

Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*)

Adam D. Leaché^{1,2,*} and Matthew K. Fujita^{1,3}

¹Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA

²Genome Center and Department of Evolution and Ecology, University of California, Davis, CA 95616, USA

³Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Genealogical data are an important source of evidence for delimiting species, yet few statistical methods are available for calculating the probabilities associated with different species delimitations. Bayesian species delimitation uses reversible-jump Markov chain Monte Carlo (rjMCMC) in conjunction with a user-specified guide tree to estimate the posterior distribution for species delimitation models containing different numbers of species. We apply Bayesian species delimitation to investigate the speciation history of forest geckos (*Hemidactylus fasciatus*) from tropical West Africa using five nuclear loci (and mtDNA) for 51 specimens representing 10 populations. We find that species diversity in *H. fasciatus* is currently underestimated, and describe three new species to reflect the most conservative estimate for the number of species in this complex. We examine the impact of the guide tree, and the prior distributions on ancestral population sizes (θ) and root age (τ_0), on the posterior probabilities for species delimitation. Mis-specification of the guide tree or the prior distribution for θ can result in strong support for models containing more species. We describe a new statistic for summarizing the posterior distribution of species delimitation models, called speciation probabilities, which summarize the posterior support for each speciation event on the starting guide tree.

Keywords: Bayesian inference; phylogeography; reversible-jump MCMC; speciation probability

1. INTRODUCTION

Species play a central role in systematic studies and comparative analyses in ecology, evolution, conservation and biogeography (Agapow 2005). In order to accurately delimit species in nature, systematists require statistical methods for testing the alternative models that may describe the number of species in a clade. Genealogical data are among the most common sources of evidence for delimiting species, especially under certain historical scenarios such as allopatric speciation where tests of intrinsic reproductive isolation are not feasible. Analytical methods for species delimitation that use genetic data typically rely upon genetic distances or gene tree monophyly (Sites & Marshall 2003, 2004), both of which often require subjective decisions regarding the thresholds that demark the species boundary (Hey 2009). Species discovery should be amenable to statistical exploration (Carstens & Knowles 2007; O'Meara 2010), but statistical methods for estimating the probabilities associated with different species delimitations in a multilocus framework are lacking.

Speciation is a continuous process (de Queiroz 1998), and this implies that delimiting species using genealogical data should be accompanied by some degree of uncertainty. Bayesian species delimitation (Yang & Rannala in press) accomplishes this goal in a multilocus, multi-species coalescent framework while accommodating gene

tree uncertainty, which can be substantial at the recent time scales that are when many species delimitation problems arise (Knowles & Carstens 2007). This method is based on a population genetics perspective that includes prior information about population size and divergence times, two important parameters for making inferences about population history (Rannala & Yang 2003). The method uses reversible-jump Markov chain Monte Carlo (rjMCMC) to estimate the posterior distribution for species delimitation models. Each of these delimitation models assumes different numbers of species, with each species composed of three key parameters: θ (the product of effective population size N and mutation rate μ per site), τ_A (the time at which the species arose) and τ_D (the time at which the species splits into two descendent species). The joint posterior distribution of species delimitations and species trees is

$$f(S, \Lambda|D) = \frac{1}{f(D)} f(D|S) f(S|\Lambda) f(\Lambda),$$

where S denotes the species trees (and therefore the parameters θ , τ_A , and τ_D), Λ denotes the species delimitation models and D represents the multilocus data. Constraining S to a set of nested species trees based on a user-specified guide tree makes the problem computationally tractable. The guide tree is a fully resolved species tree, and the rjMCMC algorithm evaluates subtrees generated by collapsing (or splitting) nodes on the guide tree without performing any type of branch swapping. Under this method, we would expect strong support for species that have been isolated for an

* Author for correspondence (aleache@ucdavis.edu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.0662> or via <http://rspb.royalsocietypublishing.org>.

