Running head: Comparative biogeography of Philippine geckos

The comparative biogeography of geckos challenges predictions from a paradigm of climate-driven vicariant diversification across an island archipelago

Jamie R. Oaks *[1], Cameron D. Siler[2], and Rafe M. Brown[3]

[1]Department of Biological Sciences & Museum of Natural History, Auburn University, Auburn, Alabama 36849, USA
[2]Sam Noble Oklahoma Museum of Natural History and Department of Biology, University of Oklahoma, Norman, Oklahoma 73072
[3]Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045, USA

August 18, 2018

Abstract

A primary goal of biogeography is to understand how large-scale environmental processes, like climate change, affect diversification. One often-invoked but seldom tested process is the so-called "species-pump" model, in which repeated bouts of co-speciation is driven by oscillating climate-induced habitat connectivity cycles. For example, over the past three million years, the landscape of the Philippine Islands has repeatedly coalesced and fragmented due to sea-level changes associated with the glacial cycles. This repeated climate-driven vicariance has been proposed as a model of speciation across evolutionary lineages codistributed throughout the islands. This model predicts speciation times that are temporally clustered around the times when interglacial rises in sea level fragmented the islands. Given the significance and conceptual impact the model has shown, surprisingly few tests of this prediction have been provided. We collected comparative genomic data from 16 pairs of insular gecko populations to test the prediction of temporally clustered divergences. Specifically, we analyze these data in a full-likelihood, Bayesian model-choice framework to test for shared divergence times among the pairs. Our results provide support against the species-pump model prediction in favor of an alternative interpretation, namely that each pair of gecko populations diverged independently. These results suggest the repeated bouts of climate-driven landscape fragmentation has not been an important mechanism of

*Corresponding author: joaks@auburn.edu

speciation for gekkonid lizards on the Philippine Islands. Interpretations of shared mechanisms of diversification historically have been pervasive in biogeography, often advanced on the basis of taxonomy-based depictions of species distributions. Our results call for possible re-evaluation of other, classic co-diversification studies in a variety of geographic systems.

KEY WORDS: phylogeography, biogeography, Philippines, diversification

1 Introduction

Understanding how environmental changes affect diversification is an important goal in evolutionary biology, biogeography, and global change biology. Environmental processes that operate at or above the level of communities can simultaneously cause speciation or extinction across multiple evolutionary lineages, and thus have a pronounced effect on the diversity and distribution of species. Island archipelagos that harbor diverse communities of co-distributed lineages and have a relatively well-understood geological history present powerful systems for understanding such shared processes of diversification (Gillespie, 2007; Losos and Ricklefs, 2009; Vences et al., 2009; Brown et al., 2013). The Philippine Islands represent such a model system, with more than 7100 islands that arguably harbor the highest concentration of terrestrial biodiversity on Earth (Catibog-Sinha and Heaney, 2006; Brown and Diesmos, 2009; Heaney and Regalado, 1998; Brown et al., 2013); how this diversity accumulated has been a topic of discussion since the early development of the field of biogeography (Wallace, 1869; Huxley, 1868).

The landscape of the Philippines has experienced a complex history. Climatological oscillations, primarily during the Pleistocene, led to the repeated formation and fragmentation of Pleistocene Aggregate Island Complexes (PAICs; Inger, 1954; Heaney, 1985; Brown and Diesmos, 2002, 2009; Esselstyn and Brown, 2009; Lomolino et al., 2010; Siler et al., 2010; Brown and Siler, 2014). During lower sea levels of glacial periods, islands coalesced into seven major landmasses (PAICs) that were fragmented into individual islands during interglacial periods. These climate-driven cycles have occurred at least six times during the last 500,000 years (Rohling et al., 1998; Siddall et al., 2003), with additional cycles occurring in the late Pliocene and early Pleistocene (Haq et al., 1987; Miller et al., 2005).

For nearly three decades, the repeated formation and fragmentation of PAICs has been a prominent model of diversification in the Philippines (Inger, 1954; Heaney, 1985; Brown and Guttman, 2002; Evans et al., 2003; Heaney et al., 2005; Roberts, 2006; Linkem et al., 2010; Siler et al., 2010, 2011, 2012; Brown and Siler, 2014). However, there is growing recognition of the complexity of historical processes that were involved in the diversification of this megadiverse archipelago. For example, there is evidence that older tectonic processes contributed to vertebrate diversification on precursor paleoislands that predates the modern distribution of landmasses in the Philippines (∼30–5 mya; Jansa et al., 2006; Blackburn et al., 2010; Siler et al., 2012; Brown and Siler, 2014). Additionally, the region's biodiversity has likely been shaped further by dispersal events from mainland source populations via recognized colonization routes (see Brown and Siler, 2014, for a review) and finer-scale isolating mechanisms that led to in situ diversification (Heaney et al., 2011; Linkem et al., 2011). Nonetheless, the question remains: Was climate-driven fragmentation of the islands an important process of speciation?

A "species-pump" model of diversification via repeated vicariance predicts that divergences across taxa that occur on historically connected islands (i.e., islands within the same PAIC) should tend to be clustered around times when rising sea levels fragmented PAICs into the islands of today. Therefore, if we compare the divergence times of multiple pairs of populations or closely related species that occur on two islands that were connected during glacial periods of lower sea levels, we expect some to be contemporaneous with interglacial fragmentation events. Such patterns of shared divergences would be difficult to explain by

other mechanisms, such as over-water dispersal.

Oaks et al. (2013) initially tested this prediction by inferring how many unique divergence times best explained mitochondrial sequence data from 22 pairs of populations from across the Philippines, using a model choice method based on approximate-likelihood Bayesian computation (ABC). However, using simulations, they found that this popular ABC approach was demonstrably sensitive to prior assumptions, with a bias toward over-clustering divergence times, both of which rendered the results difficult to interpret and potentially skewed toward interpretations of simultaneous diversification. Oaks (2014) reanalyzed these data with an ABC method that alleviated these issues; however, they found that reducing the genetic data to a small number of summary statistics left ABC methods with little information to update prior assumptions.

