma*) phylogeny inferred from
568 mitochondrial genes and morphological characters: understanding conflicts using
569 multiple approaches. *Molecular Phylogenetics and Evolution*, **31**, 961–971.

570 Jezkova T, Jaeger JR, Oláh-Hemmings V *et al.* (2016) Range and niche shifts in response to past
571 climate change in the desert horned lizard *Phrynosoma platyrhinos*. *Ecography*, **39**, 437–
572 448.

573 Jones GR (2014) STACEY: species delimitation and phylogeny estimation under the
574 multispecies coalescent. *bioRxiv*, 010199.

575 Kapli P, Lutteropp S, Zhang J *et al.* (2016) Multi-rate Poisson Tree Processes for single-locus
576 species delimitation under Maximum Likelihood and Markov Chain Monte Carlo.
577 *bioRxiv*, 063875.

578 Kekkonen M, Hebert PDN (2014) DNA barcode-based delineation of putative species: efficient
579 start for taxonomic workflows. *Molecular Ecology Resources*, **14**, 706–715.

580 Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis
581 version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.

582 Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets.
583 *Bioinformatics*, btu531.

584 Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos
585 (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological*
586 *Sciences*, rspb20100662.

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

589 Leaché AD, Koo MS, Spencer CL *et al.* (2009) Quantifying ecological, morphological, and
590 genetic variation to delimit species in the coast horned lizard species complex
591 (*Phrynosoma*). *Proceedings of the National Academy of Sciences*, **106**, 12418–12423.

592 Leaché AD, Linkem CW (2015) Phylogenomics of horned lizards (Genus: *Phrynosoma*) using
593 targeted sequence capture data. *Copeia*, **103**, 586–594.

594 Leaché AD, McGuire JA (2006) Phylogenetic relationships of horned lizards (*Phrynosoma*)
595 based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene
596 tree. *Molecular Phylogenetics and Evolution*, **39**, 628–644.

597 Lohse K (2009) Can mtDNA barcodes be used to delimit species? A response to Pons *et al.*
598 (2006). *Systematic Biology*, **58**, 439–442.

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

602 Monaghan MT, Wild R, Elliot M *et al.* (2009) Accelerated species inventory on Madagascar
603 using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

604 Montanucci RR (2015) A taxonomic revision of the *Phrynosoma douglasii* species complex
605 (Squamata: Phrynosomatidae). *Zootaxa*, **4015**, 001–177.

606 Mulcahy DG, Spaulding AW, Mendelson JR, Brodie ED (2006) Phylogeography of the flat-
607 tailed horned lizard (*Phrynosoma mcallii*) and systematics of the *P. mcallii-platyrrhinos*
608 mtDNA complex. *Molecular Ecology*, **15**, 1807–1826.

609 Nieto-Montes de Oca A, Arenas-Moreno D, Beltrán-Sánchez E, Leaché AD (2014) A new
610 species of horned lizard (Genus *Phrynosoma*) from Guerrero, México, with an updated
611 multilocus phylogeny. *Herpetologica*, **70**, 241–257.

612 Paradis E (2010) pegas: an R package for population genetics with an integrated-modular
613 approach. *Bioinformatics*, **26**, 419–420.

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

616 Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for
617 the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

618 Puillandre N, Lambert A, Brouillet S, Achaz G (2012a) ABGD, Automatic Barcode Gap
619 Discovery for primary species delimitation. *Molecular Ecology*, **21**, 1864–1877.

620 Puillandre N, Modica MV, Zhang Y *et al.* (2012b) Large-scale species delimitation method for
621 hyperdiverse groups. *Molecular Ecology*, **21**, 2671–2691.

622 Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) *Tracer v1. 6*.

623 Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode
624 Index Number (BIN) system. *PLoS One*, **8**, e66213.

625 Reeder TW, Montanucci RR (2001) Phylogenetic analysis of the horned lizards
626 (Phrynosomatidae: *Phrynosoma*): evidence from mitochondrial DNA and morphology.
627 *Copeia*, **2001**, 309–323.

628 Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy of
629 species delimitation: a Bayesian implementation of the general mixed Yule-coalescent
630 model. *BMC Evolutionary Biology*, **12**, 196.

631 Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of
632 morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae,
633 *Aliatypus*). *Systematic Biology*, **62**, 805–823.

634 Sherbrooke WC (2003) *Introduction to Horned Lizards of North America*. University of
635 California Press.

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

638 Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web
639 servers. *Systematic Biology*, **57**, 758–771.

640 Talavera G, Dincă V, Vila R (2013) Factors affecting species delimitations with the GMYC
641 model: insights from a butterfly survey. *Methods in Ecology and Evolution*, **4**, 1101–
642 1110.