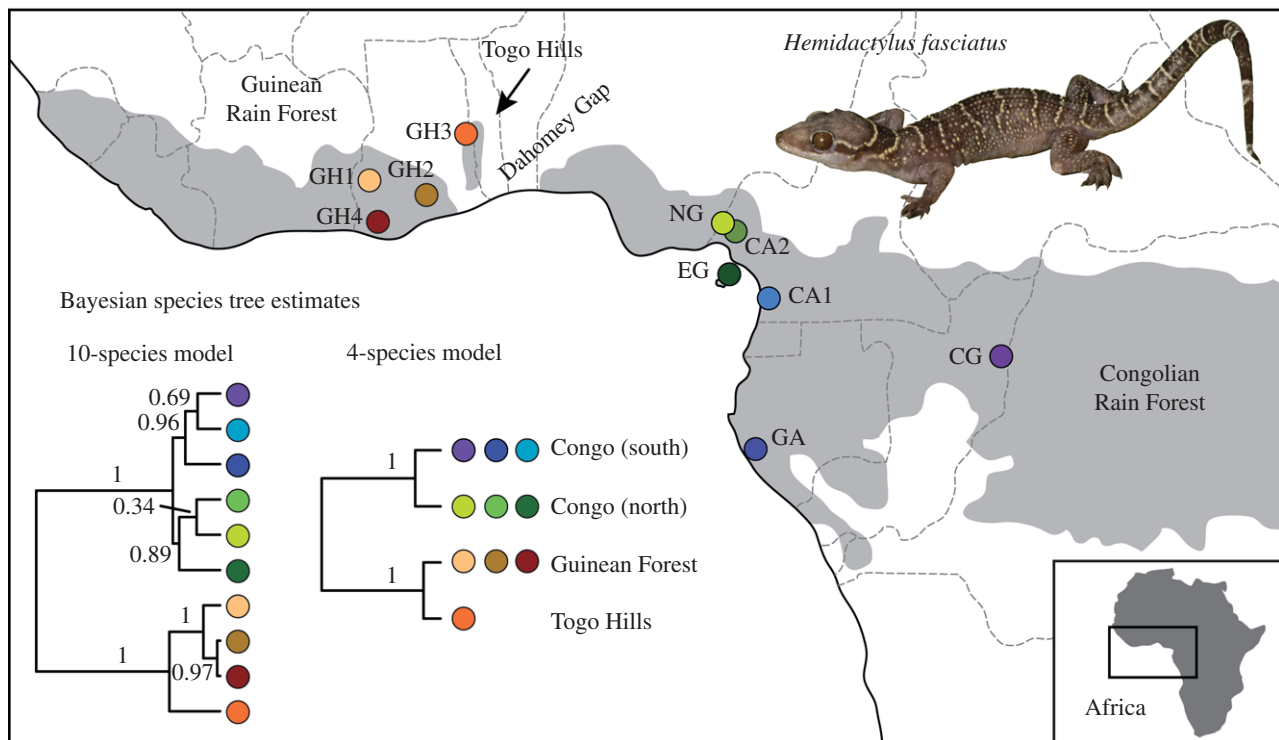


Figure 1. Populations of *Hemidactylus fasciatus* included in the study. The generalized distributions of the major rainforest fragments are shown in grey, which also approximate the distribution of *H. fasciatus*. Bayesian species trees are inferred using *BEAST, and numbers on nodes are posterior probability values.

extended period of time, and weak support for species that have experienced extensive gene flow.

Here, we investigate Bayesian species delimitation in west African forest geckos (*Hemidactylus fasciatus*). The genus *Hemidactylus* contains 80+ species comprising five major clades around the globe. *Hemidactylus fasciatus* occupies an isolated position in the phylogeny and is not allocated into any of these clades (Carranza & Arnold 2006). *Hemidactylus fasciatus* occurs in west African rainforests that are fragmented largely due to the expansion of dry forest and savannah during the last glacial maximum of the Pleistocene (figure 1; Hamilton & Taylor 1991; Dupont *et al.* 2000). This system presents several opportunities for testing allopatric divergence with genealogical data. The largest west African rainforest blocks are currently separated by the Dahomey Gap, a stretch of dry savannah extending from central Ghana through western Nigeria (Salzmann & Hoelzmann 2005; figure 1). Situated in the Dahomey Gap are the Togo Hills, a mountainous area containing moist semi-deciduous rainforest that forms a rainforest 'island' harbouring many rainforest specialists, including *H. fasciatus*, that are isolated from the more expansive rainforest blocks to the west and east (Leaché *et al.* 2006). In addition, *H. fasciatus* occurs in the Gulf of Guinea on Bioko Island. Bioko Island is separated from the west African mainland by only 32 km and sea depths of less than 60 m, and was probably connected to the mainland during the last glaciation (Lee *et al.* 1994).

2. MATERIAL AND METHODS

(a) Population sampling

Our analyses included 51 *H. fasciatus* specimens representing 10 populations (figure 1; electronic supplementary material,

table S1). We collected specimens from Ghana and Nigeria between 2003 and 2006, and the remaining samples were obtained as loans from natural history museum collections. The number of specimens sampled per population of *H. fasciatus* ranges from 1 to 19 (average = 5.1 specimens per population).

(b) Molecular methods

We collected sequence data for six loci: one mitochondrial (*12S*) and five nuclear (*BDNF*, *PNN*, *NGFB*, *FRIH*, *PRDX4*). Three of the nuclear markers are protein-coding (*BDNF*, *PNN*, *NGFB*; Townsend *et al.* 2008), while two are introns developed from a cDNA library (*FRIH*, *PRDX4*; Fujita *et al.* 2010). Intron 3 of *PRDX4* is new to this study, and the primers used for amplification are (5' → 3') TTATGGAGTTTATCTGGAAGATCAAGG (forward, situated in exon 2) and GGTAGGTCATTTCATTGT-TATCTGTCG (reverse, situated in exon 3). The data matrix is 95 per cent complete, missing only 16 sequences from a possible total of 318. All sequences are deposited at GenBank (accession numbers HM180090–HM180390).

