View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Bangor University Research Portal

PRIFYSGOL BANGOR / BANGOR UNIVERSITY

PRIFYSGOL
BANGOR
UNIVERSITY

Convergence of multiple markers and analysis methods defines the genetic distinctiveness of cryptic pitvipers

Mrinalini, M.; Thorpe, R.S.; Creer, S.; Lallias, D.S.; Dawnay, L.; Stuart, B.L.; Malhotra, A.

Molecular Phylogenetics and Evolution

DOI:
10.1016/j.ympev.2015.06.001

Published: 08/07/2015

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Mrinalini, M., Thorpe, R. S., Creer, S., Lallias, D. S., Dawnay, L., Stuart, B. L., & Malhotra, A. (2015). Convergence of multiple markers and analysis methods defines the genetic distinctiveness of cryptic pitvipers. *Molecular Phylogenetics and Evolution*, *92*, 266-279. https://doi.org/10.1016/j.ympev.2015.06.001

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

09. Oct. 2020

1 **Convergence of multiple markers and analysis methods defines the genetic**

2 **distinctiveness of cryptic pitvipers**

3

4 Mrinalini[a,*], Roger S. Thorpe[a], Simon Creer[a], Delphine Lallias[a], Louise Dawnay[a], Bryan L.

5 Stuart[b], and Anita Malhotra[a]

6

7 [a]School of Biological Sciences, College of Natural Sciences, Bangor University, Gwynedd,

8 LL57 2UW, UK.

9 [b] North Carolina Museum of Natural Sciences, Raleigh, North Carolina, USA.

10

11 *Corresponding Author

12 Mrinalini

13 Present Address: Department of Biological Sciences

14 National University of Singapore

15 Science Drive 4

16 Singapore 117543

17 Email: dbsmrin@nus.edu.sg

18

19 Abbreviated Title: Cryptic species delimitation using multiple markers and analysis methods.

20

21

22

23 **Abstract**

24 Using multiple markers and multiple analytical approaches is critical for establishing
25 species boundaries reliably, especially so in the case of cryptic species. Despite development
26 of new and powerful analytical methods, most studies continue to adopt a few, with the
27 choice often being subjective. One such example is routine analysis of Amplified Fragment
28 Length Polymorphism (AFLP) data using population genetic models despite disparity
29 between method assumptions and data properties. The application of newly developed
30 methods for analyzing this dominant marker may not be entirely clear in the context of
31 species delimitation. In this study, we use AFLPs and mtDNA to investigate cryptic
32 speciation in the *Trimeresurus macrops* complex that belongs to a taxonomically difficult
33 lineage of Asian pitvipers. We analyze AFLPs using population genetic, phylogenetic,
34 multivariate statistical, and Bayes Factor Delimitation methods. A gene tree from three
35 mtDNA markers provided additional evidence. Our results show that the inferences about
36 species boundaries that can be derived from population genetic analysis of AFLPs have
37 certain limitations. In contrast, four multivariate statistical analyses produced clear clusters
38 that are consistent with each other, as well as with Bayes Factor Delimitation results, and
39 with mtDNA and total evidence phylogenies. Furthermore, our results concur with allopatric
40 distributions and patterns of variation in individual morphological characters previously
41 identified in the three proposed species: *T. macrops sensu stricto, T. cardamomensis,* and *T.*
42 *rubeus*. Our study provides evidence for reproductive isolation and genetic distinctiveness
43 that define these taxa as full species. In addition, we re-emphasize the importance of
44 examining congruence of results from multiple methods of AFLP analysis for inferring
45 species diversity.

46 **Keywords:** AFLP; Dominant marker; Pitviper; Taxonomy; Population structure;
47 *Trimeresurus macrops.*

48 **1 Introduction**

49 Current efforts to discover and delimit species are usually facilitated by DNA
50 sequence-based evolutionary reconstructions. However, this can be biased by locus-specific
51 evolutionary constraints and genomic non-representation. In the case of recently diverged
52 species, short nuclear sequences are often phylogenetically uninformative (e.g. Bardeleben et
53 al. 2005, Weisrock et al. 2010). Multilocus markers are therefore highly recommended and
54 offer quantitative advantages and genome-wide coverage (Zhang and Hewitt 2003, Meyer
55 and Paulay 2005, Brito and Edwards 2009, Dupuis et al. 2012, Leaché et al. 2014). Routine
56 phylogenomic analysis is still constrained by issues such as differences in results across
57 methods, the need to integrate evolutionary histories of multiple loci, the lack of guidelines
58 for best practices, and extensive computational requirements (Song et al. 2012, Gatesy and
59 Springer 2013, Dell'Ampio et al. 2014, Faria et al. 2014, Leaché et al. 2014). Nonetheless,
60 new methods have been developed and high-throughput sequence analysis is gaining
61 popularity in evolutionary and speciation research (Morin et al. 2010, Springer et al. 2012,
62 McCormack 2013, Misof et al. 2014).

63 Alternatively, the use of non-sequence-based multilocus markers has also increased,
64 often revealing surprisingly clear, fine-scale genetic structure undetected by morphology and
65 sometimes even by mtDNA markers (Brown et al. 2007, Egger et al. 2007, Kingston et al.
66 2009, Meudt et al. 2009, Milá et al. 2010). Among these, Amplified Fragment Length
67 Polymorphism (AFLP) (Vos et al. 1995) is a time-tested, cost-effective, and powerful
68 technique requiring no sequence knowledge. AFLPs have continued to prove useful for
69 resolving species-level taxonomy, recovering patterns of speciation, evolutionary histories
70 and inter-relationships, inferring population structure and genetic diversity analyses in a
71 wide-range of animal species, such as butterflies (Kronforst and Gilbert 2008, Quek et al.

72 2010), cichlids (Albertson et al. 1999), salamanders (Wooten et al. 2010), lizards (Ogden and
73 Thorpe 2002), dolphins (Kingston et al. 2009), and pinnipeds (Dasmahapatra et al. 2009).
74
75 **1.1 Dominant marker analysis**
76 Usually, multilocus markers are analyzed using genetic clustering and diversity
77 analysis methods. These are implemented in population genetics models using F-statistics
78 calculations based on allele-frequencies. The uses of these methods with respect to co-
79 dominant datasets have been evaluated in both spatial and non-spatial models (Latch et al.
80 2006, Chen et al. 2007, Frantz et al. 2009). AFLPs, however, are dominant markers and do
81 not allow distinction between homozygous and heterozygous states of an allele. Therefore,
82 AFLP analysis using allele frequency-based population genetics models requires several
83 assumptions to be made. As this results in analytical limitations, it is highly recommended
84 that multiple analysis methods are applied, and inferences are made with high confidence
85 only when results show congruence across methods (Carstens et al. 2013). However, the
86 majority of AFLP studies continue to apply population genetic methods and derive biological
87 inferences with rare discussion of possible analytical biases (Hollingsworth and Ennos 2004,
88 Bonin et al. 2007). One such example is the routine use of the popular, non-spatial clustering
89 program, STRUCTURE that uses a Bayesian MCMC algorithm to infer K – the number of
90 populations. Other such programs include TESS, GENECLUST, and GENELAND that
91 perform Bayesian cluster analysis under spatial models (Guillot et al. 2005a, 2005b, François
92 et al. 2006, Chen et al. 2007, Guillot 2008, Guillot et al. 2008).
93 The algorithm used in STRUCTURE accommodates dominant data by assuming the
94 presence of recessive alleles at a subset of loci that provide partial information about diploid
95 genotypes for the entire dataset (Falush et al. 2007). GENELAND uses geographic
96 coordinates and identifies groups of individuals in Hardy-Weinberg Equilibrium (HWE)

97 (Guillot et al. 2005a, 2005b). GENELAND was upgraded to correct allele frequency
98 estimates from dominant data by taking into account observed genotypes and estimating
99 unknown genotypes using model-based MCMC simulations (Guillot and Santos 2010).
100 However, both STRUCTURE and GENELAND still assume that AFLP null-alleles (i.e.,
101 band absences) are recessive alleles for allele frequency calculations and subsequent K
102 estimation. Common assumptions that drive these analyses (such as HWE in a population and
103 linkage disequilibrium between populations but not within populations) are conceptually not
104 applicable to dominant data. The models and assumptions used for K estimation in
105 STRUCTURE are less than straightforward and need to be used with caution as they could
106 yield inaccurate results (Pritchard et al. 2000). Furthermore, the accuracy of K estimation
107 using dominant datasets in GENELAND is lower than in co-dominant datasets (Guillot and
108 Santos 2010). In some cases, tree-building analysis of AFLPs has performed better at cluster
109 identification due to the absence of population genetics model assumptions (Meudt et al.
110 2009). A certain degree of uncertainty is therefore unavoidable when using population
111 genetic models to estimate the number of K, to assign individuals to each K, and to assess the
112 genetic structure of each K. Hence, the need for new and more appropriate methods for
113 dominant marker analysis has been identified (Hollingsworth and Ennos 2004, Excoffier and
114 Heckel 2006, Bonin et al. 2007, Meudt et al. 2009).

115 Multivariate methods (such as factor and cluster analysis, principal component
116 analysis, Multi-Dimensional Scaling, Molecular Analysis of Variance) implemented outside
117 the confines of population genetics models have been extensively used to analyze AFLPs.
118 More recently, two tools that use a combination of multivariate procedures to analyze
119 multilocus genetic data were developed. Discriminant Analysis of Principal Components
120 (DAPC) was developed in *adegenet* (an R package) as a method for inferring genetic clusters
121 and genetic diversity using dominant data (Jombart et al. 2010). Hausdorf and Hennig (2010)

122 developed *prabclus,* also an R package, for species delimitation and ordination-cluster
123 analysis using both dominant and co-dominant datasets. Both *adegenet* and *prabclus,* have
124 performed better than STRUCTURE in initial studies (Hausdorf and Hennig 2010, Jombart et
125 al. 2010). These methods could be useful for cluster and population structure analyses and
126 speciation research that employ dominant markers.