Here we use comparative genomic data and a new full-likelihood Bayesian method to test the hypothesis that repeated fragmentation of islands was a causal mechanism of vicariant diversification among terrestrial vertebrates in the Philippines. By using all the information in thousands of loci from 16 inter-island pairs of gecko populations, we demonstrate a new method that provides the first robust evaluation of this central tenet of the PAIC model of diversification. Our results support independent diversification among pairs of gecko populations, provide evidence against fundamental predictions of the PAIC model of diversification, and underscore the importance for caution against adhering to overly simplistic models of diversification based on little evidence when studying dynamic and biodiverse regions such as the Philippines.

2 Methods

2.1 Sampling

For two genera of geckos, *Cyrtodactylus* and *Gekko*, we sampled individuals from pairs of populations that occur on two different islands. Because the climate-mediated fragmentation of the islands was a relatively recent phenomenon, we selected pairs of populations that were inferred to be closely related from previous genetic data (Siler et al., 2012, 2014; Welton et al., 2010b,a; Siler et al., 2010). In other words, we avoided pairs that we knew *a priori* diverged well before the connectivity cycles, because these cannot provide insight into whether divergences were clustered *during* these cycles.

We also sought to sample pairs that spanned islands connected during glacial periods, as well as islands that were never connected. We included the latter as "controls." Because these islands were never connected, the distribution of the populations inhabiting them can only be explained by inter-island dispersal. The divergence between these populations was either due to that dispersal, or an earlier intra-island divergence. Either way, the timing of divergences across islands that were never connected are not expected to be clustered across pairs. These controls are important given the tendency for previous approaches to this inference problem to over-estimate shared divergences (Oaks et al., 2013, 2014). Finding shared shared divergence times among pairs for which there is no tenable mechanism for shared divergences will indicate a problem and prevent us from misinterpreting shared divergences among pairs spanning islands that were fragmented as evidence for the PAIC model of vi-

cariant diversification. Applying these criteria, for both genera, we identified eight pairs of populations, including a mix of pairs spanning islands that were never connected, islands that were connected, and islands that were possibly connected during glacial lowstands (Figure 1; Table 1 & S1). The Islands of Bohol and Camiguin Sur, Palawan and Borneo, Sibuyan and Tablas, and Sabtang and Batan are not believed to have been fully joined during glacial periods. However, even if these pairs of islands did not have a complete land connection, they may have been close enough to permit some limited gene flow given their immediate proximity and the relative dispersal ability of gekkonid lizards. As a result, we leave their connection status during glacial lowstands as ambiguous (Figure 1 & Table 1).

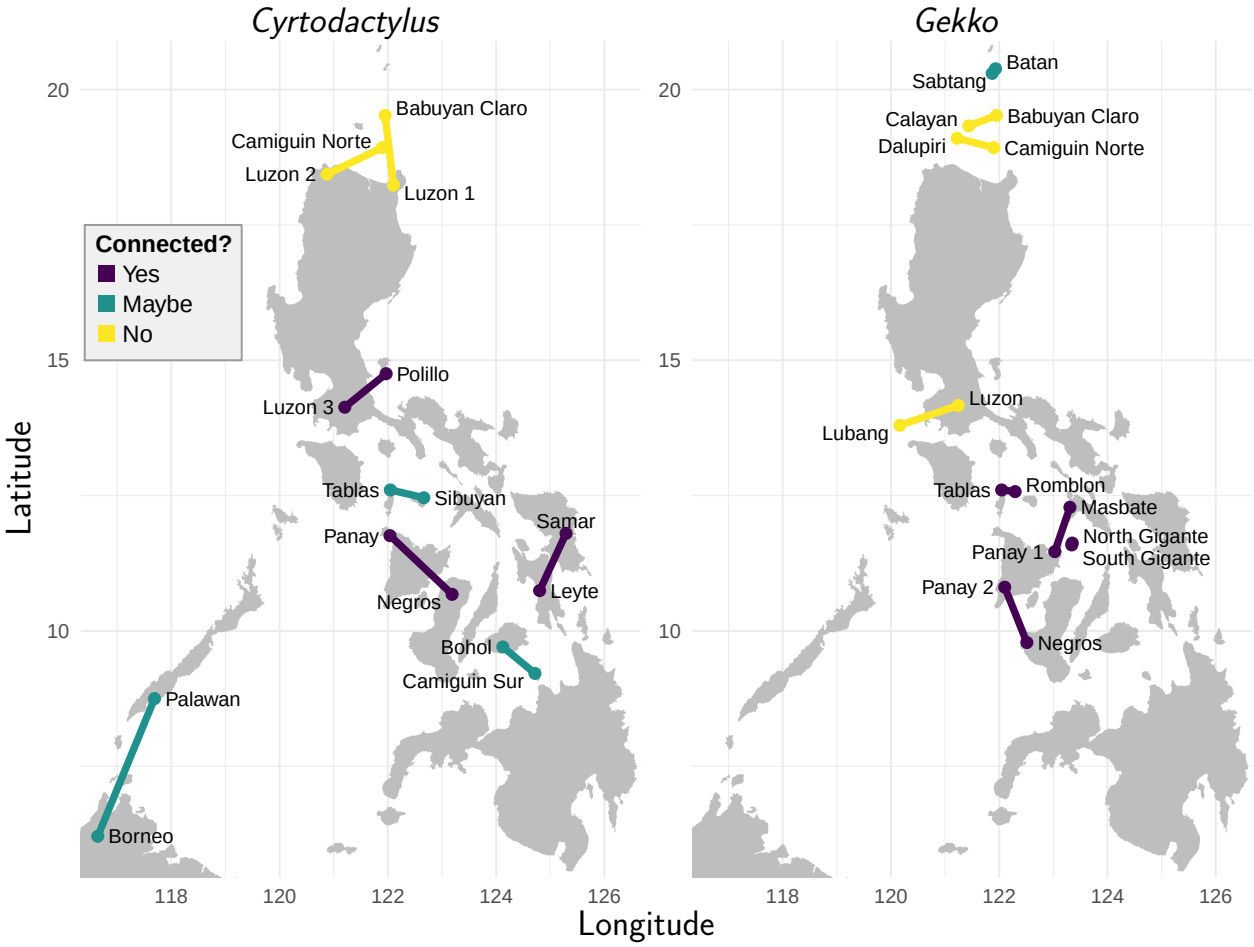

Figure 1. Philippine sampling localities for the eight pairs of *Cyrtodactylus* (left) and *Gekko* (right) populations. Localities for each pair are connected by a line and color-coded (see key) to indicate whether the islands were connected via terrestrial dry land bridges that formed during Pleistocene glacial periods. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

Table 1. Pairs of *Cyrtodactylus* and *Gekko* populations included in our full-likelihood Bayesian comparative biogeographic analyses (`ecoevolity`). Each row represents a pair of populations sampled from two islands that either were or were not connected during low sea levels of glacial periods.