The same general PCR protocols applied to all of the markers and are described in Fujita *et al.* (2010). The annealing temperatures (T_A) for each marker were 57 (*12S*), 57 (*PNN*), 53 (*BDNF*), 58 (*NGFB*), 60.7 (*FRIH*) and 66 (*PRDX4*). We purified successful PCR amplifications using ExoSAPIT (USB Corp.) sequenced with the original PCR primers using BigDye v. 3.1 chemistry, and collected the sequence data on an ABI3730 Genetic Analyzer (Applied Biosystems, Inc).

We used CODONCODE ALIGNER (v. 2; CodonCode Corp.) to assemble contigs. We aligned each set of sequences using MUSCLE (v. 3.6; Edgar 2004). Models of molecular evolution were chosen for each locus using MRMODELTEST v. 2.3, using the Akaike information criterion (AIC; Nylander 2004). To phase the nuclear data, we used the program

PHASE v. 2.1.1 (Stephens & Donnelly 2003), keeping the most probable haplotypes for each genotype for downstream analyses.

(c) Bayesian species delimitation

We explore two approaches to Bayesian species delimitation using multilocus data. The first approach involves three steps: (i) population structure inference provides an estimate of the number of populations and assignment of individuals to those populations; (ii) the phylogenetic relationships among populations are estimated using Bayesian species tree inference; and (iii) Bayesian species delimitation is performed using the population assignments and species tree estimated in steps 1 and 2. Our second approach is to treat each of the 10 sampled localities as a separate population (figure 1). This procedure eliminates the need to perform population structure inference, and is therefore composed of only steps 2 and 3 above.

Bayesian species delimitation was conducted using Bayesian Phylogenetics and Phylogeography (BPP v.2.0; Yang & Rannala in press) with the full phased dataset for the five nuclear loci. The model assumes no admixture following speciation, which is an assumption motivated by the biological species concept. Running the rjMCMC analyses for 500 000 generations (sampling interval of five) with a burn-in period of 10 000 produced consistent results across separate analyses initiated with different starting seeds. Ensuring adequate rjMCMC mixing involves specifying a reversible jump algorithm to achieve dimension matching between species delimitation models with different numbers of parameters, and we used algorithm 0 with the fine-tuning parameter $\varepsilon = 15.0$. Each species delimitation model was assigned equal prior probability.

The prior distributions on the ancestral population size (θ) and root age (τ_0) can affect the posterior probabilities for models, with large values for θ and small values for τ_0 favouring conservative models containing fewer species (Yang & Rannala in press). We evaluated the influence of these priors by considering three different combinations of prior. Both priors are assigned a gamma $G(\alpha, \beta)$ distribution, with a prior mean = α/β and prior variance = α/β^2 . The first combination of priors assumed relatively large ancestral population sizes and deep divergences: $\theta \sim G(1, 10)$ and $\tau_0 \sim G(1, 10)$, both with a prior mean = 0.1 and variance = 0.01. The second combination of priors assumed relatively small ancestral population sizes and shallow divergences among species: $\theta \sim G(2, 2000)$ and $\tau_0 \sim G(2, 2000)$, both with a prior mean = 0.001 and variance = 5×10^{-7} . The final combination is a mixture of priors that assume large ancestral populations sizes $\theta \sim G(1, 10)$ and relatively shallow divergences among species $\tau_0 \sim G(2, 2000)$, which is a conservative combination of priors that should favour models containing fewer species, and may be the most appropriate combination of priors for *H. fasciatus*.

(d) Population structure

We used the nuclear allele information for each individual to infer the number of populations and the assignment of individuals to those populations using STRUCTURAMA v. 1.0 (Huelsenbeck & Andolfatto 2007). The number of populations (K) is treated as a random variable with a Dirichlet process prior (DPP). We ran analyses with the number of populations fixed (e.g. $K = 1-10$), and with the number of populations following a DPP with a prior mean of

$E(K) = 1, 5$ and 10. All MCMC analyses were run for a total of 10 million generations (sampling 1000 steps and excluding 200 as burn-in).

(e) Species tree inference

We used the hierarchical Bayesian model implemented in *BEAST v. 1.5.3 (Heled & Drummond 2010) to estimate species trees for the populations identified by STRUCTURAMA and for the 10 sampled localities (figure 1). *BEAST estimates the species tree directly from the sequence data, and incorporates uncertainty associated with gene trees, nucleotide substitution model parameters and the coalescent process (Heled & Drummond 2010). Phased nuclear alleles were used in *BEAST analyses. We repeated each analysis twice and MCMC analyses were run for a total of 50 million generations (sampling 1000 steps and excluding the first 20% as burn-in). We assessed convergence by examining the likelihood plots through time using TRACER v. 1.4.1 (Rambaut & Drummond 2007).

(f) Impact of guide trees

The guide tree is a fully resolved species tree that specifies the relationships among the species included in the analysis. Ideally, uncertainty in the guide tree could be incorporated into the analysis, but this option is not currently implemented. To investigate the impacts of a mis-specified guide tree on species delimitation, we conducted analyses with a guide tree that was similar in shape to the 10-species guide tree (to retain the same prior distribution), but with random rearrangement of the 10 species at the tips.