127 Finally, Leaché et al. (2014) developed a new approach for AFLP and SNP based
128 species delimitation by adapting a method called Single Nucleotide Polymorphism and AFLP
129 Phylogenies (SNAPP) (Bryant et al. 2012). SNAPP produces posterior probability
130 distributions of allele frequency changes and allows species tree estimation without the need
131 for gene tree reconstruction and integration (Bryant et al. 2012). Grummer et al. (2013) first
132 developed sequence-based Bayes Factor Delimitation (BFD) to perform marginal likelihood
133 estimations (MLE) and test multiple species delimitation hypotheses. Species delimitation
134 models are tested at the same time as species tree estimation, forgoing the need to specify a
135 guide species tree (Grummer et al. 2013). Leaché et al. (2014) modified this and developed
136 SNAPP BFD for species delimitation using SNPs and AFLPs. SNAPP BFD is implemented
137 using MLE path sampling analysis in version 2 of Bayesian Evolutionary Analysis Sampling
138 Trees (BEAST) software (Drummond et al. 2012, Bouckaert et al. 2014).
139

140 **1.2 The study group – *Trimeresurus (Trimeresurus) macrops***

141 Asian green pitvipers from the genus *Trimeresurus* (Serpentes: Crotalidae: Crotalinae)
142 (Lacépède 1804) are well known for cryptic speciation (e.g. Malhotra and Thorpe 2000,
143 Vogel et al. 2004, Malhotra and Thorpe 2004a). The genus was divided into several genera in
144 2004 (Malhotra and Thorpe 2004b), among which was *Cryptelytrops* (Cope 1860). Recently,
145 as a result of new information on the type species of *Trimeresurus*, *Cryptelytrops* was shown
146 to be a junior synonym of *Trimeresurus* (David et al. 2011). Therefore, the species placed in

147 *Cryptelytrops* by Malhotra and Thorpe (2004b) are now correctly placed within *Trimeresurus,*
148 whether defined in a broader sense (by subsuming Malhotra and Thorpe's proposed genera as
149 subgenera) or narrower sense (continuing to accept the existence of several well-defined,
150 ecologically, genetically, and morphologically diagnosable generic-level units within the
151 former larger genus).

152 *Trimeresurus macrops sensu lato (s.l.)* is distributed across Thailand, Laos, Cambodia,
153 and Viet Nam, and was shown to consist of three cryptic species with disjunct geographic
154 ranges in the highlands of Cambodia (Fig. 1), distinguished by variations in several
155 individual morphological characters corresponding to their allopatric distributions (Malhotra
156 et al. 2011a). The populations have been proposed as three distinct species: (i) *T. macrops*
157 *sensu stricto (s.s.)* found in Thailand, south & central Laos, and northeast Cambodia*,* (ii) *T.*
158 *cardamomensis* (Cardamom Mountains green pitviper), from southeast Thailand and the
159 Cardamom mountains of southwest Cambodia, and (iii) *T. rubeus* (Ruby-eyed green pitviper),
160 found in southern Viet Nam and eastern Cambodia (Malhotra et al. 2011a). The morphology
161 and species ranges for each of these putative species have been fully described (Malhotra et
162 al. 2011a). A multivariate morphometric analysis, however, was not completely successful in
163 separating the three species (Fig. A.1), possibly due to geographic variation within each of
164 the species. Thus, whether the three populations are genetically distinct lineages and are
165 reproductively isolated needs to be clarified to fully support their species status.

166 Here we use multiple genetic markers and analysis methodologies to investigate the
167 genetic distinctiveness of the three proposed species in the *T. macrops* complex. We use three
168 mtDNA markers and multilocus nuclear marker set from AFLPs. We employ eight methods
169 of AFLP analysis incorporating population genetic, phylogenetic, multivariate statistical, and
170 Bayes Factor Delimitation approaches to confirm species boundaries. We provide a

171 comprehensive description of methods and results, and discuss them within the framework of
172 each method for a better understanding of AFLP analysis.
173

174 **2 Materials and methods**

175 **2.1 Sampling scheme**

176 All samples were obtained from field collections, museums and private collections.
177 Geographic distribution of samples within putative species ranges is as shown in Fig.1. Each
178 locality is represented by multiple specimens in most cases. A total of 39 samples were used
179 for mtDNA sequence analysis, including 22 samples of *T. macrops s.s.,* 6 samples of *T.*
180 *cardamomensis* and 7 samples of *T. rubeus*. Samples covered the geographic distribution of
181 much of the known range of *T. macrops s. l.,* and many of these also had morphological data
182 available. Two closely related species were also included: *T. venustus* (3 samples from South
183 Thailand and West Malaysia); and *T. kanburiensis* (1 sample, the only one available in our
184 collection).
185 For AFLP analysis, a total of 50 individuals were genotyped among which 34 were
186 shared with mtDNA analysis. Putative species in *T. macrops s.l.* were represented by: *T.*
187 *macrops s.s.* = 22 samples (20 shared with mtDNA analysis); *T. cardamomensis* = 6 (all
188 shared with mtDNA analysis); and *T. rubeus* = 7 (six shared with mtDNA analysis).
189 Additionally, 15 *T. venustus* were genotyped (three shared with mtDNA analysis). However,
190 the single *T. kanburiensis* sample available could not be successfully genotyped due to very
191 low DNA yield. Full sample details including museum voucher numbers are provided in
192 Table A.1
193

194 **2.2 Experimental methods**

195 ***2.2.1 mtDNA amplification and sequencing***

196 Liver or muscle tissue in 80% ethanol, clippings of ventral scales in 80% ethanol, or
197 blood obtained from the caudal vein preserved in 5% EDTA and SDS–Tris buffer (100mM
198 Tris, 3% SDS) were used. Whole genomic DNA was extracted using standard salt
199 precipitation protocols (Sambrook et al. 1989). Three mitochondrial genes, 12S rRNA (12S),
200 16S rRNA (16S), and NADH4 (ND4), were amplified (as described in Malhotra et al. 2011b),
201 cleaned with shrimp alkaline phosphatase and Exonuclease I (Werle et al. 1994), and
202 sequenced using dye-labelled terminators (ABI PRISM™ BigDye™ Terminator Cycle
203 Sequencing Ready Reaction Kit) on an ABI 3730XL automated sequencer.
204

205 ***2.2.2 AFLP Genotyping***

206 Genomic DNA was extracted using GenElute™ Mammalian Genomic DNA Miniprep
207 Kit (Sigma-Aldrich). Extracts were duplicated for six samples using the same tissue type, and
208 three samples using different tissue types, for repeatability tests, and negative controls
209 (lacking any tissue) were included to monitor contamination. Extract quality was checked on
210 1% Agarose-EtBr gels, DNA was quantified on a NanoDrop ND-1000 Spectrophotometer,
211 and corrected to 10 ngμl^{-1} using 0.1M TE.

212 AFLPs were generated following the general protocol from Whitlock et al. (2008), but
213 with specific modifications as follows. 100 ng DNA was used per sample and 6.9 µl
214 digestion-ligation mix (final concentrations: 1X TA buffer, 0.17 µgμl^{-1} bovine serum albumin,
215 0.059 Uμl^{-1} each of *Eco*RI and *Mse*I enzymes, 0.3X T4 ligase buffer, 0.03 Uμl^{-1} T4 DNA
216 ligase, 0.74 µM each of *Eco* and *Mse* adaptors with 3 µl d_2H_2O) was added to make up a final
217 volume of 16.9 µl. This was incubated at 16°C for 16 hours in a preconditioned water bath in
218 ThermoFast® 96-well plates (ABgene) and diluted by a factor of 1:4 (i.e. to a final volume of
219 50 µl) with d_2H_2O.

220 Pre-selective and selective primer sequences are provided in Table A.2. Fluorophore
221 labelling of selective primers at the 5' end was performed by Applied Biosystems® using
222 6FAM, VIC, and PET labels. Combinations of selective primers and fluorophore scheme are
223 given in Table A.2. Pre-selective amplification (PA) reactions were performed using 1 µl
224 diluted ligated product in 10 µl reactions (final concentrations: 1X PCR Buffer, 2 mM $MgCl_2$,
225 0.2 mM dNTP, 0.5 µM each of pre-selective *Eco*RI and *Mse*I primers, 0.025 $U\mu l^{-1}$
226 Thermoprime *Taq*) with 4.15 µl d_2H_2O. Thermocycling parameters were initial warm-up at
227 94°C for 2 minutes, 20 cycles of denaturing at 94°C for 30 seconds, annealing at 56°C for
228 one minute, extension at 72°C for 2 minutes, and a final extension of 72°C for 10 minutes
229 and 20°C for 5 minutes. PA products were diluted 1:10, and 1 µl of each, along with 5 µl
230 loading buffer, was run on 1.5% Agarose-EtBr gels. Successful PAs resulted in a smear across
231 the whole range of a 500bp ladder. 1 µl of diluted PA product was used for selective
232 amplification (SA) in10 µl reactions with final concentrations same as PA except the primers
233 were replaced by fluorophores and reverse selective primer. Thermocycling parameters were
234 initial warm-up at 94°C for 2 minutes, 12 cycles of denaturing at 94°C for 30 seconds,
235 annealing at 65°C Δ-0.7°C/cycle for 30 seconds, extension at 72°C for 1 minute, 23 cycles of
236 94°C for 30 seconds, annealing at 56°C for 30 seconds, extension at 72°C for 1 minute, and a
237 final extension of 72°C for 10 minutes and 20°C for 5 minutes. SA products were diluted at
238 1:100 and 1 µl from each primer pair was poolplexed in 10 µl formamide along with 0.5 µl
239 GeneScan™ 500 LIZ® Size Standard and the samples were processed on an ABI 3130XL
240 Genetic Analyzer.

241

242 **2.3 AFLP peak scoring**

243 AFLP profiles were visualized and processed in GeneMapper® Software v4.0, and
244 samples with amplification problems for one or more markers were discarded. Several

245 automated and semi-automated AFLP scoring methods have been proposed in an effort to
246 reduce time, error, and subjectivity of peak calling (reviewed in Meudt and Clarke (2007)).
247 We used the semi-automated method proposed by Whitlock et al. (2008) but found that direct
248 application of this method to raw or filtered data (using the specified phenotype-calling
249 threshold) resulted in significant numbers of inaccurate peak-calling from unaccounted false
250 peaks (artefacts, inter-dye pull-ups, shoulder peaks, saturation peaks), peak mobility, and
251 clear peaks failing to get called. Therefore, the data was first checked by eye and corrected
252 for peak mobility, false peaks and uncalled peaks. Mean Peak Height (MPH) was calculated
253 at each locus, a locus-selection threshold of 100 relative fluorescence units was applied. A
254 relative phenotype-calling threshold of 20% of MPH was applied, i.e., all peaks ≥ 20% MPH
255 were marked as present (1) and peaks ≤ 20% MPH were marked absent (0). We found our
256 method, albeit time-consuming, significantly enhanced genotyping accuracy. Repeatability
257 was measured as the number of loci with corresponding band presences across duplicated
258 samples compared to the total number of loci scored.
259

260 **2.4 Data analysis**

261 ***2.4.1 Bayesian phylogenetic analyses***

262 ***2.4.1.1 mtDNA phylogeny*** The mtDNA dataset consisted of 1450 bp including 350 bp of 12S,
263 474 bp of 16S, and 626 bp of ND4. The dataset was partitioned into 12S, 16S, and first,
264 second, and third codon positions for ND4. Models of sequence evolution were inferred in
265 jModelTest 2.1 (Guindon and Gascuel 2003, Darriba et al. 2012) using the Akaike
266 Information Criterion (Posada and Crandall 1998). Mixed-model Bayesian analysis was
267 implemented in MrBayes 3.1 (Ronquist and Huelsenbeck 2003) using the following models:
268 Generalized time-reversible with gamma-distributed rate variation (GTR+G) for 12S and first
269 and second codon positions of ND4; and GTR for 16S and third codon position of ND4.