Species	Island 1	Island 2	Connected?	Sample sizes		# loci	# sites	# variable	# polyallelic
C. annulatus	Bohol	Camiguin Sur	?	4	4	15,500	1,411,669	12,469	61
C. redimiculus-baluensis	Palawan	Borneo	?	4	3	8989	815,005	25,700	239
C. sumuroi-gubaot	Samar	Leyte	yes	5	5	18,759	1,709,440	38,862	347
C. philippinicus	Luzon 1	Babuyan Claro	no	2	2	3855	350,748	2620	4
C. philippinicus	Luzon 2	Camiguin Norte	no	3	4	15,519	1,412,286	10,184	35
C. philippinicus	Polillo	Luzon 3	yes	5	5	19,561	1,781,649	27,857	171
C. philippinicus	Panay	Negros	yes	3	2	8256	751,746	6536	20
C. philippinicus	Sibuyan	Tablas	?	3	3	21,426	1,951,966	14,010	54
G. crombota-rossi	Babuyan Claro	Calayan	no	5	5	16,901	1,538,408	5737	50
G. gigante	North Gigante	South Gigante	yes	4	3	17,393	1,583,712	4674	21
G. mindorensis	Lubang	Luzon	no	5	4	18,137	1,651,186	12,092	68
G. mindorensis	Masbate	Panay 1	yes	3	4	20,570	1,873,140	11,662	49
G. mindorensis	Negros	Panay 2	yes	3	5	17,636	1,605,943	6527	30
G. porosus	Sabtang	Batan	?	4	4	16,345	1,488,491	5378	31
G. romblon	Romblon	Tablas	yes	5	4	7074	643,155	5859	34
G. sp. B-sp. A	Camiguin Norte	Dalupiri	no	5	5	15,199	1,383,596	5612	31

2.2 Genomic library preparation and sequencing

We extracted DNA from tissue using the guanidine thiocyanate method described by Esselstyn et al. (2008). We diluted the extracted DNA for each individual to a concentration of 5 ng/μL based on the initial concentration measured with a Qubit 2.0 Fluorometer. We generated three restriction-site associated DNA sequence (RADseq) libraries, each with 96 individuals, using the multiplexed shotgun genotyping (MSG) protocol of Andolfatto et. al. (Andolfatto et al., 2011). Following digestion of 50 ng of DNA with the NdeI restriction enzyme, we ligated each sample to one of 96 adaptors with a unique six base-pair barcode. After pooling the 96 samples together, we selected 250–300bp fragments to remain in the library using a Pippen Prep. For each pool of 96 size-selected samples, we performed eight separate polymerase chain reactions for 14 cycles (PCR) using Phusion High-Fidelity PCR Master Mix (NEB Biolabs) and primers that bind to common regions in the adaptors. Following PCR, we did two rounds of AMPure XP bead cleanup (Beckman Coulter, Inc.) using a 0.8 bead volume to sample ratio. Each library was sequenced in one lane of an Illumina Hiseq 2500 high-output run, with single-end 100bp reads. We provide information on all of the individuals included in the three RADseq libraries in Table S1, a subset of which were included in the population pairs we analyzed for this study (Table 1 & S2). We deposited the demultiplexed, raw sequence reads into the NCBI Sequence Read Archive (Bioproject PRJNA486413, SRA Study SRP158258), and the assembled data matrices are available in our project repository (`https://github.com/phyletica/gekgo`).

2.3 Data assembly

We used ipyrad version 0.7.13 (Eaton, 2017) to demultiplex and assemble the raw RADseq reads into loci. To maximize the number of assembled loci, we de novo assembled the reads separately for each pair of populations. All of the scripts and ipyrad parameters files we used to assemble the data are available in our project repository (`https://github.com/phyletica/gekgo`), and the ipyrad settings are listed in Table S3.

2.4 Testing for shared divergences

We approach the inference of temporally clustered divergences as a problem of model choice. Our goal is to treat the number of divergence events shared (or not) among the pairs of populations, and the assignment of the pairs to those events, as random variables to be estimated from the aligned sequence data. For eight pairs, there are 4,140 possible divergence models. I.e., there are 4,140 ways to partition the eight pairs to $k = 1, 2, \ldots, 8$ divergence events (Bell, 1934; Oaks, 2014, 2018). Whereas divergences caused by sea-level rise would not happen simultaneously, we expect that on a timescale of the lizards' mutation rate, treating them as simultaneous should be a better explanation of data generated by such a process than treating them as independent.

Given the large number of models, and our goal of making probability statements about them, we used a Bayesian model-averaging approach. Specifically, we used the full-likelihood Bayesian comparative biogeography method implemented in the software package `ecoevolity` version 0.1.0 (commit b9f34c8) (Oaks, 2018). This method models each pair of populations

as a two-tipped "species" tree, with an unknown, constant population size along each of the three branches, and an unknown time of divergence. This method can directly estimate the likelihood of values of these unknown parameters from orthologous biallalic characters by analytically integrating over all possible gene trees and mutational histories (Bryant et al., 2012; Oaks, 2018). Within this full-likelihood framework, this method uses a Dirichlet process prior on the assignment of our pairs to an unknown number of divergence times. The Dirichlet process is specified by a (1) concentration parameter, α, which determines how probable it is for pairs to share the same divergence event, *a priori*, and (2) base distribution, which serves as the prior on the unique divergence times.

Importantly, because the pairs of populations are modeled as disconnected species trees, the relative rates of mutation among the pairs is not identifiable. This requires us to make informative prior assumptions about the relative rates of mutation among the pairs. Because *Cyrtodactylus* and *Gekko* are deeply divergent (> 80 mya; Gamble et al., 2011), and nothing is known about their relative rates of mutation, we analyzed the two genera separately. Within each genus, the populations are all closely related (Welton et al., 2010b,a; Siler et al., 2010, 2012, 2014) allowing us to make the simplifying assumption that the rates of mutation are equal across pairs *within* each genus.