643 Tang CQ, Humphreys AM, Fontaneto D, Barraclough TG, Paradis E (2014) Effects of
644 phylogenetic reconstruction method on the robustness of species delimitation using
645 single-locus data. *Methods in Ecology and Evolution*, **5**, 1086–1094.

646 Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data.
647 *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.

648 Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from
649 multiple loci. *Molecular Biology and Evolution*, msu279.

650 Zamudio KR, Jones KB, Ward RH (1997) Molecular systematics of short-horned lizards:
651 Biogeography and taxonomy of a widespread species complex. *Systematic Biology*, **46**,
652 284–305.

653 Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with
654 applications to phylogenetic placements. *Bioinformatics*, **29**, 2869–2876.

655

656 **Author Contributions**

657 Designed study (CB), performed statistical analyses (CB), contributed to manuscript writing and
658 figure preparation (CB, RWB).

659

660 **Data Accessibility**

661 All multiple sequence alignments and gene trees can be found on the Dryad Digital Repository
662 ([doi:10.5061/dryad.r7989](https://doi.org/10.5061/dryad.r7989)).

Table 1. Number of horned lizard species (*Phrynosoma*) inferred by each single-locus species delimitation method tested on the full data set of unique haplotypes. Results from GMYC and PTP are from genealogies containing 220 ND4 haplotypes (n) with highly heterogeneous sampling intensity of target taxa. All bPTP and mPTP results are from Bayesian MCMC analyses. Confidence intervals for totals are in parentheses. ABGD results are based on the initial partitioning scheme with a maximum intraspecific diversity value of 0.012915 (K80 distances). Singletons were pruned prior to ABGD analysis. Taxonomy for *P. douglasii* complex follows Montanucci (2015) and includes *P. douglasii*, *P. hernandesi*, *P. bauri*, *P. brevirostris*, and *P. ornatissimum*. Also shown are average corrected pairwise distances (substitutions/site) for each species containing multiple haplotypes and Watterson's theta ($\theta = 4N_e\mu$). NA= Not applicable. 'Species matched' refers to the proportion of delimited species matching defined taxonomic species; 'Species lumped' indicates the proportion of taxonomic species classified within a delimited species; 'Species splits' represents the proportion of taxonomic species splits by each delimitation method.

Taxon	n	Mean Tamura-Nei distance	Watterson's theta	GMYC single	GMYC multiple	bPTP	mPTP	ABGD
<i>P. douglasii</i> complex	41	0.0774	115.6936	2	29	26	2	4
<i>P. orbiculare</i>	34	0.0633	56.985	1	14	11	6	5
<i>P. mcallii</i>	29	0.0094	12.4772	1	2	1	1	1
<i>P. platyrhinos</i> *	111	0.0211	35.2124	1	31	4	4	2
<i>P. asio</i>	1	NA	NA	1	1	1	1	-
<i>P. cornutum</i>	1	NA	NA	1	1	1	1	-
<i>P. coronatum</i>	1	NA	NA	1	1	1	1	-
<i>P. solare</i>	1	NA	NA	1	1	1	1	-
<i>P. taurus</i>	1	NA	NA	1	1	1	1	-
Total	220	0.1258 [¶]	121.1346 [†]	10 (7-98) ^a	81 (69-103) ^b	52.40 (37-84) ^c	18 (16-21) ^d	12
Species matched (percentage of total)				0.80	0.07	0.15	0.33	0.17
Species lumped (percentage of total)				0.50	0.02	0.04	0.28	0.42
Species splits (percentage of total)				0.20	0.95	0.74	0.67	0.83
Match ratio				0.70	0.13	0.24	0.39	0.20

663 *a* = likelihood ratio versus null model 10.90506, *p* = 0.0043; *b* = likelihood ratio versus null model 23.85364, *p* < 0.001; *c* = mean number of species; *d* =
664 Central Credible Interval; * = ~3x number of haplotypes versus *P. douglasii* complex, *P. orbiculare*, *P. mcallii*; [†] = theta for all 220 sequences; [¶] = mean Tamura-
665 Nei distance for all 220 sequences.

Table 2. Calculation of the Taxonomic Index of Congruence (C_{tax}) and Relative Taxonomic Resolving Power Index (R_{tax}) for all species delimitation methods compared based on the full 220 unique haplotype data set.

	C_{tax}					Mean C_{tax}	R_{tax}	# species
	sGMYC	mGMYC	bPTP	mPTP	ABGD			
sGMYC	-	-	-	-	-	0.22	0.05	5
mGMYC	0.05	-	-	-	-	0.22	0.96	76
bPTP	0.12	0.50	-	-	-	0.3	0.53	42
mPTP	0.33	0.16	0.29	-	-	0.36	0.15	13
ABGD	0.36	0.15	0.27	0.64	-	0.36	0.14	12

666

Total number of speciation events across all methods (excluding singletons) = 78; Mean C_{tax} = average value for method across all pairwise comparisons. Singletons were excluded from calculations to allow fair comparisons (all singletons were pruned prior to ABGD analyses). See text for additional details.