270 *Viridovipera vogeli* was used as outgroup. Four independent MCMC analyses of 15,000,000
271 cycles each (sampled every 3,000 generations) were performed with one cold chain and three
272 heated chains. The first 25% of trees were discarded as burn-in and a 50% majority-rule
273 consensus tree was constructed from combined post-burn-in trees. Trace plots of clade
274 probabilities were viewed using AWTY (Wilgenbusch et al. 2004).
275
276 ***2.4.1.2 Total evidence and AFLP phylogenies*** Bayesian phylogenetic inference was
277 performed on both AFLP and total evidence (mtDNA + AFLP) datasets. A complex Bayesian
278 model has been developed specifically for AFLP evolution (Luo et al. 2007). However the
279 lack of algorithms and extreme computational burden (40,000 times slower than restriction
280 site model implemented in MrBayes), make its implementation impractical (Koopman et al.
281 2008). We therefore used MrBayes 3.1 under the standard restriction data model by setting
282 the coding bias to 'noabsencesites' for AFLP data to correct data for unobserved all-absence
283 sites (Ronquist and Huelsenbeck 2003). The total evidence dataset was partitioned into AFLP,
284 12S, 16S, and ND4 partitions and model parameters estimated for mtDNA genes (Posada
285 2008). Four independent MCMC analyses were performed with 3 million cycles (sampled
286 every 1000 generations) using one cold chain and three heated chains. The first 300,000 and
287 500,000 runs were discarded as burnin for AFLP and total evidence datasets respectively. A
288 consensus tree was derived from post-burnin trees using all compatible groups. Final 50%
289 majority rule consensus trees were constructed and re-rooted using the basal clade from the
290 mtDNA reconstruction in FigTree v1.1.2 (Rambaut and Drummond 2008).
291
292 ***2.4.2 Bayes Factor Delimitation***
293 Bayes Factor species delimitation was performed using SNAPP BFD (Leaché et al.
294 2014) implemented in BEAST2 (Bouckaert et al. 2014). Five different speciation hypotheses

295 were tested by lumping putative species in *T. macrops s.l.* in several combinations. In each
296 hypothesis, *T. venustus* was included as a separate species. Details of speciation models are as
297 follows with lumping of putative species indicated in parentheses: Model A: *T. macrops s.s.,*
298 *T. cardamomensis, T. rubeus;* Model B: *(T. macrops s.s. + T. cardamomensis + T. rubeus);*
299 Model C: *(T. macrops s.s.+ T. cardamomensis), T. rubeus;* Model D: *(T. macrops s.s. + T.*
300 *rubeus), T. cardamomensis;* Model E: *T. macrops s.s., (T. cardamomensis + T. rubeus).*
301 Marginal likelihood estimation for each model was performed by path sampling conducted in
302 48 steps. MCMC chain length of 100,000 with pre-burnin of 10,000 for each step was
303 sufficient to establish stationarity. The strength of support for competing models was
304 evaluated as per Leaché et al. (2014), using the Kass and Raftery (1995) framework.
305
306 ***2.4.3 Population genetics methods***
307 ***2.4.3.1 Descriptive statistics*** Estimating genetic diversity from dominant data under non-
308 HWE is possible by incorporating population specific inbreeding co-efficients (F_{is}) into
309 calculations of diversity indices such as F_{st} (Yeh et al. 1997, Foll et al. 2008). Although F_{is}
310 values from small populations (less than 10 individuals) could be unreasonable (Holsinger
311 and Lewis 2007), the difficulty of estimating allele frequencies from small populations could
312 be overcome by analyzing a large number of loci (Krauss 2000). Given that sample sizes for
313 two of our putative species were small (six and seven individuals), we incorporated F_{is} values
314 from ABC4F (Foll et al. 2008) into F_{st} calculations in POPGENE (Yeh et al. 1997) and also
315 calculated F_{st} in the *f* - free model in HICKORY 1.1 (Holsinger and Lewis 2007). An
316 AMOVA was performed in *GenAlex v6.3* (Peakall and Smouse 2006) to calculate % genetic
317 variance and Φ_{PT} (a distance-based analog of F_{st}) of populations, based on 9999 random
318 permutations.
319

320 ***2.4.3.2 Detection of outliers*** BayeScan v1.0 (Foll and Gaggiotti 2008) was used to test for
321 outlier loci in AFLP data (Pérez-Figueroa et al. 2010). Model parameters were automatically
322 estimated based on 10 pilot runs (length = 5,000), using default chain parameters (sample
323 size = 5,000, thinning interval = 20, and additional burn-in = 50,000). Jefferey's scale of
324 evidence was set to maximum (decisive) and loci with $\log_{10}$ (Bayes Factor) = 2.0
325 (corresponding to $p = 0.99$) were considered outliers.
326
327 ***2.4.3.3 Estimation of historical gene flow*** Historical gene flow (N_m) among populations was
328 estimated indirectly by the average of effective number of migrants exchanged between
329 populations in each generation. Crow and Aoki's correction was applied, i.e., $N_m = (1-F_{st})/4\ \alpha$
330 F_{st}, where N_m = the number of migrants per generation, and correction factor $\alpha = [n/(n-1)]^2$
331 where n = number of populations (Crow and Aoki 1984).
332
333 ***2.4.3.4 STRUCTURE*** Cluster analysis for dominant data was implemented in the admixture
334 model using correlated allele frequencies – a more accurate model for assigning individuals
335 to closely related groups (Pritchard et al. 2000, Falush et al. 2003, Falush et al. 2007). Ten
336 runs of 100,000 iterations each were performed with K ranging from 1 to 10, and burn-in of
337 10,000 iterations. Since estimating the probability of K, Pr(X|K), is computationally difficult,
338 two *ad hoc* methods were used: Ln Pr(X|K), prescribed by Pritchard et al. (2000), and ΔK
339 based on the second order rate of change of likelihood function with respect to K, proposed
340 by Evanno et al. (2005). Ln Pr(X|K) and ΔK for each K were plotted using STRUCTURE
341 HARVESTER v0.56.4 (Earl and vonHoldt 2012). Assignment tests were performed to obtain
342 the accuracy of assignment of individuals to putative species by including prior population
343 membership information for each sample, and setting K=3 and 4 based on ΔK and Ln
344 Pr(X|K) plots. Q matrices of population membership from 10 replicates were permuted in the

345 GREEDY_OPTION of CLUMPP v1.1.2 for a mean permuted matrix (Jakobsson and
346 Rosenberg 2007). Results were visualized in *Distruct* (Rosenberg 2004).
347
348 ***2.4.3.5 GENELAND*** Preliminary test runs of 200,000 iterations were used to check for
349 appropriateness of correlated and uncorrelated allele frequency model assumptions under
350 both spatial and non-spatial priors. Based on these results, four independent MCMC runs of
351 500,000 iterations each were performed using a spatial prior with coordinate uncertainty
352 fixed at 1km, uncorrelated allele frequencies, minimum and maximum K fixed at 1 and 16,
353 and a burnin of 50,000 generations. An additional run of 1 million iterations and burnin of
354 100,000 generations was also performed to check for differences in K estimation due to
355 increase in number of iterations. We tested the influence of spatial priors by rerunning the
356 analysis after swapping the geographic co-ordinates of sample B44 (whose cluster
357 assignment was incongruent with that of STRUCTURE) with that of sample A144 from north
358 Thailand.
359
360 ***2.4.4 Multivariate statistical approaches***
361 ***2.4.4.1 Principal co-ordinate analysis*** Principal co-ordinate analysis (PCoA), also known as
362 Classical Multidimensional Scaling, is a traditionally used cluster analysis that transforms
363 distances among objects into similarity/dissimilarity matrix (Gower 2005). This matrix is
364 used to position objects in a space of reduced dimensionality while retaining the relationships
365 between them. We performed a PCoA in MVSP v3.13n using Gower General Similarity Co-
366 efficient (Gower 1966, 1971) to derive the similarity index among pairs of taxa. The resulting
367 principal co-ordinates were plotted to visualize taxon clusters.
368

369 ***2.4.4.2 Discriminant analysis of principal components*** Discriminant Analysis of Principal
370 Components (DAPC), implemented in *adegenet* in R, uses the *find.clusters* function to
371 perform a Principal Component Analysis (PCA) and estimate overall genetic variance. When
372 groups (such as species) are unknown, a K-means clustering algorithm divides total variance
373 into among-group and within-group components. We ran K-means clustering several times,
374 allowing K to vary from 1 to 49 in the first instance. Optimal number of clusters/putative
375 species was determined by the lowest Bayesian Information Criteria (BIC). Discriminant
376 Analysis (DA) of Principal Components (PCs) then defines a model which finds groups that
377 maximize among-group genetic variability and minimize within-group variability (Jombart et
378 al. 2010). We retained as many principal components in the preliminary data transformation
379 step as necessary to represent 75% of total genetic variation. The optimal number of PCs to
380 obtain a robust discrimination is estimated in *optim.a.score,* and the quality of discrimination
381 is indicated by *a.score* for each cluster. We performed 30 DAPC simulations in *optim.a.score*
382 for each of the 10 PCs retained. A second PCA, with K allowed to vary from 1 to 10, was
383 performed to better visualize BIC results. A DAPC was performed using appropriate number
384 of PCs to maximize *a.scores* according to *optim.a.score* result.
385
386 ***2.4.4.3 Gaussian clustering*** In *prabclus,* a Non-Metric Multi-Dimensional Scaling (NMDS)
387 is performed on a distance matrix to derive Euclidean variables of genetic dissimilarity
388 between individuals. Jaccard distances (Jaccard 1908) between individuals were calculated
389 from the binary matrix using the *prabinit* function. As with DAPC, BIC was used as an
390 indicator to estimate the number of clusters/putative species. A Gaussian mixture model
391 determined clusters of individuals corresponding to mixtures of normal distributions that
392 account for variation in data. Ten permutations of NMDS were performed on the distance
393 matrix by *kruskal* method in three dimensions using *prabclust* function. To visualize the