Based on previous data (Welton et al., 2010b,a; Siler et al., 2010) we assumed a prior of Exponential(mean $= 0.005$) on divergence times for our eight pairs of *Cyrtodactylus* populations, in units of substitutions per site. To explore the sensitivity of our results to this assumption, we also tried a prior of Exponential(mean $= 0.05$) on the divergence times. Based on previous data (Siler et al., 2012, 2014), we assumed a prior of Exponential(mean $= 0.0005$) on divergence times for our eight pairs of *Gekko* populations, in units of substitutions per site. To explore the sensitivity of our results to this assumption, we also tried priors of Exponential(mean $= 0.005$) and Exponential(mean $= 0.05$) on the *Gekko* divergence times.

For the concentration parameter of the Dirichlet process, we assumed a hyperprior of Gamma$(1.1, 56.1)$ for both genera. This places approximately half of the prior probability on the model with no shared divergences ($k = 8$). By placing most of the prior probability on the model of independent divergences, if we find posterior support for shared divergences, we can be more confident it is being driven by the data, as opposed to the prior on divergence times penalizing additional divergence-time parameters (Jeffreys, 1939; Lindley, 1957; Oaks et al., 2013, 2014). To explore the sensitivity of our results to this assumption, we also tried a hyperprior of Gamma$(1.5, 3.13)$ and Gamma$(0.5, 1.31)$. The former corresponds with a prior mean number of divergence events of five, whereas the latter places 50% of the prior probability on the single divergence ($k = 1$) model.

For all analyses of both the *Cyrtodactylus* and *Gekko* data, we assumed equal mutation rates among the pairs, a prior distribution of Gamma(shape $= 4.0$, mean $= 0.004$) on the effective effective size of the populations scaled by the mutation rate ($N_e\mu$), and a prior distribution of Gamma(shape $= 100$, mean $= 1$) on the relative effective size of the ancestral population (relative to the mean size of the two descendant populations). The model implemented in `ecoevolity` assumes each character is unlinked (i.e., evolved along a gene tree that is independent conditional on the population tree). However, by analyzing simulated data, Oaks (2018) showed the method performs better when all linked sites are used than when data are excluded to avoid violating the assumption of unlinked sites. Accordingly, we analyzed all of the sites of our RADseq loci. The model implemented in `ecoevolity` is

also restricted to characters with two possible states (biallelic). Thus, for sites with three or more nucleotides (hereafter referred to as polyallelic sites), we compared how sensitive our results were to two different strategies: (1) Removing polyallelic sites, and (2) recoding the sites as biallelic by coding each state as either having the first nucleotide in the alignment or a different nucleotide. We assumed the biallelic equivalent of a Jukes-Cantor model of character substitution (Jukes and Cantor, 1969) so that our results are not sensitive to how nucleotides are coded as binary (Oaks, 2018).

For each analysis, we ran 10 independent MCMC chains for 150,000 generations, sampling every 100th generation. We assessed convergence and mixing of the chains by inspecting the potential scale reduction factor Brooks and Gelman (the square root of Equation 1.1 in 1998) and effective sample size (Gong and Flegal, 2016) of the log likelihood and all continuous parameters using the `pyco-sumchains` tool of `pycoevolity`. We also visually inspected the trace of the log likelihood and parameters over generations with the program Tracer version 1.6 (Rambaut et al., 2014).

3 Results

3.1 Data collection and MCMC convergence

Table 1 summarize the number of individuals sampled for each pair of islands, along with the number of assembled loci, and the number of total, variable, and polyallelic characters. The 10 independent MCMC chains of all our `ecoevolity` analyses appeared to have converged almost immediately. We conservatively removed the first 101 samples, leaving 1400 samples from each chain (14,000 samples for each analysis). With the first 101 samples removed, across all our analyses, all ESS values were greater than 2000, and all PSRF values were less than 1.005.

3.2 Testing for shared divergences

3.2.1 *Cyrtodactlus* population pairs

For *Cyrtodactylus*, our `ecoevolity` results support the model of no shared divergences, i.e., all eight pairs of populations diverged independently (Figures 2 & 3). This support is consistent across all three priors on the concentration parameter of the Dirichlet process (Figures S1 & S2). The support is also consistent across both priors on divergence times and whether polyallelic sites are recoded or removed (Figures S3 & S4). Estimates of effective population sizes are also very robust to priors on α and τ, and whether polyallelic sites are recoded or removed (Figures S5 & S6).

3.2.2 *Gekko* population pairs

For *Gekko*, posterior probabilities weakly support no shared divergences, but Bayes factors weakly support seven divergence events across the eight pairs (Figure 4), suggesting a possible shared divergence between *G. mindorensis* on the Islands of Panay and Masbate and *G. porosus* on the Islands of Sabtang and Batan (Figure 5). Under the intermediate prior on