667

Table 3. Number of horned lizard species (*Phrynosoma*) inferred by each single-locus species delimitation method tested on the pruned data set. Results from GMYC and PTP are from genealogies containing 149 ND4 haplotypes (n) with approximately even sampling intensity of target taxa. All bPTP and mPTP results are from Bayesian MCMC analyses. Confidence intervals for totals are in parentheses. ABGD results are based on the initial partitioning scheme with a maximum intraspecific diversity value of 0.012915 (K80 distances). Singletons were pruned prior to ABGD analysis. Taxonomy for *P. douglasii* complex follows Montanucci (2015) and includes *P. douglasii*, *P. hernandesi*, *P. bauri*, *P. brevirostris*, and *P. ornatissimum*. Also shown are average corrected pairwise distances (substitutions/site) for each species containing multiple haplotypes and Watterson's theta ($\theta = 4N_e\mu$). NA= Not applicable. 'Species matched' refers to the proportion of delimited species matching defined taxonomic species; 'Species lumped' indicates the proportion of taxonomic species classified within a delimited species; 'Species splits' represents the proportion of taxonomic species splits by each delimitation method.

Taxon	n	Mean Tamura-Nei distance	Watterson's theta	GMYC single	GMYC multiple	bPTP	mPTP	ABGD
<i>P. douglasii</i> complex	41	0.0774	115.6936	2	11	22	2	4
<i>P. orbiculare</i>	34	0.0633	56.985	1	10	12	6	5
<i>P. mcallii</i>	29	0.0094	12.4772	1	1	1	1	1
<i>P. platyrhinos</i>	40	0.0256	34.7945	1	1	3	3	2
<i>P. asio</i>	1	NA	NA	1	1	1	1	-
<i>P. cornutum</i>	1	NA	NA	1	1	1	1	-
<i>P. coronatum</i>	1	NA	NA	1	1	1	1	-
<i>P. solare</i>	1	NA	NA	1	1	1	1	-
<i>P. taurus</i>	1	NA	NA	1	1	1	1	-
Total	149	0.1468 [¶]	129.0831 [†]	10 (3-19) ^a	28 (9-66) ^b	46.74 (34-68) ^c	17 (11-19) ^d	12
Species matched (percentage of total)				0.80	0.29	0.19	0.35	0.17
Species lumped (percentage of total)				0.50	0.14	0.00	0.29	0.42
Species splits (percentage of total)				0.20	0.71	0.73	0.65	0.83
Match ratio				0.70	0.39	0.30	0.40	0.20

a = likelihood ratio versus null model 10.02172, *p* = 0.0067; *b* = likelihood ratio versus null model 12.4399, *p* = 0.002; *c* = mean number of species; *d* = Central Credible Interval; [†] = theta for all 149 sequences; [¶] = mean Tamura-Nei distance for all 149 sequences.

Table 4. Calculation of the Taxonomic Index of Congruence (C_{tax}) and Relative Taxonomic Resolving Power Index (R_{tax}) for all species delimitation methods compared based on the pruned 149 unique haplotype data set.

	C_{tax}					Mean C_{tax}	R_{tax}	# species
	sGMYC	mGMYC	bPTP	mPTP	ABGD			
sGMYC	-	-	-	-	-	0.25	0.11	5
mGMYC	0.18	-	-	-	-	0.40	0.59	23
bPTP	0.11	0.59	-	-	-	0.33	1.00	38
mPTP	0.36	0.38	0.30	-	-	0.43	0.30	12
ABGD	0.36	0.43	0.30	0.69	-	0.45	0.30	12

Total number of speciation events across all methods (excluding singletons) = 37; Mean C_{tax} = average value for method across all pairwise comparisons. Singletons were excluded from calculations to allow fair comparisons (all singletons were pruned prior to ABGD analyses). See text for additional details.

Table 5. Comparison of the number of delimited horned lizard (*Phrynosoma*) species by the single- (sGMYP) and multiple-threshold (mGMYP) GMYP models based on different methods of annotating node height and sampling regimes. 'Unique' = only unique haplotypes (220); 'All' = all sequences (368). Three ways to summarize node heights on BEAST maximum clade credibility trees were evaluated ('Mean heights', 'Common ancestor heights', 'Median heights'). Values represent the number of ML entities with confidence intervals.