(g) Speciation probabilities

Whereas typical Bayesian phylogenetic analyses provide posterior probability estimates for clades, the species delimitation method assumes a fixed tree topology and does not provide such values. Instead, the rjMCMC algorithm estimates the posterior distribution for species delimitation models that differ in the number of species, all of which are compatible with the starting guide tree. However, we can estimate the marginal posterior probability of speciation associated with each bifurcation in the guide tree, which we call speciation probabilities, by summing the probabilities for all models that support a particular speciation event in the guide tree. A speciation probability of 1 on a node indicates that every species delimitation model visited by the rjMCMC algorithm supports the two lineages descending from that node as species. Conversely, a speciation probability of 0 reflects the situation where all of the species delimitation models in the posterior distribution collapsed that particular node to one species. We consider speciation probability values ≥ 0.95 as strong support for a speciation event.

3. RESULTS

(a) Population structure

STRUCTURAMA estimates for population structure in *H. fasciatus* based on the five nuclear loci support four populations (table 1). The marginal likelihoods for varying numbers of populations (ranging from $K = 1$ to 10) reach a high point at $K = 4$ (table 1). In the Bayesian analyses using the Dirichlet process prior with different mean values, the maximum posterior probability estimate for the number of populations is also $K = 4$ (table 1). The individuals assigned to the four populations are constant across the

Table 1. Estimates of population structure in *Hemidactylus fasciatus* based on five nuclear loci. The number of populations with the highest posterior probability and marginal likelihood is $K = 4$. Population assignments are conserved across analyses: $\text{pop}_1\{\text{GH1, GH2, GH4}\}$, $\text{pop}_2\{\text{GH3}\}$, $\text{pop}_3\{\text{CA1, CG, GA}\}$, $\text{pop}_4\{\text{CA2, EG, NG}\}$.

| K | posterior probability distributions | | | marginal likelihoods $\text{Pr}(X K)$ |
|-----|-------------------------------------|-------------|-------------|--|
| | $E(K) = 1$ | $E(K) = 5$ | $E(K) = 10$ | |
| 1 | — | — | — | -944.33 |
| 2 | — | — | — | -804.23 |
| 3 | 0.01 | — | — | -691.12 |
| 4 | 0.99 | 0.83 | 0.55 | -670.36 |
| 5 | — | 0.16 | 0.38 | -673.44 |
| 6 | — | 0.02 | 0.07 | -676.35 |
| 7 | — | — | 0.01 | -679.56 |
| 8 | — | — | — | -686.75 |
| 9 | — | — | — | -681.63 |
| 10 | — | — | — | -685.000 |

separate analyses: Guinean Forest = $\text{pop}_1\{\text{GH1, GH2, GH4}\}$, Togo Hills = $\text{pop}_2\{\text{GH3}\}$, Congolian Forest south = $\text{pop}_3\{\text{CA1, CG, GA}\}$ and Congolian Forest north = $\text{pop}_4\{\text{CA2, EG, NG}\}$.

(b) Species tree inference

The five nuclear genes that we sequenced for *H. fasciatus* contained between 13 and 19 segregating sites, whereas the mtDNA gene (*12S*) contained 128 segregating sites (see the electronic supplementary material, table S2). In general, gene genealogies estimated for the nuclear loci support a separation between populations from the Guinean and Congolian rainforests, and the mtDNA gene tree supports each population as distinct (methodological details and gene trees are provided in the electronic supplementary material).

The species tree resulting from the analysis of the four populations identified in STRUCTURAMA provides strong support (posterior probability = 1.0) for a split between populations on different sides of the Dahomey Gap, a split between the Togo Hills and the Guinean Forest, and a split between the northern and southern Congolian Forest (figure 1). The *BEAST analysis with the 10 sampled localities each treated as separate species supports the same clades recovered in the species tree analysis with four populations, which indicates that there is consistency between the major clades supported by the two approaches (figure 1).

(c) Bayesian species delimitation

The Bayesian species delimitation results for *H. fasciatus* are shown in figure 2. When assuming four species, Bayesian species delimitation supports the guide tree with speciation probabilities of 1.0 on all nodes (figure 2a). Different prior distributions for θ and τ_0 do not affect this outcome (figure 2a). For the analyses assuming 10 species, the posterior probability distributions of models support up to 15 different species delimitation models with posterior probability values ≥ 0.01 (see the electronic supplementary material, table S3). Five species are supported by speciation probabilities of 1.0 on the 10-species guide tree; these species match those supported