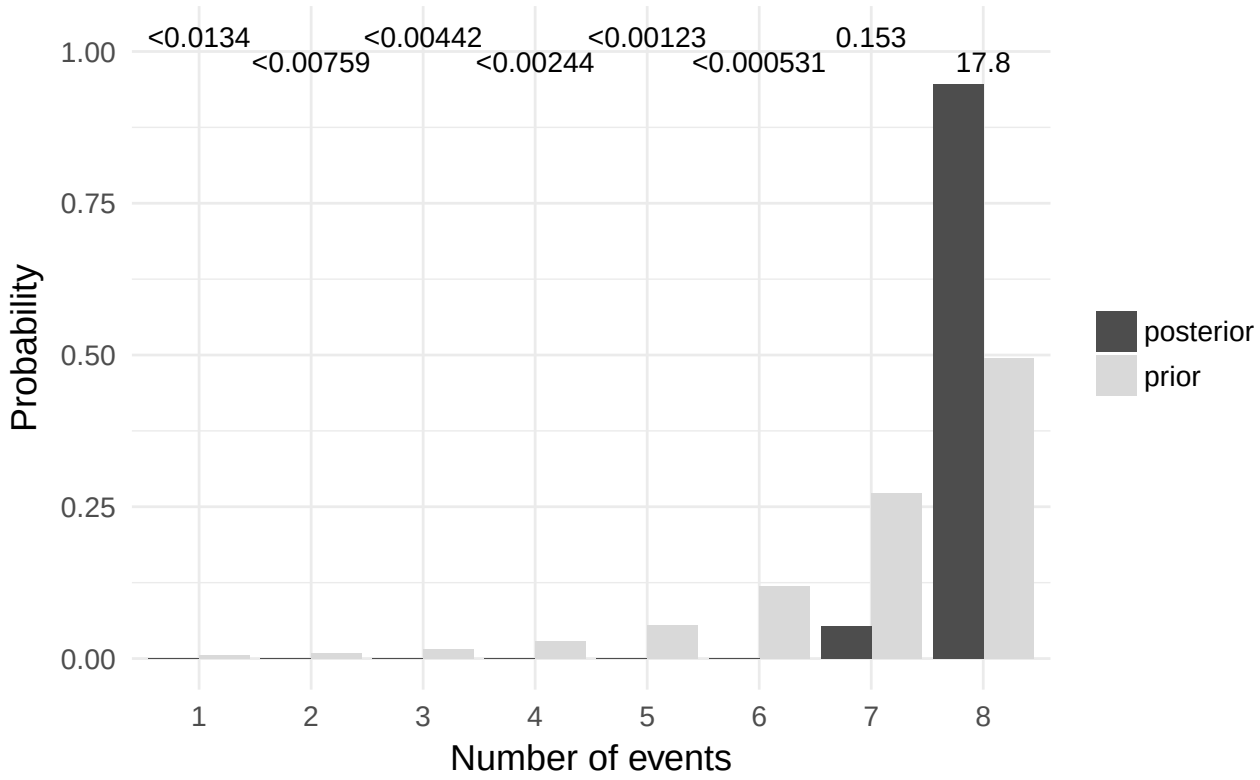

Figure 2. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

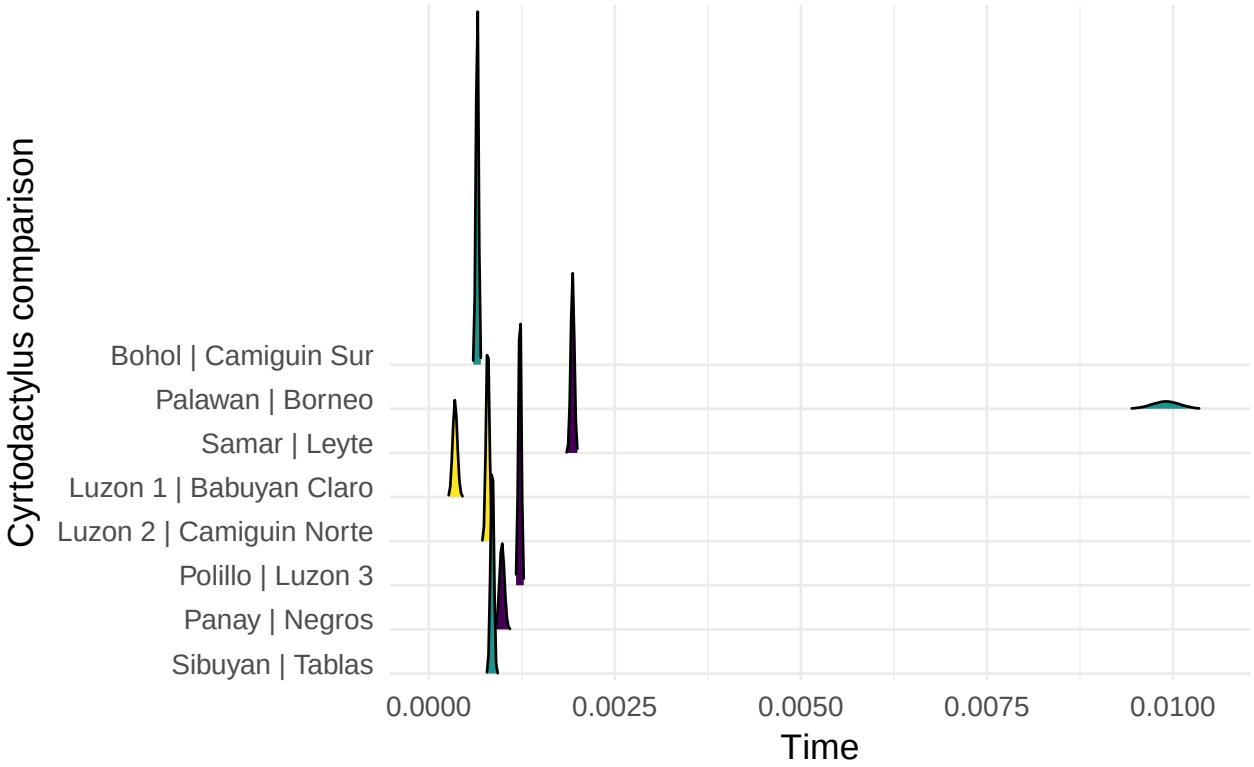

Figure 3. Approximate marginal posterior densities of divergence times for each pair of *Cyrtodactylus* populations. The density plot of each pair is color-coded to indicate whether the islands were connected during glacial periods (Fig. 1). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

the concentration parameter, support increases for this shared divergence increases (Figures S7 & S8). Under the prior that puts most of the probability on one shared event, posterior probabilities prefer six divergences (Figure S7) with another shared divergence between *G. mindorensis* on the Islands of Babuyan Claro and Calayan and *G. porosus* on the Islands of Romblon and Tablas (Figure S8), however, Bayes factors still prefer seven divergences. Similarly, as the prior on divergence times becomes more diffuse, the results shift from ambiguity between seven or eight divergence events, to ambiguity between six or seven events, to strong support for six events, with the same island pairs sharing divergences (Figures S9 & S10).