394 clusters, the clusters object, showing assignment of individuals to each cluster, was exported
395 to *Rcmdr* (Fox 2005).
396
397 ***2.4.4.4 Hierarchical clustering with seriation*** Finally, AFLP band presence and absence was
398 visualized in *PermutMatrix V1.9.3,* a package originally developed for gene expression
399 analysis (Caraux and Pinloche 2005). A Euclidean distance matrix of dissimilarity was
400 generated and samples & loci were clustered based on McQuitty's unsupervised hierarchical
401 clustering with a multiple heuristic seriation rule. Row-wise (taxa) and column-wise (locus)
402 enumeration was optimized to improve visualization.
403
404 **3 Results**
405 **3.1 Bayesian phylogenetic analyses**
406 ***3.1.1 mtDNA phylogeny*** Partitioned Bayesian mtDNA phylogeny showed *T. macrops s.l.*
407 samples forming three distinct clades (Fig. 2a). Samples from Thailand, Laos, and northeast
408 Cambodia were assigned to the nominate species *T. macrops sensu stricto (T. macrops s.s.)*
409 since this clade included specimens from the type locality (Bangkok, Thailand). The second
410 clade consists of samples from south-eastern provinces of Thailand and southwest Cambodia.
411 Finally, samples from southern Viet Nam and eastern Cambodia formed a sister group to all
412 above clades, and to *T. kanburiensis*. These relationships are strongly supported with
413 posterior probabilities of 100%. Recognizing the second and the third clades as *T. macrops*
414 *s.l.* would render the species polyphyletic. The three clades therefore represent *T. macrops*
415 *s.s., T. cardamomensis,* and *T. rubeus* respectively.
416
417 ***3.1.2 Total evidence phylogeny*** The total evidence phylogeny consisted of mtDNA
418 sequences and AFLP genotypes. The AFLP data comprised of 298 polymorphic loci from

419 330 loci with a repeatability score of 97% (for both duplicates for a given sample and for
420 different tissue types with different storage conditions for a given sample). The total evidence
421 tree (Fig. 2b) was topologically well-resolved and had robust support values. The three clades
422 in *T. macrops s.l.* were well-differentiated, with 100% support values at deeper nodes and
423 mostly high support values at the tips.
424
425 ***3.1.3 AFLP phylogeny*** The Bayesian AFLP tree (Fig. A.2) was poorly resolved at deeper
426 nodes, with low support values. *T. cardamomensis* and *T. venustus* were not distinct from
427 specimens of *T. macrops s.s*.
428
429 **3.2 Bayes Factor Delimitation**
430 For each speciation hypothesis, marginal likelihood values from SNAPP BFD
431 analysis are shown in Table 2. All models derived by lumping of the three putative species in
432 *T. macrops s.l.* in various combinations were decisively rejected based on $2 \times \log_e BF > 10$.
433 Model A consisting of four species, where *T. macrops s.s., T. cardamomensis,* and *T. rubeus*
434 were hypothesized to be three distinct species, was the most favored model.
435
436 **3.3 Population genetics methods of AFLP analysis**
437 ***3.3.1 Genetic differentiation and historical gene flow***
438 F_{st} was 0.5 on average in both POPGENE and HICKORY. The % variation within and
439 among putative species from AMOVA was 46% and 54% respectively. Overall Φ_{PT} was 0.538
440 and Φ_{PT} between populations are given in Table 1. The % polymorphic loci for *T. macrops*
441 *s.s., T. cardamomensis and T. rubeus* were 54%, 21% and 25% respectively. A plot of $\log_{10}$
442 (BF) against F_{st} in BayeScan showed no outliers indicating neutrality of loci (Fig. A.3).
443 Estimated historical gene flow, N_m, among populations was 0.14.

444

445 ***3.3.2 Non-spatial and spatial Bayesian MCMC cluster analysis***

446 In STRUCTURE, Ln Pr(X|K) increased by large increments up to K=4, while the
447 increase was small at K=5 and decreased after K=6 (Fig. A.4a). With Evanno's methods, ΔK
448 clearly peaked at three populations (Fig. A.4b). A graph estimating population structure for
449 K=4 (including *T. venustus*) is given in Fig. 3. The probability of each individual belonging to
450 assigned species is given in Table A.3. The probabilities of individual assignments when K=3
451 (Evanno's method) were nearly all equal to 1.0.

452 In GENELAND, three clusters could be visualized from the posterior distribution.
453 Maps of individual posterior probabilities of membership to each cluster are given in Fig. 4.
454 All individuals from the *T. cardamomensis* cluster were assigned to *T. macrops s.s.* Posterior
455 probabilities of cluster membership for individual samples are given in Table A.4. The sample
456 switching experiment showed that probability of membership of sample B44 to *T. macrops*
457 *s.s.* increased to p=0.693 and to *T. venustus* decreased to p=0.158. Probability of membership
458 of sample A144 to *T. macrops s.s.* decreased to p=0.550 (from p=0.664) and to *T. venustus*
459 increased to p=0.304 (from p=0.158).

460

461 **3.4 Multivariate statistical analysis of AFLPs**

462 PCoA showed that ~50% of total variation was explained by the first three axes.
463 These accounted for 24%, 18% and 7% of observed variation. A scatterplot showed the three
464 proposed species formerly within *T. macrops* to be well separated on axes 1 and 3 (Fig. 5a).

465

466 The DAPC scatterplot clearly demarcated the three clusters corresponding to the
467 putative species, with *T. venustus* as the fourth cluster (Fig. 5b). Assignment of individuals to
468 each cluster in the former *T. macrops* (Fig. A.5a) agreed completely with geographic

469 distributions and mtDNA clades. The *a.score* for each cluster was: *T. macrops s.s.*=0.15, *T.*
470 *cardamomensis*=0.96, *T. rubeus*=0.97, and *T. venustus*=0.68.

471 In *prabclus* four species clusters (including *T. venustus*) were detected and clearly
472 separated in dimensions 1 and 3 with 100% accuracy of individual assignment to respective
473 clusters (Fig. 5c). No datapoints were classified as noise components in this analysis (Fraley
474 and Raftery 1998, 2002, Hausdorf and Hennig 2010).

475 Euclidean distance based hierarchical clustering and seriation also showed three
476 clusters within *T. macrops s.l.* A partial map of AFLP banding patterns and the clustering
477 results are presented in the graphical abstract, and full results are provided in Fig. A.6. *T.*
478 *macrops s.s.* and *T. cardamomensis* samples differentiated into two distinct clusters within a
479 single large cluster, except for ambiguous placement of A144. The position of *T. venustus,*
480 nested within *T. macrops s.l.,* was congruent with phylogenetic analyses.

481

482 4 Discussion

483 4.1 Overview

484 In the majority of cryptic speciation studies, morphological conservativeness, parallel
485 and convergent evolution of phenotypic traits, and/or mimicry (Sanders et al. 2006), driven
486 by natural selection, sexual selection, and ecological adaptation, confound species
487 delimitation. Southeast Asian green or "bamboo" pitvipers are typical in this sense due to
488 their general morphological conservativeness or environmentally driven morphological
489 convergence (Sanders et al. 2004). This is further complicated by sampling difficulties, due to
490 cryptic life-styles of snakes, which can be a major drawback in population genetics analyses.
491 In such cases, multiple markers and analysis methods may provide critical information to
492 derive robust inferences on species diversity (Dupuis et al. 2012, Carstens 2013).
493 Traditionally used in population genetics, AFLPs have recently found application in species

494 delimitation of a variety of plants (e.g. Prebble et al. 2012, Medrano et al. 2014) and animals
495 (e.g. Nie et al. 2012, Arthofer et al. 2013). Surprisingly, very few studies compare results
496 from multiple analysis methods (eg: Meudt et al. 2009, Reeves and Richards 2011). Most
497 comparative studies have only assessed genetic similarity co-efficients and multivariate
498 clustering methods such as UPGMA and NJ (Meyer et al. 2004, Kosman and Leonard 2005,
499 Dalirsefat et al. 2009).

500 In our investigation of cryptic speciation in *T. macrops,* we used three mtDNA
501 markers and 298 polymorphic AFLP loci and eight analysis methods. Of these, results from
502 two Bayesian phylogenetic analyses (total evidence and mtDNA), Bayes Factor Delimitation,
503 four multivariate statistical methods of AFLPs (PCoA, DAPC, Gaussian clustering, and
504 hierarchical clustering), and to a certain degree one population genetic method
505 (STRUCTURE) were congruent with each other. Moreover, these genetic results correspond
506 to allopatric geographic ranges and individual morphological characters described for each
507 species (Malhotra et al. 2011a). This gives us high confidence in the genetic distinctiveness of
508 *T. macrops s.s., T. cardamomensis,* and *T. rubeus* (Malhotra et al. 2011a).

509 In contrast, GENELAND was the only method that gave a contradictory result. This
510 brings into question the appropriateness of using GENELAND for a system such as this, and
511 we discuss this further in subsequent sections. The AFLP-only phylogenetic analysis, with
512 low support values, was unreliable. It is likely that low confidence can be placed in these two
513 results. We discuss our results and evaluate inferences in the context of dominant marker
514 analysis methods and provide a comparative assessment of phylogenetic, Bayes Factor
515 Delimitation, spatial and non-spatial Bayesian clustering methods, DAPC, Gaussian
516 clustering, and finally hierarchical clustering.