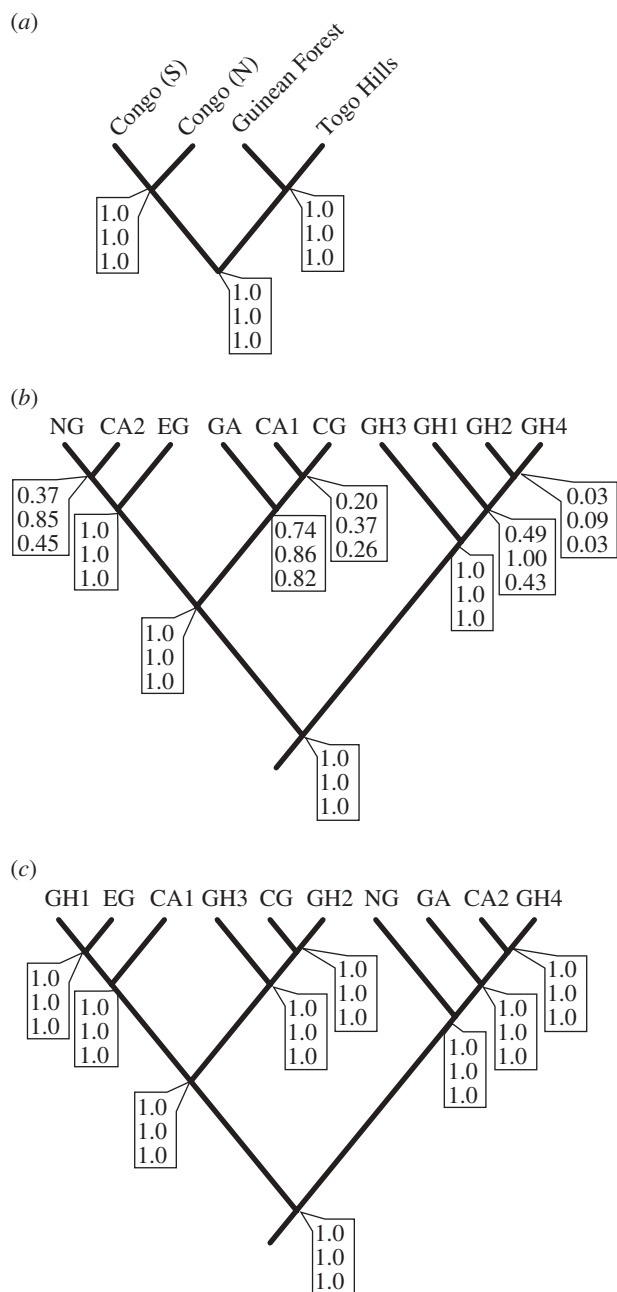


Figure 2. Bayesian species delimitation results for *Hemidactylus fasciatus* assuming 4-species (a) and 10-species (b and c) guide trees. The speciation probabilities are provided for each node under each combination of priors for θ and τ_0 : top, prior means = 0.1; middle, prior means = 0.001; bottom, prior mean $\theta = 0.1$, prior mean $\tau_0 = 0.001$. A small prior mean for θ results in higher speciation probabilities, although similar changes to the prior distribution of τ_0 have minor impacts. Randomly rearranging the labels on the guide tree (c) results in speciation probabilities of 1.0 for each node.

on the 4-species guide tree, with the addition of a split within the northern Congo clade that supports the Bioko Island population as distinct (figure 2b). The remaining speciation events are not strongly supported and are influenced by the priors (figure 2b).

(d) Impact of priors and guide trees

Randomizing the species at the tips of the 10-species guide tree results in speciation probabilities of 1 for

every speciation event across all three combinations of priors tested (figure 2c). Our explanation for this result is that the chance placement of divergent populations as sister taxa creates artificially large divergences among descendent species, which the model interprets as speciation events. This result illustrates the large impact that an incorrectly specified guide tree can have on Bayesian species delimitation. Additional Bayesian species delimitation analyses using all 15 possible topologies for the northern Congo samples as guide trees show that even moderate changes to the guide tree can impact the support for models (see the electronic supplementary material, table S4).

For the 10-species guide tree, the prior distribution for θ applied to these empirical data has a large impact on the species delimitation results. Assuming a small θ prior mean (0.001) and variance (5×10^{-7}) results in increased speciation probabilities compared with analyses using a larger θ prior mean (0.1) and variance (0.01) (figure 2b). Similar changes to τ_0 have little to no effect on speciation probabilities (figure 2b).

4. DISCUSSION

(a) Bayesian species delimitation

The guide tree plays a critical role in the outcome of the species delimitation model. We used guide trees generated from species tree analyses of the nuclear loci, although any number of species tree inference approaches could be used. Randomizing the relationships among species produced speciation probabilities of 1.0 for every speciation event (figure 2c). Although there is no biological motivation for conducting species delimitations using random trees, this example illustrates the negative impacts of artificially increasing the level of divergence among sister species by guide tree mis-specification. When the best estimate for the species tree contains ambiguous relationships, we recommend comparing the species delimitation results obtained from guide trees representing all possible resolutions of those polytomies.