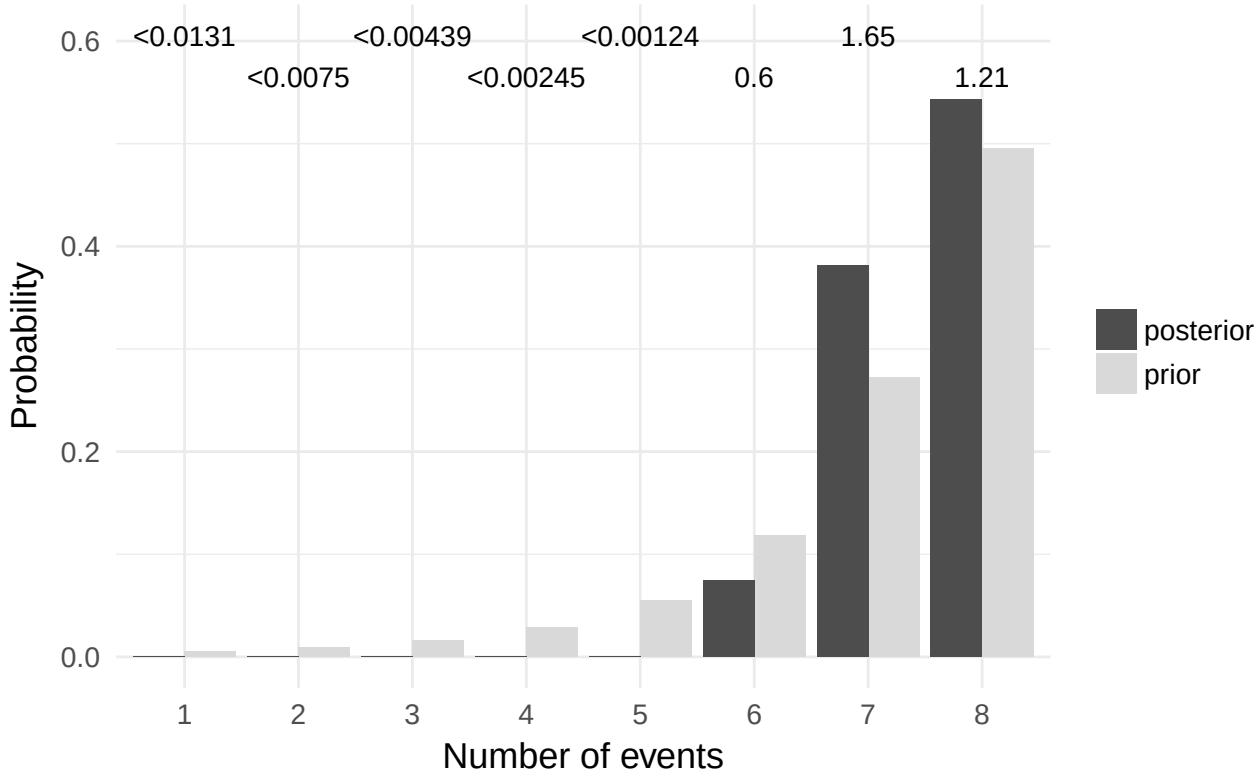

Figure 4. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

As with *Cyrtodactylus*, the estimates of divergence times are robust to whether polyallelic sites are recoded or removed (Figure S10), and population size estimates are robust to priors on α and τ, and whether polyallelic sites are recoded or removed (Figures S11 & S12).

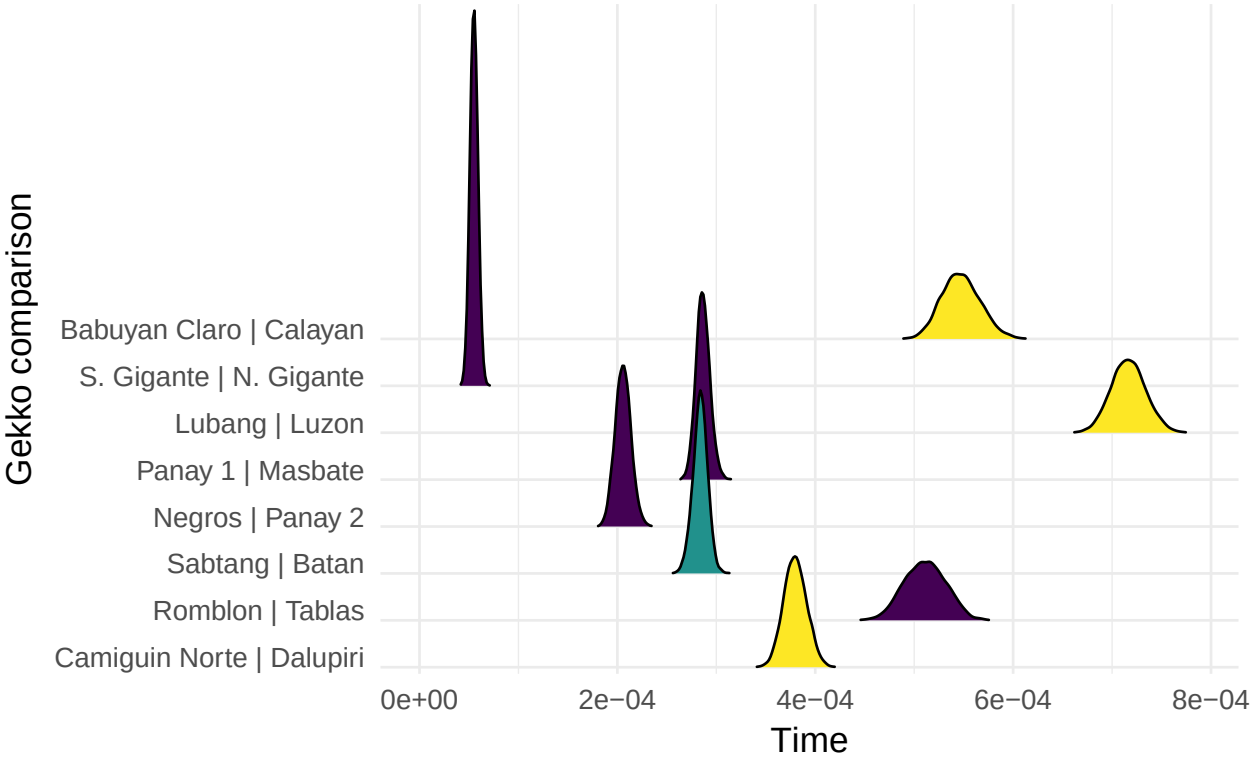

Figure 5. Approximate marginal posterior densities of divergence times for each pair of *Gekko* populations. The density plot of each pair is color-coded to indicate whether the islands were connected during glacial periods (Fig. 1). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

4 Discussion

4.1 Is there evidence for shared divergences among *Gekko* pairs?

Under some of the priors we explored, there is support for two possible shared divergences among the pairs of *Gekko* populations: (1) *G. mindorensis* on the Islands of Panay and Masbate and *G. porosus* on the Islands of Sabtang and Batan, and (2) *G. mindorensis* on the Islands of Babuyan Claro and Calayan and *G. porosus* on the Islands of Romblon and Tablas. The Islands of Babuyan Claro and Calayan were never connected, and we only inferred support for the second shared divergence under the most extreme priors on α and τ that are expected to favor shared divergences (Figures S7, S8, S9, & S10). Thus, support for the second shared divergence scenario is likely an artifact of prior sensitivity. However, the weak support for a shared divergence between *G. mindorensis* on the Islands of Panay and Masbate and *G. porosus* on the Islands of Sabtang and Batan under more reasonable priors is interesting because both pairs of islands were either connected or potentially close enough during glacial periods to allow gene flow.