517

518 **4.2 Comparative assessment of results from AFLP analysis**

4.2.1 Phylogenetic analyses

The resolution in total evidence Bayesian phylogenetic reconstruction matched that of traditionally used mtDNA markers, producing a robust phylogeny with three distinct clades in *T. macrops* (Fig. 2a & b). The AFLP tree, however, showed poor resolution of species and low support values at deeper nodes (Fig. A.2). This difference in result indicates a stronger phylogenetic signal from the mtDNA sequence data. AFLP-based phylogenetic analyses have been largely confined to distance-based methods such as Neighbor-Joining and UPGMA in the past. However, despite homology and non-independence of fragments, AFLP phylogenies can successfully delimit species, sometimes performing better than STRUCTURE (Meudt et al. 2009) and even mtDNA markers (Mendelson and Simons 2006, Kingston et al. 2009). Combined AFLP and mtDNA datasets can yield robust phylogenies and provide evidence for interspecific hybridization (Després et al. 2003, Pelser et al. 2003, Koopman 2005, Meudt and Clarke 2007, Kingston et al. 2009). There have been conflicting reports on the utility of AFLPs at deeper phylogenetic or interspecific levels as it appears to be affected by a drastic increase in non-homologous shared fragments resulting in loss of phylogenetic signal (Althoff et al. 2007, Dasmahapatra et al. 2009, Graves 2009, Kingston et al. 2009, García-Pereira et al. 2014). Moreover, choice of bands and tree-building methods, and application of restriction sites models, could over-simplify complex evolutionary processes, thus affecting resolution of deeper nodes (Dasmahapatra et al. 2009, Graves 2009). Finally, the more complex model developed specifically for AFLPs (Luo et al. 2007) has been found to be extremely computationally burdensome, making it impractical to implement (Koopman et al. 2008, Dasmahapatra et al. 2009). The failure of our AFLP reconstruction under a restriction site model reinforces the need for better, practical, and exclusive phylogenetic methods (Graves 2009).

544 ***4.2.2 Bayes Factor Delimitation***

545 SNAPP-BFD offers the advantage of testing multiple species hypothesis by
546 integrating topologies during marginal likelihood estimation and thereby avoiding the need to
547 predefine the species tree and biasing support (Leaché et al. 2014). From our SNAPP-BFD
548 analysis, Bayes Factors ranked the four species model (three from *T. macrops* complex and *T.*
549 *venustus*) as the most highly favored model (Marginal Likelihood=-3757) (Table 2). Model E,
550 where we lumped *T. macrops s.s., T. cardamomensis and T. rubeus* into a single species,
551 received the lowest support (Marginal Likelihood = -4587; 2x $\log_e$ Bayes Factor = 1660) and
552 was ranked last among the five models tested. Therefore, the hypotheses that consider *T.*
553 *macrops s.l.* as a single species, or as consisting of two species, are not supported by SNAPP-
554 BFD analysis.

555

556 ***4.2.3 Non-spatial and spatial Bayesian MCMC clustering analyses***

557 Both STRUCTURE and GENELAND use Bayesian MCMC methods to assign
558 individuals probabilistically to populations based on allele frequencies. They cluster groups
559 of individuals into populations by assuming that they are in HWE and linkage equilibrium. In
560 the estimation of number of clusters (K), both STRUCTURE and GENELAND failed to
561 differentiate *T. cardamomensis* as a separate species cluster and returned K = 3 by grouping
562 these specimens in a single cluster with *T. macrops s.s.* (Fig. 3 & 4).

563 In STRUCTURE, the Pritchard et. al. (2000) method was inconclusive as to whether
564 K equalled 3, 4, or even 5, as results depended on what cut off was applied to the Ln Pr(X|K)
565 increase (Fig A.4a). In contrast, Evanno's method gave us a clear result of K = 3 (Fig. A.4b).
566 Applying Evanno's correction is the norm for K estimation, since it is more formal and is
567 endorsed by Pritchard et al. (2007). However, the Pritchard et. al. (2000) method of K
568 estimation, which is said to be unreliable, subjective, and sometimes biologically

569 meaningless, proved more realistic when the Ln Pr(X|K) values were compared. The Ln
570 Pr(X|K) increase from K = 3 to 4 was 8.57%, approximately 50% of increase from both K = 1
571 to 2 and K = 2 to 3, whereas it dropped to 2.1% for K = 4 to 5 (Table A.5). Moreover, there
572 are several cases of K underestimation, and STRUCTURE results are said to be conservative
573 when Evanno's method is applied (e.g., Frantz et al. 2009, Blanquer and Uriz 2010).
574 Therefore, based on our phylogenetic results and taking into consideration that K = 4 proved
575 a better estimation in the Pritchard et al. (2000) method of K estimation, we calculated the
576 individual assignment probabilities by assigning individuals in *T. macrops s.l.* to three
577 clusters and *T. venustus* individuals to a fourth cluster. The mean permuted assignment
578 probability value *(p)* for *T. cardamomensis* individuals to a separate cluster (as per CLUMPP)
579 was only slightly lower at p = 0.969 compared to *T. macrops s.s* (p = 1.0), *T. rubeus* (p =
580 0.999), and *T. venustus* (p = 0.999). This further increased our confidence that Evanno's
581 method underestimated K by grouping *T. cardamomensis* with *T. macrops s.s.*
582 In GENELAND, however, all individuals of *T. cardamomensis* as well as *T. macrops*
583 *s.s.* were assigned to *T. macrops s.s.* with a probability of assignment p = 0.644. Among the
584 *T. macrops s.s.* samples, specimen B44, reportedly from Nakhon Si Thammarat in southern
585 Thailand, was an exception with a lower probability of assignment to *T. macrops s.s* (p =
586 0.441). Moreover, while the probability of assignment of all other *T. macrops s.s.* specimens
587 to *T. venustus* was p = 0.158, that of sample B44 was higher at p = 0.363. In contrast, the
588 STRUCTURE analysis assigned B44 to *T. macrops s.s.* with a maximum mean permuted
589 probability of 1.0 by CLUMPP. This specimen was the only *T. macrops s.s.* specimen
590 reported from southern Thailand and as it was obtained from a dealer, the presence of this
591 species at the reported locality is unconfirmed. Our suspicion, that this difference in
592 assignment between GENELAND and STRUCTURE was due to the influence of spatial

593 priors on individual assignment probabilities in GENELAND, was supported by a sample
594 switching analysis.

595 The apparent interruption of gene flow represented by the Gulf of Thailand, which
596 currently separates populations in the south of Thailand from populations in Cambodia and
597 Vietnam, might be considered to invalidate the GENELAND results. However, it is important
598 to consider the context at the time these species were evolving. The Gulf of Thailand has a
599 maximum depth of 80m, and Voris (2000) showed that for c. 35% of the last 170,000 years, it
600 would have largely been dry land. While this precision is only possible for relatively recent
601 timescales, it is also known that there have been sea level fluctuations of similar magnitudes
602 during the last 30 my (Hall and Holloway, 1998).

603 Allele-frequency estimations from small populations could be compensated by
604 analyzing large number of loci (Krauss 2000). Our AFLP dataset was large (298 loci), yet the
605 genetic diversity indices show a clear bias caused by small sample sizes of *T. cardamomensis*
606 and *T. rubeus,* since the % polymorphic loci decreased considerably relative to population
607 size. F_{st} values were much lower for *T. cardamomensis* and *T. rubeus* (0.0507 and 0.00729,
608 $p<0.00001$) as opposed to *T. macrops s.s.* and *T. venustus* (0.9, $p<0.00001$). However, Φ_{PT}
609 values showed *T. cardamomensis* to be less genetically distinct from *T. macrops s.s.* than *T.*
610 *rubeus* (Table 1). These estimates appear to be non-representative, with biases arising from
611 sampling deficiency, probably insufficient number of loci, as well as dominant nature of the
612 marker. Hence deriving any strong biological inferences from diversity indices would be
613 highly dubious. Our estimation of historical gene flow among *T. macrops* complex
614 populations was low (N_m=0.14), although N_m calculation is based on F_{st}. Moreover,
615 considering the wider geographic distribution of *T. macrops* complex and geographical
616 barriers in southeast Asia, it is likely that there is established allopatry among populations.
617 Therefore, it seems highly likely that implementing GENELAND (which is more appropriate

618 for systems with contemporary gene flow) may not have been ideal for the *T. macrops*
619 complex. It is difficult to predict whether increasing sample size may have provided better
620 resolution for *T. cardamomensis* in STRUCTURE and GENELAND. *Trimeresurus rubeus*
621 (represented by only seven individuals), was still sufficiently genetically diversified to form a
622 separate cluster as well as achieve 100% individual assignment success in all analysis
623 methods.
624
625 ***4.2.4 Multivariate statistical methods***
626 Multivariate techniques proved superior to Bayesian MCMC clustering in terms of
627 sensitivity and confidence as PCoA, DAPC, Gaussian clustering, and hierarchical clustering
628 all split *T. macrops s.l.* into three clusters and also assigned individuals to their corresponding
629 mtDNA clades with 100% success (Fig. 5a, b, & c and Fig. A.6).
630 In DAPC analysis, *optim.a.score* recognized that the first three PCs would give
631 highest *a.scores*. Although they represented only 45% of total variance, we used the first
632 three PCs to obtain a strong and stable DAPC solution (Thibaut Jombart, personal
633 communication). *a.scores* were very high for *T. cardamomensis* and *T. rubeus* (0.96 and
634 0.97), and this increases our confidence that these two groups are genetically distinct. *T.*
635 *macrops s.s.* had an *a. score* of 0.15, but given that some issues still exist with *optim.a.score*
636 and *a.score* functions (e.g., with repeatability), a more critical review of these scores is not
637 justified at this stage. Gaussian clustering in *prabclus* was highly successful, defining three
638 clusters of *T. macrops* and assigning individuals accurately to them. It is important to note
639 that both DAPC and Gaussian clustering were not affected by any sampling deficiency or by
640 lower level of genetic diversity of *T. cardamomensis*, which was always identified as a
641 separate cluster, in contrast to Bayesian clustering.