Incorporating separate prior probability distributions for ancestral population sizes (θ) and the root age (τ_0) is a benefit of the Bayesian species delimitation method as long as the priors are not overwhelmingly misleading in relation to the relative amount of information contained in the data. We varied the prior distributions for θ and τ_0 to have mean values that differed by two orders of magnitude (i.e. 0.1–0.001), and even larger differences in variance (0.01 – 5×10^{-7}). Based on our empirical data for *H. fasciatus*, the species delimitation results are not affected by these changes to τ_0 ; however, assuming a small prior mean and variance for θ resulted in higher speciation probabilities for recently diverged species (figure 2b). In general, these results agree with the suggestion offered by Yang & Rannala (in press) that a combination of large θ and small τ_0 should favour fewer species.

(b) Speciation in forest geckos

Our investigation of species delimitation in *H. fasciatus* using multilocus nuclear data reveals the potential for recognizing at least four genetically distinct species (figure 3). These species have distributions that match those of the major blocks of rainforests, suggesting that

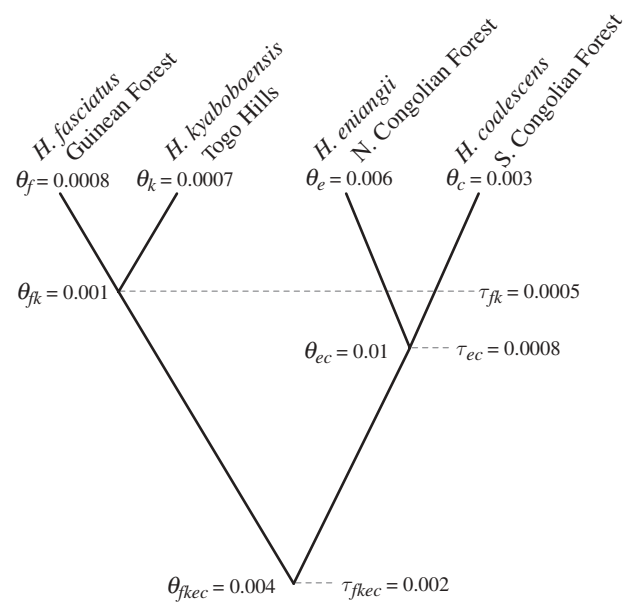


Figure 3. Bayesian species delimitation supports 4 species within *Hemidactylus fasciatus*. The posterior estimates (mean of the distribution) for θ and τ are provided on the species tree. The 10 species guide tree supports 5 species, with the extra species representing the population on Bioko Island (part of *H. eniangii* under the model proposed here).

new lineages have become reproductively isolated from each other as a result of habitat fragmentation, which has driven allopatric speciation. Populations from the same forest fragment are collapsed into a single species by the Bayesian species delimitation method, which may be an indication that these populations are still connected by gene flow. This is not the case with respect to the central Congolian forest where there appears to be a species boundary in Cameroon near the Sanaga River. Other vertebrate groups with distributions spanning tropical west Africa that show similarly high levels of genetic structure include amphibians (Blackburn 2008), birds (Bowie *et al.* 2004; Marks 2009), murid rodents (Nicolas *et al.* 2008) and chimpanzees (Hey 2010).

(c) Systematics

Systematists generally disagree on which criteria should be used to judge whether a population is a species, and these differences in opinion are more acute when considering allopatric speciation (Wiens 2004). Species delimitation is controversial when populations are allopatric because of difficulties associated with assessing properties inherent to the biological species concept, including natural reproduction resulting in viable and fertile offspring and intrinsic reproductive isolation (Mayr 1942; Dobzhansky 1950). This is not a major concern from the perspective of a lineage-based species concept (de Queiroz 1998), since reproductive isolation represents just one of the many criteria available to delimit species in nature (de Queiroz 2007). Ideally, a combination of genetic, morphological and ecological criteria can be used to delimit species (e.g. Leaché *et al.* 2009; Ross *et al.* 2009). In terms of *H. fasciatus*, we are not aware of any morphological or ecological characteristics that differentiate these lineages.

Given the fairly strict conditions that the Bayesian species delimitation method assumes to designate species, we feel that recognizing four species is a conservative estimate. Additional data from morphology, ecology and physiology can only strengthen the evidence that these four species are distinct. This is the situation today; it may be that future climate change and demographic fluctuations will cause some of these species to merge, diversify or go extinct, and a reappraisal of species assignments may change, but the contemporary scenario supports *four species*.

(i) *Hemidactylus coalescens* *sp. nov.*

Holotype. Zoologisches Forschungsinstitut und Museum Alexander Koenig (ZFMK) 87680, adult male; Cameroon, Campo Region, Nkoelon, 2.3972° N, 10.04515° E, 85 m; collected by Michael F. Barej and Julia Wurstner, 27 October 2007. Paratype = ZFMK 87679.

Diagnosis. This species includes all populations that cluster with those from the southern portion of the Congolian rainforest included in this study (southern Cameroon, Gabon and Congo), with strong support in the Bayesian species delimitation model.

Etymology. This species is named after the coalescent process used to delimit the species.