Under the priors we initially chose as appropriate (as opposed to those used to assess prior sensitivity), the posterior probability that the Panay-Masbate and Sabtang-Batan pairs co-diverged is 0.385. To evaluate support for this co-divergence, we could calculate a Bayes factor using the prior probability that any two pairs share the same divergence time, which is approximately 1.66 in favor of the co-divergence (Figure 4). However, this would not be appropriate, because we did not identify the Panay-Masbate and Sabtang-Batan pairs of interest *a priori*, but rather our attention was drawn to these pairs based on the posterior results. Thus, the probability that *any* two pairs share the same divergence is no longer the appropriate prior probability for our Bayes factor calculation. Rather, we need to consider the prior probability that the two pairs with the most similar divergence times share the same divergence. To get this prior probability, we can take advantage of the fact that this condition is met anytime the number of divergence events is less than eight. Thus, the prior probability that the two pairs with the most similar divergence times share the same divergence is equal to one minus the prior probability that all eight pairs diverge independently. Under our prior on the concentration parameter of Gamma$(1.1, 56.1)$, this prior probability is approximately 0.5. Thus, our posterior probability for the co-divergence between the Panay-Masbate and Sabtang-Batan pairs is actually *less* than the prior probability, resulting in a weak Bayes factor of approximately 1.6 in support *against* the co-divergence. Thus, based on probability theory, we should favor the explanation that all eight pairs of *Gekko* populations diverged independently.

4.2 Implications for the PAIC diversification model

For each genus we 5–6 pairs of populations that span two islands that were connected (or nearly so) by terrestrial dry land bridges that formed during Pleistocene glacial periods. The connectivity between each pair of islands was likely fragmented by rising sea levels six or more times over the past three million years (Rohling et al., 1998; Siddall et al., 2003). Given that, for each genus, we have fewer pairs than the number of times the islands were fragmented, the support we found for independent divergence times among the pairs we

analyzed does not obviate all correlates of the PAIC diversification model; our pairs could have diverged at different fragmentation events. Comparative genomic data from more pairs of populations would be necessary to explore this possibility.

Another possibility is that some of our pairs of populations diverged during the same interglacial period, but the time when gene flow was cut off by rising sea levels was different enough to be estimated as separate divergences in `ecoevolity`. Based on estimates of rates of interglacial sea-level rise in Southeast Asia (Hanebuth et al., 2000; Sathiamurthy and Voris, 2006), it is unlikely that the timing of island fragmentation across the Philippines would have differed more than 5,000 years. If we assume a rate of mutation an order of magnitude faster than that estimated by Siler et al. (2012) for the phosducin gene of Philippine *Gekko* (1.18×10^{-9} substitutions per site per year), we would not expect to see a difference in divergence times greater than 5.9×10^{-6} between two pairs that diverged during the same interglacial. It seems reasonable to assume that the difference in divergence times between *G. mindorensis* on the Islands of Panay and Masbate and *G. porosus* on the Islands of Sabtang and Batan is close to the minimum resolution of `ecoevolity` given our data; there is little posterior variance in the divergence times for these pairs (Figure 5), and we note posterior uncertainty in whether these pairs co-diverged or not (Figure 4). The posterior mean absolute difference in divergence time between these pairs, conditional on them not co-diverging, is 9.66×10^{-6}. These numbers are similar enough to suggest that it might be possible that `ecoevolity` would have the temporal resolution given our data to distinguish the divergence times of two pairs that diverged during the same interglacial fragmentation event. However, it does not seem very likely, especially when we consider the strong support inferred for distinct divergences among all but two of the pairs of populations analyzed.

4.3 Sensitivity to the prior on divergence times

It is interesting that in analyses of both genera we see support for shared divergences increase as the prior on divergence times becomes more diffuse (Figures S3 & S9). While less extreme here, this is the same pattern seen in approximate-likelihood Bayesian approaches to this problem (Oaks et al., 2013; Hickerson et al., 2014; Oaks et al., 2014). Hickerson et al. (2014) proposed this pattern was caused by numerical problems, whereas Oaks et al. (2014) interpreted the problem as being more fundamental: as more prior density is placed in regions of divergence-time space where the likelihood tends to be low, models that have fewer divergence-time parameters have greater marginal likelihoods because their likelihoods are "averaged" over less space with low likelihood and substantial prior weight. Our results clearly support the latter explanation, as the MCMC approach used here does not suffer from the insufficient prior sampling proposed by Hickerson et al. (2014).

Whereas the full-likelihood Bayesian approach used here is much more robust to prior assumptions than the ABC approaches, our results demonstrate that is still important to assess sensitivity of the results to the priors (Oaks et al., 2013). This is especially true for the posterior probabilities of divergence models or the number of divergence events. The marginal likelihoods of the divergence models are what update our prior probabilities to give use these posterior probabilities, and the marginal likelihood is averaged over the entire parameter space of a model, and weighted by the prior. As a result, they can be sensitive to the priors regardless of the informativeness the data (Oaks et al., 2018).

4.4 Conclusions

Climate-driven fragmentation of the Philippine Islands has been invoked as a model of pulsed co-speciation throughout the archipelago. This model predicts that population divergences between fragmented islands should be temporally clustered around interglacial rises in sea levels. We analyzed comparative genomic data from 16 pairs of insular gecko populations within a full-likelihood, Bayesian model-choice framework to test for shared divergence events. Our results support independent divergences among the pairs of gecko populations. While comparative genomic data from more taxa will allow us to address additional questions, our results suggest the repeated cycles of climate-driven island fragmentation has not been an important shared mechanism of speciation for gekkonid lizards in the Philippines.