642 Finally, the Euclidean distance-based hierarchical clustering and seriation also showed
643 distinct clustering of four species with *T. venustus* nested in *T. macrops s.l.*. *T. macrops s.s.*
644 and *T. cardamomensis* each formed a distinct sub-cluster within a single large cluster (Fig.
645 A.6). Locus-based (column-wise) clustering identified regions of banding dissimilarity across
646 the three species in *T. macrops s.l.* (Graphical Abstract). Sample A144, which clustered with
647 *T. cardamomensis* is the only representative from North Thailand (Jae Sorn NP, Lampang
648 province) and likely a genetic outlier. It is interesting that this was only apparent in the
649 hierarchical clustering analysis.
650

651 **5 Conclusions**

652 Our study provides the genetic evidence required to complete investigations into the
653 morphologically cryptic species in the *Trimeresurus macrops* complex. Congruence of results
654 between multiple markers and methodologies clearly demarcates the three proposed species
655 as genetically and reproductively isolated. We therefore confirm *Trimeresurus macrops s.s.,*
656 *T. cardamomensis, and T. rubeus* as full species. Further, our study reinforces the importance
657 of using appropriate and multiple analysis methods and performing a comparative assessment
658 before deriving inferences on species diversity. The study also demonstrates the continued
659 utility of AFLPs for cryptic species delimitation and discovery, when high cost and
660 sequencing noise can be deterrents for using high-throughput sequencing (e.g. RAD-Seq)
661 (Davey et al. 2013).
662

663 **ACKNOWLEDGEMENTS**

664 AM and RST are grateful to the large numbers of people who have assisted in the field,
665 including T. Chan-ard, J. Nabhitabhata (National Science Museum of Thailand), L.
666 Chanhome (Queen Savoabha Memorial Institute, Thailand), K. Thirakhupt and P-P. van Dijk
667 (Chulalongkorn University, Thailand), C. Rangsiyanon (Chiang Mai University), S. Paglia,
668 M. Cox, J. Murray, and T-X. Kiem (Cho–Ray Hospital, Viet Nam). We gratefully
669 acknowledge the National Science Council of Thailand and the Ministry of Health, Viet Nam
670 for permission to carry out fieldwork. This study was funded (AM/RST) by the Leverhulme
671 Trust (F174/I), the Wellcome Trust (057257/Z/99/Z and 060384/Z/00/Z), the Darwin
672 Initiative (162/6/65), a Sir Kenneth Blaxter Scholarship (British Society of Animal Science,
673 2007) to Mrinalini, the Royal Society, the Percy Sladen Trust, the Bonhote Trust, and the
674 Carnegie Trust. We also thank K. Udomritthiruj and G. Vogel for donation of specimens. The
675 opportunity for BLS to work in Laos was made possible by the Wildlife Conservation Society
676 / Division of Forest Resource Conservation Cooperative Program, and in Cambodia by the
677 Wildlife Conservation Society / Ministry of Agriculture, Forestry and Fisheries / Ministry of
678 Environment Cooperative Program. The Ministry of Agriculture and Forestry (Vientiane,
679 Laos) and Ministry of Agriculture, Forestry and Fisheries (Phnom Penh, Cambodia)
680 permitted export of specimens to the Field Museum. BLS was supported by the John D. and
681 Catherine T. MacArthur Foundation, the National Geographic Society (grant No. 6247–98),
682 the Wildlife Conservation Society, and the U.S. National Science Foundation (DEB-
683 1145922). H. Heatwole, David Emmett, T. Neang, S. Platt, K. Sok, and B. Thaovanseng are
684 thanked for fieldwork assistance, and M. Hedemark, A. Johnson, T. Hansel, J. Walston, and
685 C. Poole are thanked for logistical support.

686

687

688 **References**

689 Albertson, R.C., Markert, J.A., Danley, P.D., Kocher, T.D. 1999. Phylogeny of a rapidly
690 evolving clade: The cichlid fishes of Lake Malawi, East Africa. Proc. Natl. Acad. Sci. U.S.A.
691 96:5107–5110.
692
693 Althoff, D.M., Gitzendanner, M.A., Segraves, K.A. 2007. The utility of amplified fragment
694 length polymorphisms in phylogenetics: a comparison of homology within and between
695 genomes. Syst. Biol. 56:477–484.
696
697 Arthofer, W., Rauch, H., Thaler-Knoflach, B., Moder, K., Muster, C., Schlick-Steiner, B.C.,
698 Steiner, F.M. 2013. How diverse is *Mitopus morio*? Integrative taxonomy detects cryptic
699 species in a small-scale sample of a widespread harvestman. Mol. Ecol. 22:3850–3863.
700
701 Bardeleben, C., Moore, R.L., Wayne, R.K. 2005. A molecular phylogeny of the Canidae
702 based on six nuclear loci. Mol. Phylogenet. Evol. 37:815–831.
703
704 Blanquer, A., Uriz, M.J. 2010. Population genetics at three spatial scales of a rare sponge
705 living in fragmented habitats. BMC Evol. Biol. 10:13.
706
707 Bonin, A., Ehrich, D., Manel, S. 2007. Statistical analysis of amplified fragment length
708 polymorphism data: a toolbox for molecular ecologists and evolutionists. Mol. Ecol.
709 16:3737–3758.
710
711 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, MA.,
712 Rambaut, A., & Drummond, A. J. 2014. BEAST 2: A Software Platform for Bayesian

713 Evolutionary Analysis. PLoS Comp. Biol. 10(4): e1003537. doi:
714 10.1371/journal.pcbi.1003537.
715
716 Brito, P.H., Edwards, S.V. 2009. Multilocus phylogeography and phylogenetics using
717 sequence-based markers. Genetica 135:439–455.
718
719 Brown, D.M., Brenneman, R.A., Koepfli, K-P., Pollinger, J.P., Milá, B., Georgiadis, N.J.,
720 Louis Jr, E.E., Grether, G.F., Jacobs, D.K., Wayne, R.K. 2007. Extensive population genetic
721 structure in the giraffe. BMC Biol. 5:57.
722
723 Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and RoyChoudhury, A. 2012.
724 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full
725 coalescent analysis. Mol. Biol. Evol. 29:1917–1932.
726
727 Caraux G., Pinloche S. 2005. A software for analysing and visualising data. Bioinformatics.
728 21:1280-1281.
729
730 Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D. 2013. How to fail at species
731 delimitation. Mol. Ecol. 22:4369–4383.
732
733 Chen, C., Durand, E., Forbes, F., François, O. 2007. Bayesian clustering algorithms
734 ascertaining spatial population structure: a new computer program and a comparison study.
735 Mol. Ecol. Notes. 7:747–756.
736
737 Cope, E.D. 1860. Catalogue of the venomous serpents in the museum of the Academy of

738 Natural Sciences of Philadelphia, with notes on the families, genera, and species. Proc. Acad.
739 Natl. Sci. Phil. 11:332–347.
740
741 Crow, J.F., Aoki, K. 1984. Group selection for a polygenic behavioral trait: estimating the
742 degree of populations subdivision. Proc. Natl. Acad. Sci. U.S.A. 81:6073–6077.
743
744 Dalirsefat, S., Meyer, A., Mirhoseini, S. 2009. Comparison of similarity coefficients used for
745 cluster analysis with amplified fragment length polymorphism markers in the silkworm,
746 *Bombyx mori*. J. Insect Sci. 9:71.
747
748 Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new
749 heuristics and parallel computing. Nature Methods 9(8):772.
750
751 Dasmahapatra, K.K., Hoffman, J.I., Amos, W. 2009. Pinniped phylogenetic relationships
752 inferred using AFLP markers. Heredity 103:168–177.
753
754 Davey, J.W., Cezard, T., Fuentes-Urtilla, P., Eland, C., Gharbi, K., Blaxter, M.L. 2013.
755 Special features of RAD Sequencing data: implications for genotyping. Mol. Ecol. 22:3151-
756 3164.
757
758 David, P., Vogel, G., Dubois, A. 2011. On the need to follow rigorously the Rules of the
759 *Code* for the subsequent designation of a nucleospecies (type species) for a nominal genus
760 which lacked one: the case of the nominal genus *Trimeresurus,* Lacépède, 1804 (Reptilia:
761 Squamata: Viperidae). Zootaxa 2992:1-51.
762

763 Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J.,
764 Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., von Haeseler, A.,
765 Ebersberger, I., Pass, G., Misof, B. 2014. Decisive data sets in phylogenomics: Lessons from
766 studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evol.
767 31:239-249.
768
769 Després, L., Gielly, L., Redoutet, B., Tabarlet, P. 2003. Using AFLP to resolve phylogenetic
770 relationships in a morphologically diversified plant species complex when nuclear and
771 chloroplast sequences fail to reveal variability. Mol. Phylogenet. Evol. 27:185–196.
772
773 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with
774 BEAUti and the BEAST 1.7 Mol. Biol. Evol. 29: 1969–1973.
775 Dupuis, J.R., Roe, A.D., Sperling, F.A. 2012. Multi-locus species delimitation in closely
776 related animals and fungi: one marker is not enough. Mol. Ecol. 21:4422–36.
777
778 Earl, D.A., vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for
779 visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet.
780 Resour. 4:359-361.
781
782 Egger, B., Koblmüller, S., Sturmbauer, C., Sefc, K.M. 2007. Nuclear and mitochondrial data
783 reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. BMC
784 Evol. Biol. 7:137.
785
786 Evanno, G., Regnaut, S., Goudet, J. 2005. Detecting the number of clusters of individuals
787 using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

788

789 Excoffier, L., Heckel, G. 2006. Computer programs for population genetics data analysis: a
790 survival guide. Nat. Rev. Genet. 7:745–758.
791
792 Falush, D., Stephens, M., Pritchard, J.K. 2003. Inference of population structure using
793 multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–
794 1587.
795
796 Falush, D., Stephens, M., Pritchard, J.K. 2007. Inference of population structure using
797 multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes. 7:574–578.
798
799 Faria, R., Renaut, S., Galindo, J., Pinho, C., Melo-Ferreira, J., Melo, M., Jones, F.,
800 Salzburger, W., Schluter, D., Butlin, R. 2014. Advances in Ecological Speciation: an
801 integrative approach. Mol. Ecol. 23: 513–521.
802
803 Foll, M., Baumont, M.A., Gaggiotti, O.E. 2008. An approximate Bayesian computation
804 approach to overcome biases that arise when using AFLP markers to study population
805 structure. Genetics 179:927–939.
806
807 Foll, M., Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for
808 both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–995.
809
810 Fox, J. 2005. The R Commander: A basic-statistics graphical user interface to R. J. Stat. Soft.
811 14:1–42.
812

813 Fraley, C., Raftery, A.E. 1998. How many clusters? Which clustering methods? Answers via
814 model-based cluster analysis. Comput. J. 41:578-588.
815
816 Fraley, C., Raftery, A.E. 2002. Model-based clustering, discriminant analysis, and density
817 estimation. J. Am. Stat. Assoc. 97:611–631.
818
819 François, O., Ancelet, S., Guillot, G. 2006. Bayesian clustering using hidden Markov random
820 fields in spatial population genetics. Genetics 174:805–816.
821
822 Frantz, A.C., Cellina, S., Krier, A., Schley, L., Burke, T. 2009. Using spatial Bayesian
823 methods to determine the genetic structure of a continuously distributed population: clusters
824 or isolation by distance? J. Appl. Ecol. 46:493–505.
825
826 García-Pereira, M.J., Caballero, A., Quesada, H. 2010. Evaluating the relationship between
827 evolutionary divergence and phylogenetic accuracy in AFLP data sets. Mol. Biol. Evol.
828 27:988–1000.
829
830 García-Pereira, M.J., Carvajal-Rodríguez, A., Whelan, S., Caballero, A., Quesada, H. 2014.
831 Impact of deep coalescence and recombination on the estimation of phylogenetic
832 relationships among species using AFLP markers. Mol. Phylogenet. Evol. 76:102–109.
833
834 Gatesy, J. and Springer, M.S. 2013. Concatenation versus coalescence versus
835 “concatalescence”. Proc. Natl. Acad. Sci. U.S.A. 110(10). doi: 10.1073/pnas.1221121110
836
837 Gower, J.C. 1966. Some distance properties of latent root and vector methods used in

838 multivariate analysis. Biometrika 53:325–328.