(ii) *Hemidactylus eniangii* *sp. nov.*

Holotype. Museum of Vertebrate Zoology (MVZ) 253215, adult male; Nigeria, Cross River State, Cross River National Park, Oban Hills Sector, Southern Portion, Erokut Park entry gate, 05.3639° N, 08.43341° E, 143 m; collected by Adam D. Leaché, Anne M. Leaché, and Edem A. Eniang, 6 April 2006. Paratypes = MVZ 253213, 253214.

Diagnosis. This species includes all populations that cluster with those from the northern portion of the Congolian rainforest included in this study (eastern Nigeria, Equatorial Guinea, and northern Cameroon) with strong support in the Bayesian species delimitation model.

Etymology. This species is named in honour of Nigerian conservation biologist and herpetologist Dr Edem A. Eniang.

(iii) *Hemidactylus fasciatus* Gray, 1842

Holotype. The locality of the type specimen (Natural History Museum, London; xxi.24.a) is unknown. We restrict the type locality to the western Guinean rainforest from Sierra Leone to the Dahomey Gap in Ghana. By doing so the synonyms also rest in this form (*Leiurus ornatus* Gray 1845 [type locality = W. Africa]; *Hemidactylus formosus* Hallowell 1857 [type locality = Liberia]). We assign the subspecies *H. fasciatus ituriensis* Schmidt 1919 as distinct from *H. fasciatus* based on the non-overlapping characters presented in Loveridge (1947) and Schmidt (1919).

Diagnosis. This species includes all populations that cluster with those from the western Guinean rainforest included in this study, with strong support in the Bayesian species delimitation model.

(iv) *Hemidactylus kyaboboensis* *sp. nov.*

Holotype. Museum of Vertebrate Zoology (MVZ) 245 291, adult male; Ghana, Volta Region, Togo Hills, Kyabobo National Park, Waterfall, 08.33019° N, 00.59411° E, 515 m; collected by Adam D. Leaché, Raul Diaz and Matthew K. Fujita, 16 June 2004. Paratypes = MVZ 245292–245299.

Diagnosis. This species includes all populations that cluster with those from the Togo Hills included in this study with strong support in the Bayesian species delimitation model.

Etymology. This species is named after Kyabobo National Park, Togo Hills, Volta Region, Ghana.

We thank the Ghanaian Wildlife Division, Nigerian Biodiversity Research Center, Conservation International, Nature Conservation Research Center and IUCN for assistance with fieldwork. For tissue loans, we thank the Ambrose Monell Cryo Collection at the AMNH, the Smithsonian, LSUMNS, A. Bauer, K. Jackson and M. Barej (ZFMK). We thank Z. Yang and B. Rannala for providing us with beta versions of the BPP software package and for including speciation probabilities as a default output in their program. For useful comments we thank D. Iskandar, J. McGuire, C. Moritz, B. Rannala, S. Singhal, Z. Yang and two anonymous reviewers. Funding was provided by the National Science Foundation (DEB-0508929 and DBI-0805455 to A.D.L., and DBI-0905714 to M.K.F.).

REFERENCES

- Agapow, M. 2005 Species: demarcation and diversity. In *Phylogeny and Conservation* (eds A. Purvis, J. L. Gittleman & T. Brooks), pp. 57–75. Cambridge, UK: Cambridge University Press.
- Blackburn, D. C. 2008 Biogeography and evolution of body size and life history of African frogs: phylogeny of squeakers (*Arthroleptis*) and long-fingered frogs (*Cardioglossa*) estimated from mitochondrial data. *Mol. Phylogenet. Evol.* **49**, 806–826. (doi:10.1016/j.ympev.2008.08.015)
- Bowie, R. C. K., Fjeldså, J., Hackett, S. J. & Crowe, T. M. 2004 Molecular evolution in space and through time: mtDNA phylogeography of the Olive Sunbird (*Nectarinia olivacea/obscura*) throughout continental Africa. *Mol. Phylogenet. Evol.* **33**, 56–74. (doi:10.1016/j.ympev.2004.04.013)
- Carranza, S. & Arnold, E. N. 2006 Systematics, biogeography, and evolution of *Hemidactylus* geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **38**, 531–545. (doi:10.1016/j.ympev.2005.07.012)
- Carstens, B. C. & Knowles, L. L. 2007 Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Syst. Biol.* **56**, 400–411. (doi:10.1080/10635150701405560)
- de Queiroz, K. 1998 The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In *Endless forms: species and speciation* (eds D. J. Howard & S. H. Berlocher), pp. 57–75. Oxford, UK: Oxford University Press.
- de Queiroz, K. 2007 Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886. (doi:10.1080/10635150701701083)
- Dobzhansky, T. 1950 Mendelian populations and their evolution. *Am. Nat.* **84**, 401–418. (doi:10.1086/281638)
- Dupont, L. M., Jahns, S., Marret, F. & Ning, S. 2000 Vegetation change in equatorial West Africa: time-slices for