Data set	Node heights	sGMYP	mGMYP
Unique	Mean heights	10 (7-98)**	81 (69-103)***
Unique	Common ancestor heights	64 (58-72)***	66 (61-66)***
Unique	Median heights	9 (7-69)**	76 (68-103)***
All	Mean heights	13 (8-138)*	164 (157-164)***
All	Common ancestor heights	79 (76-87)***	100 (83-100)***
All	Median heights	9 (7-16)*	191 (13-299)***

* = $P < 0.05$; ** = $P < 0.005$; *** = $P < 0.001$

669

670

671 **Figure Legends**

672

673 **Fig. 1.** Comparison of species delimitation results of horned lizards (*Phrynosoma*) based on
674 analysis of 220 unique ND4 haplotypes and highly heterogeneous taxonomic sampling. Each
675 coloured bar represents a species delimited by each method tested. Gene tree is from a BEAST
676 analysis under a strict clock and constant size coalescent tree prior. Node height was determined
677 using mean heights across the posterior distribution. Node values represent Bayesian posterior
678 probabilities (>0.95) for major clades. sGMYC = single threshold GMYC; mGMYC = multiple
679 threshold GMYC; bPTP = single-rate Poisson Tree Processes model (MCMC analysis of a
680 RAxML gene tree) fitting a single branch length distribution to coalescent events; mPTP =
681 multi-rate Poisson Tree Processes model (MCMC analysis of a RAxML gene tree) fitting
682 multiple branch length distributions to coalescent events across distinct species. ABGD =
683 Automatic Barcode Gap Discovery. Note that singletons (i.e. *P. asio*, *P. taurus*, *P. cornutum*, *P.*
684 *solare*, *P. coronatum*) were removed prior to ABGD analysis. * = OTUs (PDU71587,
685 PDU71589) clustered together as a single species based on mPTP analysis of a RAxML gene
686 tree, which yielded a slightly different topology than the BEAST genealogy shown. See Fig. 3
687 for additional details. Refer to online version for a full colour representation of figure.

688

689 **Fig. 2.** Comparison of species delimitation results of horned lizards (*Phrynosoma*) based on
690 analysis of 149 unique ND4 haplotypes with relatively even intraspecific sampling. Each
691 coloured bar represents a species delimited by each method tested. Gene tree is from a BEAST
692 analysis under a strict clock and constant size coalescent tree prior. Node height was determined
693 using mean heights across the posterior distribution. Node values represent Bayesian posterior

694 probabilities (>0.95) for major clades. sGMYC = single threshold GMYC; mGMYC = multiple
695 threshold GMYC; bPTP = single-rate Poisson Tree Processes model (MCMC analysis of a
696 RAxML gene tree) fitting a single branch length distribution to coalescent events; mPTP =
697 multi-rate Poisson Tree Processes model (MCMC analysis of a RAxML gene tree) fitting
698 multiple branch length distributions to coalescent events across distinct species. ABGD =
699 Automatic Barcode Gap Discovery. Note that singletons (i.e. *P. asio*, *P. taurus*, *P. cornutum*, *P.*
700 *solare*, *P. coronatum*) were removed prior to ABGD analysis. * = OTUs (PDU71587,
701 PDU71589) clustered together as a single species based on mPTP analysis of a RAxML gene
702 tree, which yielded a slightly different topology than the BEAST genealogy shown. See Fig. 3
703 for additional details. Refer to online version for a full colour representation of figure.

704

705 **Fig. 3.** Species delimitation of horned lizards (*Phrynosoma*) based on MCMC mPTP analysis of
706 a RAxML gene tree constructed with 220 ND4 sequences under a GTRGAMMA model.
707 Branches are colour-coded to represent speciation (black) or coalescence (red) events. Values at
708 nodes indicate probability of a speciation event based on mPTP MCMC analysis (first value) and
709 maximum likelihood bootstrap proportions using the autoMRE bootstopping criterion in RAxML
710 (second value). YPG = samples LVT9951 and LVT9952 from Yuma Proving Ground, Arizona.
711 SMOcN = northern Sierra Madre Occidental; SMOcS = southern Sierra Madre Occidental;
712 SMOrS = southern Sierra Madre Oriental; TMVB = Trans-Mexican Volcanic Belt; TMVBc =
713 central Trans-Mexican Volcanic Belt; SMOr-H = Sierra Madre Oriental-Hidalgo; SMOrN =
714 northern Sierra Madre Oriental; CMP = Central Mexican Plateau. PSS = *Phrynosoma* singleton
715 species (*P. cornutum*, *P. coronatum*, *P. solare*, *P. taurus*). *P. hernandesi* s.l. = *P. hernandesi*, *P.*

- 716 *bauri*, *P. brevirostris*, *P. ornatissimum*. Horned lizard shown is a *P. orbiculare* from Coahuila,
717 Mexico. Refer to online version for a full colour representation of figure.