839

840 Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics

841 27:857–874.

842

843 Gower, J.C. 2005. Principal coordinates analysis. Encycl. Biostat. 6.

844

845 Graves, J. 2009. Deeper AFLPs. Heredity 103:99.

846

847 Grummer, J. A., Bryson, R.W., Reeder, T.W. 2013. Species delimitation using Bayes factors:

848 simulations and application to the *Sceloporus scalaris* species group (Squamata:

849 Phrynosomatidae). Syst. Biol.63(2):119–133.

850

851 Guillot, G., Estoup, A., Mortier, F., Cosson, J.F. 2005a. A spatial statistical model for

852 landscape genetics. Genetics 170:1261–1280.

853

854 Guillot, G., Mortier, F., Estoup, A. 2005b. Geneland: A computer package for landscape

855 genetics. Mol. Ecol. Notes. 5:708–711.

856

857 Guillot, G. 2008. Inference of structure in subdivided populations at low levels of genetic

858 differentiation.The correlated allele frequencies model revisited. Bionformatics 24:2222–

859 2228.

860

861 Guillot, G., Santos, F., Estoup, A. 2008. Analysing georeferenced population genetics data

862 with Geneland: a new algorithm to deal with null alleles and a friendly graphical user

863 interface. Bioinformatics 24:1406–1407.

864

865 Guillot, G., Santos, F. 2010. Using AFLP markers and the Geneland program for the

866 inference of population genetic structure. Mol. Ecol. Res. 10:1082–1084.

867

868 Guindon, S. and Gascuel, O. 2003. A simple, fast and accurate method to estimate large

869 phylogenies by maximum-likelihood". Syst Biol. 52:696–704.

870

871 Hausdorf, B., Hennig, C. 2010. Species delimitation using dominant and codominant

872 multilocus markers. Syst. Biol. 59:491–503.

873

874 Hall, R., Holloway, J. D. (1998): Biogeography and Geological Evolution of SE Asia. —

875 Leiden (Backhuys Publishers). — 417 S. ISBN 90-73348-97-8. Zool. Reihe, 76: 158.

876 doi: 10.1002/mmnz.20000760119.

877

878 Hollingsworth, P.M., Ennos, R.A. 2004. Neighbour-joining trees, dominant markers and

879 population genetic structure. Heredity 92:490–498.

880

881 Holsinger, K.E., Lewis, P.O. 2007. HICKORY: a package for analysis of population genetic

882 data, version 1.0. Storrs : Department of Ecology and Evolutionary Biology, University of

883 Connecticut.

884

885 Jaccard, P. 1908. Nouvelle recherches sur la distribution florale. Bull. Soc. Vaudoise. Sci.

886 Nat. 44:223–270.

887

888 Jakobsson, M., Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation
889 program for dealing with label switching and multimodality in analysis of population
890 structure. Bioinformatics 23:1801–1806.
891
892 Jombart, T., Devillard, S., Balloux, F. 2010. Discriminant analysis of principal components: a
893 new method for the analysis of genetically structured populations. BMC Genet. 11:94.
894
895 Kass, R. E. and A. E. Raftery. 1995. Bayes factors. J. Amer. Stat. Assoc. 90:773–795.
896
897 Kingston, S.E., Adams, L.D., Rosel, P.E. 2009. Testing mitochondrial sequences and
898 anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group:
899 molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). BMC Evol.
900 Biol. 9:245.
901
902 Koopman, W.J.M. 2005. Phylogenetic signal in AFLP data sets. Syst. Biol. 54:197–217.
903
904 Koopman, W.J.M., Wissemann, V., De Cock, K., Van Huylenbroeck, J., De Riek, J.,
905 Sabatino, G.J.H., Visser, D., Vosman, B., Ritz, C.M., Maes, B., Werlemark, G., Nybom, H.,
906 Debener, T., Linde, M., Smulders, M.J.M. 2008. AFLP markers as a tool to reconstruct
907 complex relationships: a case study in *Rosa* (Rosaceae). Am. J. Bot. 95:353–366.
908
909 Kosman, E., Leonard, K.J. 2005. Similarity coefficients for molecular markers in studies of
910 genetic relationships between individuals for haploid, diploid, and polyploid species. Mol.
911 Ecol. 14: 415–424.
912

913 Krauss, S.L. 2000. Accurate genetic diversity estimates from amplified fragment length
914 polymorphism. Mol. Ecol. 9:1241–1245.
915
916 Kronforst, M.R., Gilbert, L.E. 2008. The population genetics of mimetic diversity in
917 *Heliconius* butterflies. Proc. R. Soc. B. 275:493–500.
918
919 Lacépède, B.G.E.L. (Comte de). 1804. Mémoire sur plusieurs animaux de la Nouvelle
920 Hollande dont la description n'a pas encore été publiée. Ann. Mus. Hist. Nat. Paris. 4:55–58.
921
922 Latch, E.K., Dharmarajan, G., Glaubitz, J.C., Rhodes Jr, O.E. 2006. Relative performance of
923 Bayesian clustering software for inferring population substructure and individual assignment
924 at low levels of population differentiation. Conserv. Genet. 7:295–302.
925
926 Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R. 2014. Species delimitation using
927 genome-wide SNP data. Syst. Biol. 63: 534-542.
928
929 Luo, R., Hipp, A.L., Larget, B. 2007. A bayesian model of AFLP marker evolution and
930 phylogenetic inference. Stat. Appl. Genet. Mol. Biol. 6:11.
931
932 Malhotra, A., Thorpe, R.S. 2000. A phylogeny of the *Trimeresurus* group of pit vipers: new
933 evidence from a mitochondrial gene tree. Mol. Phylogenet. Evol.16:199–211.
934
935 Malhotra, A., Thorpe, R.S. 2004a. Maximizing information in systematic revisions: a
936 combined molecular and morphological analysis of a cryptic green pitviper complex,
937 *Trimeresurus stejnegeri*. Biol. J. Linn. Soc. 82:219–235.

938

939 Malhotra, A., Thorpe, R.S. 2004b. A phylogeny of four mitochondrial gene regions suggests
940 a revised taxonomy for Asian pitvipers (*Trimeresurus* and *Ovophis*). Mol. Phylogenet. Evol.
941 32:83–100.

942

943 Malhotra, A., Thorpe, R.S., Mrinalini, Stuart, B.L. 2011a. Two new species of pitviper of the
944 genus *Cryptelytrops* Cope 1860 (Squamata: Viperidae: Crotalinae) from Southeast Asia.
945 Zootaxa 2757:1–23.

946

947 Malhotra, A., Dawson, K., Guo, P., Thorpe, R.S. 2011b. Phylogenetic structure and species
948 boundaries in the mountain pitviper *Ovophis monticola* (Serpentes: Viperidae: Crotalinae) in
949 Asia. Mol. Phylogenet. Evol. 59:444-457.

950

951 McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield,
952 R.T. 2013. A Phylogeny of birds based on over 1,500 loci collected by target enrichment and
953 high-throughput sequencing. PLoS One 8(1): e54848.

954

955 Medrano, M., Lopez-Perea, E., Herrera, C.M. 2014. Population genetics methods applied to a
956 species delimitation problem: Endemic trumpet daffodils (*Narcissus* Section *Pseudonarcissi*)
957 from the southern Iberian Peninsula. Int. J. Plant Sci. 175:501–517.

958

959 Mendelson, T.C., Simons, J.N. 2006. AFLPs resolve cytonuclear discordance and increase
960 resolution among barcheek darters (Percidae: *Etheostoma*: *Catonotus*). Mol. Phylogenet.
961 Evol. 41:445–453.

962

963 Meudt, H.M., Clarke, A.C. 2007. Almost forgotten or latest practice? AFLP applications,
964 analyses and advances. Trends Plants Sci. 12:106–117.
965
966 Meudt, H.M., Lockhart, P.J., Bryant, D. 2009. Species delimitation and phylogeny of a New
967 Zealand plant species radiation. BMC Evol. Biol. 9:111.
968
969 Meyer, A.S., Garcia, A.A.F., de Souza, A.P., de Souza Jr, C.L. 2004. Comparison of
970 similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays L*).
971 Genet. Mol. Biol. 27:83–91.
972
973 Meyer, C.P., Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling.
974 PLoS Biol. 3:e422.
975
976 Milá, B., Carranza, S., Guillaume, O., Clobert, J. 2010. Marked genetic structuring and
977 extreme dispersal limitation in the Pyrenean brook newt *Calotriton asper* (Amphibia:
978 Salamandridae) revealed by genome-wide AFLP but not mtDNA. Mol. Ecol. 19:108–120.
979
980 Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., *et*
981 *al.*: Phylogenomics resolves the timing and pattern of insect evolution. 2014. Science.
982 346:763-768.
983
984 Morin, P.A., Archer. F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J.,
985 Parsons, K., Pitman, R., Li, L., Bouffard, P., Abel Nielsen, S.C., Rasmussen, M.,
986 Willerslevm, E., Gilbert, M.T., Harkins, T. 2010. Complete mitochondrial genome
987 phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome

988 Res. 20:908–916.

989

990 Nie, R-E., Mochizuki, A., Brooks, S.J., Liu, Z-Q, Yang, X-K. 2012. Phylogeny of the green
991 lacewing *Chrysoperia nipponensis* species-complex (Neuroptera: Chrysopidae) in China,
992 based on mitochondrial and AFLP data. Ins. Sci. 19:633–642.

993

994 Ogden, R., Thorpe, R.S. 2002. The usefulness of amplified fragment length polymorphism
995 markers for taxon discrimination across graduated fine evolutionary levels in Caribbean
996 Anolis lizards. Mol. Ecol.11:437–445.