- the last 150 ka. *Palaeogeogr. Palaeoclimatol.* **155**, 95–122. (doi:10.1016/S0031-0182(99)00095-4)
- Edgar, R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
- Fujita, M. K., McGuire, J. A., Donnellan, S. C. & Moritz, C. 2010 Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). *Evolution*. (doi:10.1111/j.1558-5646.2010.00993.x)
- Hamilton, A. C. & Taylor, D. 1991 History of climate and forests in tropical Africa during the last 8 million years. *Clim. Change* **19**, 65–78. (doi:10.1007/BF00142215)
- Heled, J. & Drummond, A. J. 2010 Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580. (doi:10.1093/molbev/msp274)
- Hey, J. 2009 On the arbitrary identification of real species. In *Speciation and patterns of diversity* (eds R. K. Butlin, J. R. Bridle & D. Schluter), pp. 15–28. Cambridge, UK: Cambridge University Press.
- Hey, J. 2010 The divergence of Chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933. (doi:10.1093/molbev/msp298)
- Huelsenbeck, J. P. & Andolfatto, P. 2007 Inference of population structure under a Dirichlet process model. *Genetics* **175**, 1787–1802. (doi:10.1534/genetics.106.061317)
- Knowles, L. L. & Carstens, B. C. 2007 Delimiting species without monophyletic gene trees. *Syst. Biol.* **56**, 887–895. (doi:10.1080/10635150701701091)
- Leaché, A. D., Rödel, M.-O., Linkem, C. W., Diaz, R. E., Hillers, A. & Fujita, M. K. 2006 Biodiversity in a forest island: reptiles and amphibians of the West African Togo Hills. *Amphib. Reptile Conserv.* **41**, 22–45.
- Leaché, A. D., Koo, M. S., Spencer, C. L., Papenfuss, T. J., Fisher, R. N. & McGuire, J. A. 2009 Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proc. Natl Acad. Sci. USA* **106**, 12 418–12 423. (doi:10.1073/pnas.0906380106)
- Lee, D.-C., Halliday, A. N., Fitton, J. G. & Poli, G. 1994 Isotopic variations with distance and time in the volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth Planet. Sci. Lett.* **123**, 119–138. (doi:10.1016/0012-821X(94)90262-3)
- Loveridge, A. 1947 Revision of the African lizards of the family Gekkonidae. *Bull. Mus. Comp. Zool.* **98**, 1–469.
- Marks, B. D. 2009 Are lowland rainforests really evolutionary museums? Phylogeography of the green hylia (*Hylia prasina*) in the Afrotropics. *Mol. Phylogenet. Evol.* **55**, 178–184. (doi:10.1016/j.ympev.2009.10.027)
- Mayr, E. 1942 *Systematics and the origin of species*. New York, NY: Columbia University of Press.
- Nicolas, V., Mboumba, J.-F., Verheyen, E., Denys, C., Lecompte, E., Olayemi, A., Missou, A. D., Katuala, P. & Colyn, M. 2008 Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *J. Biogeogr.* **35**, 2074–2089. (doi:10.1111/j.1365-2699.2008.01950.x)
- Nylander, J. A. A. 2004 MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.
- O'Meara, B. C. 2010 New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* **59**, 59–73. (doi:10.1093/sysbio/syp077)
- Rambaut, A. & Drummond, A. J. 2007. Tracer v1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>.
- Rannala, B. & Yang, Z. 2003 Bayes Estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656.
- Ross, K. G., Gotzek, D., Ascunce, M. S. & Shoemaker, D. D. 2009 Species delimitation: a case study in a problematic ant taxon. *Syst. Biol.* **59**, 162–184. (doi:10.1093/sysbio/syp089)
- Salzmann, U. & Hoelzmann, P. 2005 The Dahomey gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* **15**, 190–199. (doi:10.1191/0959683605hl799rp)
- Schmidt, K. P. 1919 Contributions to the herpetology of the Belgian Congo based on the collection of the American Museum Congo Expedition 1909–1915. Part I. Turtles, crocodiles, lizards and chameleons. *Bull. Am. Mus. Nat. Hist.* **39**, 385–624.
- Sites Jr, J. W. & Marshall, J. C. 2003 Delimiting species: a renaissance issue in systematic biology. *Trends Ecol. Evol.* **18**, 462–470. (doi:10.1016/S0169-5347(03)00184-8)
- Sites Jr, J. W. & Marshall, J. C. 2004 Operational criteria for delimiting species. *Annu. Rev. Ecol. Syst.* **35**, 199–227. (doi:10.1146/annurev.ecolsys.35.112202.130128)
- Stephens, M. & Donnelly, P. 2003 A comparison of Bayesian methods for haplotype reconstruction. *Am. J. Hum. Genet.* **73**, 1162–1169. (doi:10.1086/379378)
- Wiens, J. J. 2004 What is speciation and how should we study it? *Am. Nat.* **163**, 914–923. (doi:10.1086/386552)
- Yang, Z. & Rannala, B. In press. Bayesian species delimitation using multilocus sequence data. *Proc. Natl Acad. Sci. USA* **107**, 9264–9269. (doi:10.1073/pnas.0913022107).