997

998 Peakall, R., Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic
999 software for teaching and research. Mol. Ecol. Notes 6:288–295.

1000

1001 Pelser, P.B., Gravendeel, B., Van der Meijden, R. 2003. Phylogeny reconstruction in the gap
1002 between too little and too much divergence: the closest relatives of *Senecio jacobaea*
1003 (Asteraceae) according to DNA sequences and AFLPs. Mol. Phylogenet. Evol. 29:613–628.

1004

1005 Pérez-Figueroa, A., Garcia-Pereira, M.J., Saura, M., Rolan-Alvarez, E., Caballer, A. 2010.
1006 Comparing three different methods to detect selective loci using dominant markers. J. Evol.
1007 Biol. 23:2267–2276.

1008

1009 Posada, D. 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25:1253–
1010 1256.

1011

1012 Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution.

1013 Bioinformatics 14:817–818.

1014

1015 Prebble, J.M., Meudt, H.M., Garnock-Jones, P.J. 2012. Phylogenetic relationships and

1016 species delimitation of New Zealand bluebells (*Wahlenbergia,* Campanulaceae) based on

1017 analyses of AFLP data. New Zeal. J. Botany 50:365–378.

1018

1019 Pritchard, J.K., Stephens, M., Donnelly, P. 2000. Inference of population structure using

1020 multilocus genotype data. Genetics 155:945–959.

1021

1022 Pritchard, J.K., Wen, X., Falush, D. 2007. Documentation for STRUCTURE software:

1023 Version 2.2. http://pritch.bsd.uchicago.edu/software/structure22/readme.pdf.

1024

1025 Quek, S-P., Counterman, B.A., de Moura, P.A., Cardoso, M.Z., Marshall, C.R., McMillan,

1026 W.O., Kronforst, M.R. 2010. Dissecting comimetic radiations in *Heliconius* reveals divergent

1027 histories of convergent butterflies. Proc. Natl. Acad. Sci. U.S.A. 107:7365–7370.

1028

1029 Rambaut, A., Drummond, A.J. 2008. FigTree v1.1.2. Available from:

1030 http://tree.bio.ed.ac.uk/software/figtree.

1031

1032 Reeves, P.A., Richards, C.M. 2011. Species delimitation under the general lineage concept:

1033 An empirical example using wild north American hops (Cannabaceae: *Hummulus lupulus*).

1034 Syst. Biol. 60:45–59.

1035

1036 Ronquist, F., Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under

1037 mixed models. Bioinformatics 19:1572–1574.

1038

1039 Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population
1040 structure. Mol. Ecol. Notes. 4:137–138.

1041

1042 Sambrook, J., Fritsch, E.F., Maniatis, T.E. 1989. Molecular cloning: A laboratory manual. 2nd
1043 edition. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

1044

1045 Sanders, K.L., Malhotra, A. and Thorpe, R.S. 2004. Ecological diversification in a group of
1046 Indomalayan pitvipers (*Trimeresurus*): convergence in taxonomically important traits has
1047 implications for species identification. J. Evol. Biol., 17:721–731.

1048

1049 Sanders, K.L., Malhotra, A. and Thorpe, R.S. 2006. Evidence for a Müllerian mimetic
1050 radiation in Asian pitvipers. Proc. Biol. Sci. 273:1135–1141.

1051

1052 Song, S., Liu, L., Edwards, S.V., Wu, S. 2012. Resolving conflict in eutherian mammal
1053 phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad.
1054 Sci. U.S.A. 109(37):14942–14947.

1055

1056 Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L. Stadler,
1057 T., Steiner, C., Ryder, O.A., Janečka,J.E., Fisher, C.A., Muryphy, W. 2012.
1058 Macroevolutionary dynamics and historical biogeography of primate diversification inferred
1059 from a species supermatrix. PLoS One. 7:1–23.

1060

1061 Vogel, G., David, P., Pauwels, O.S.G. 2004. A review of morphological variation in
1062 *Trimeresurus popeiorum* (Serpentes: Viperidae: Crotalinae) with the description of two new
1063 species. Zootaxa 727:1–63.
1064
1065 Voris, H.K. 2000. Special Paper 2: Maps of Pleistocene sea levels in Southeast Asia:
1066 shorelines, river systems and time durations. J. Biogeog. 27:1153–1167.
1067
1068 Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J.,
1069 Peleman, J., Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids
1070 Res. 23:4407–4414.
1071
1072 Weisrock, D.W., Rasoloarison, R.M., Fiorentino, I., Ralison, J.M., Goodman, S.M.,
1073 Kappeler, P.M., Yoder, A.D. 2010. Delimiting species without nuclear monophyly in
1074 Madagascar's mouse lemurs. PLoS ONE 5:e9883.
1075
1076 Werle, E., Schneider, C., Renner, M., Volker, M., Fiehn, W. 1994. Convenient single-step,
1077 one tube purification of PCR products for direct sequencing. Nucleic Acids Res. 22:4354–
1078 4355.
1079
1080 Whitlock, R., Hipperson, H., Mannarelli, M., Burke, T. 2008. A high-throughput protocol for
1081 extracting high-purity genomic DNA from plants and animals. Mol. Ecol. Res. 8:736–741.
1082 Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R.K., Burke, T. 2008. An objective,
1083 rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping
1084 error. Mol. Ecol. Res. 8:725–735.
1085

1086 Wilgenbusch, J.C., Warren, D.L., Swofford, D.L. 2004. AWTY: a system for graphical
1087 exploration of MCMC convergence in Bayesian phylogenetic inference. Available from:
1088 http://ceb.csit.fsu.edu/awty.
1089
1090 Wooten, J.A., Camp, C.D., Rissler, L.J. 2010. Genetic diversity in a narrowly endemic,
1091 recently described dusky salamander, *Desmognathus folkertsi*, from the southern
1092 Appalachian Mountains. Conserv. Genet. 11:835–854.
1093
1094 Yeh, F.C., Yang, R.C., Boyle, T.B.J., Ye, Z.H., Mao, J.X. 1997. POPGENE, the
1095 user-friendly shareware for population genetic analysis. Molecular Biology and
1096 Biotechnology Centre, University of Alberta, Canada.
1097
1098 Zhang, D-X., Hewitt, G.M. 2003. Nuclear DNA analyses in genetic studies of populations:
1099 practice, problems and prospects. Mol. Ecol. 12:563–584.
1100

FIGURE CAPTIONS

Figure 1. Map of southeast Asia showing sampling scheme for mtDNA and AFLP analysis of the three putative species formerly assigned to *Trimeresurus macrops*, plus closely related species *T. venustus and T. kanburiensis*. MtDNA analysis included 39 samples and AFLP genotyping included 50 samples, of which 34 were also used in mtDNA analysis. The datasets used for the samples are represented by: empty circles (mtDNA); solid circle (AFLP); solid circle inside empty circle (both mtDNA and AFLP). Each location may represent multiple overlapping samples. The distribution of species in *T. macrops s.l.* is based on Malhotra et al. (2011a) and represents: *T. macrops sensu stricto* (orange)*; T. cardamomensis* (blue)*;* and *T. rubeus* (red). Samples in clear represent *Trimeresurus kanburiensis* from west Thailand and *T. venustus* from southern Thailand and Malaysia.

Figure 2. Mixed-model Bayesian phylogenetic reconstructions showing multiple distinct clades in the *Trimeresurus macrops* complex. 50% majority-rule consensus phylograms with Bayesian posterior probabilities derived from a) mtDNA and b) Total evidence dataset. The phylogenies are highly congruent and show three distinct clades corresponding to *T. macrops sensu stricto*, *T. cardamomensis,* and *T. rubeus*.

Figure 3. Population structure of the *Trimeresurus macrops* complex and *T. venustus* estimated using STRUCTURE. Summary plot shows estimated population assignments based on K = 4 with each color corresponding to a species.

Figure 4. Map of posterior probabilities of species membership and spatial location of genetic discontinuities estimated in GENELAND. Three clusters (K = 3) could be

visualized from the posterior distributions. The three plots represent assignment of pixels to each cluster: a) *T. macrops sensu stricto* and *T. cardamomensis*; b) *T. rubeus*; and c) *T. venustus*. Lightest colours indicate highest probabilities of membership and contour lines represent the spatial position of genetic discontinuities between species.

Figure 5. Scatter plots from multivariate analyses showing four clusters representing three species in *T. macrops* s.l., and *T. venustus*. The four species clusters are represented by different symbols (circles: *T. macrops sensu stricto;* triangles: *T. cardamomensis*; vertical crosses: *T. rubeus*; and diagonal crosses: *T. venustus*).

a) PCoA. Eigenvectors 1 and 3 using Gower General Similarity Coefficients are plotted. b) DAPC. The first three PCs of the DAPC are plotted. c) *prabclus.* Gaussian clusters on dimensions 1 and 3 identified from Non-metric Multi-Dimensional Scaling

Table 1 Pairwise Φ_{PT} values for populations of *Trimeresurus macrops* complex and outgroup *T. venustus*. Φ_{PT} were derived from AMOVA of 298 AFLP markers using 9999 random permutations. Significant *p-values* $(p<0.05)$ are highlighted by *.

Species	*T. macrops s.s.*	*T. cardamomensis*	*T. rubeus*	*T. venustus*
T. macrops s.s.				
T. cardamomensis	0.328*			
T. rubeus	0.555*	0.627*		
T. venustus	0.494*	0.579*	0.694*	

Table 2 Bayes Factor Delimitation of species in the *Trimeresurus macrops* complex using AFLPs (298 loci). Alternate species delimitation models were tested using SNAPP BFD against the four species model (Model A). Lumping of species is indicated by parentheses. The four species model is significantly better than all other models by 2 x $\log_e$ BF $\geq$ 408.

Model	Model Details	Species	ML	Rank	2 x $\log_e$ BF
A	*T. macrops s.s., T. cardamomensis, T. rubeus, & T. venustus*	4	-3757	1	
B	*(T. macrops s.s. + T. cardamomensis + T. rubeus) & T. venustus*	2	-4587	5	+1660
C	*(T. macrops s.s.+ T. cardamomensis), T. rubeus, & T. venustus*	3	-3961	2	+408
D	*(T. macrops s.s. + T. rubeus), T. cardamomensis, & T. venustus*	3	-4384	4	+1254
E	*T. macrops s.s., (T. cardamomensis + T. rubeus), & T. venustus*	3	-4138	3	+762

ML = Marginal likelihood
BF = Bayes factor