5 Data Accessibility

All of the demultiplexed, raw sequence reads are available from the NCBI Sequence Read Archive (Bioproject PRJNA486413, SRA Study SRP158258). A detailed history of all aspects of this project was recorded in a version-controlled repository, which is publicly available at https://github.com/phyletica/gekgo.

6 Acknowledgments

The computation work was made possible by Auburn University (AU) Hopper Cluster supported by the AU Office of Information Technology. This work was possible thanks to funding provided to JRO from the National Science Foundation (DBI 1308885 and DEB 1656004). RMB's research has been supported by NSF DEB 0073199, 0743491, 0640737, 1418895, and 1654388; and the University of Kansas Biodiversity Institute. This paper is contribution number XXXX of the Auburn University Museum of Natural History.

References

Andolfatto, P., D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, and D. L. Stern, 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research 21:610–617.

Bell, E. T., 1934. Exponential numbers. American Mathematical Monthly 41:411–419.

Blackburn, D. C., D. P. Bickford, A. C. Diesmos, D. T. Iskandar, and R. M. Brown, 2010. An ancient origin for the enigmatic flat-headed frogs (Bombinatoridae: *Barbourula*) from the Islands of Southeast Asia. PLoS ONE 5:10.

Brooks, S. P. and A. Gelman, 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Brown, R. M. and A. C. Diesmos, 2002. Application of lineage-based species concept to oceanic island frog populations: The effects of differing taxonomic philosophies on the estimation of Philippine biodiversity. Silliman Journal 42:133–162.

———, 2009. Philippines, biology. Pp. 723–732, *in* R. Gillespie and D. Clague, eds. Encyclopdia of Islands. University of California Press, Berkeley.

Brown, R. M. and S. I. Guttman, 2002. Phylogenetic systematics of the *Rana signata* complex of Philippine and Bornean stream frogs: reconsideration of Huxley's modification of Wallace's line at the Oriental-Australian faunal zone interface. Biological Journal of the Linnean Society 76:393–461.

Brown, R. M. and C. D. Siler, 2014. Spotted stream frog diversification at the Australasian faunal zone interface, mainland versus island comparisons, and a test of the Philippine 'dual-umbilicus' hypothesis. Journal of Biogeography 41:182–195.

Brown, R. M., C. D. Siler, C. H. Oliveros, J. A. Esselstyn, A. C. Diesmos, P. A. Hosner, C. W. Linkem, A. J. Barley, J. R. Oaks, M. B. Sanguila, L. J. Welton, R. G. Moyle, A. T. Peterson, and A. C. Alcala, 2013. Evolutionary processes of diversification in a model island archipelago. Annual Review of Ecology, Evolution, and Systematics 44:411–435.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. Roychoudhury, 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29:1917–1932.

Catibog-Sinha, C. S. and L. R. Heaney (eds.) 2006. Philippine Biodiversity: Principles and Practice. Haribon Foundation, Quezon City, Philippines.

Eaton, D., 2017. ipyrad version 0.7.13. http://ipyrad.readthedocs.io.

Esselstyn, J. A. and R. M. Brown, 2009. The role of repeated sea-level fluctuations in the generation of shrew (Soricidae: *Crocidura*) diversity in the Philippine Archipelago. Molecular Phylogenetics and Evolution 53:171–181.

Esselstyn, J. A., H. J. Garcia, M. G. Saulog, and L. R. Heaney, 2008. A new species of *Desmalopex* (Pteropodidae) from the Philippines, with a phylogenetc analysis of the Pteropodini. Journal of Mammalogy 89:815–825.

Evans, B., R. Brown, J. Mcguire, J. Supriatna, N. Andayani, A. Diesmos, D. Iskandar, D. Melnick, and D. Cannatella, 2003. Phylogenetics of fanged frogs: Testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. Systematic Biology 52:794–819.

Gamble, T., A. M. Bauer, G. R. Colli, E. Greenbaum, T. R. Jackman, L. J. Vitt, and A. M. Simons, 2011. Coming to America: multiple origins of new world geckos. Journal of Evolutionary Biology 24:231–244.

Gillespie, R. G., 2007. Oceanic islands: Models of diversity. Pp. 223–241, *in* S. A. Levin, ed. Encyclopedia of Biodiversity. Academic Press, Oxford, UK.

Gong, L. and J. M. Flegal, 2016. A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. Journal of Computational and Graphical Statistics 25:684–700.

Hanebuth, T., K. Stattegger, and P. M. Grootes, 2000. Rapid flooding of the Sunda Shelf: A late-glacial sea-level record. Science 288:1033–1035.

Haq, B. U., J. Hardenbol, and P. R. Vail, 1987. Chronology of fluctuating sea levels since the Triassic. Science 235:1156–1167.

Heaney, L. R., 1985. Zoogeographic evidence for middle and late pleistocene land bridges to the philippine islands. Mod Quatern Res SE Asia 9:127–144.

Heaney, L. R., D. S. Balete, E. A. Rickart, P. A. Alviola, M. R. M. Duya, M. V. Duya, M. J. Veluz, L. VandeVrede, and S. J. Steppan, 2011. Chapter 1: Seven new species and a new subgenus of forest mice (Rodentia: Muridae: *Apomys*) from Luzon Island. Fieldiana Life and Earth Sciences 2:1–60.

Heaney, L. R. and J. C. Regalado, Jr., 1998. Vanishing treasures of the Philippine rain forest. Field Museum, Chicago, Illinois.

Heaney, L. R., J. S. Walsh, and A. T. Peterson, 2005. The roles of geological history and colonization abilities in genetic differentiation between mammalian populations in the Philippine Archipelago. Journal of Biogeography 32:229–247.

Hickerson, M. J., G. N. Stone, K. Lohse, T. C. Demos, X. Xie, C. Landerer, and N. Takebayashi, 2014. Recommendations for using msbayes to incorporate uncertainty in selecting an ABC model prior: A response to Oaks et al. Evolution 68:284–294.

Huxley, T. H., 1868. On the classification and the distribution of the Alectoromorphae and Heteromorphae. Proceedings of the Zoological Society of London 6:249–319.

Inger, R. F., 1954. Systematics and zoogeography of Philippine Amphibia. Fieldiana 33:182–531.

Jansa, S. A., F. K. Barker, and L. R. Heaney, 2006. The pattern and timing of diversification of Philippine endemic rodents: Evidence from mitochondrial and nuclear gene sequences. Systematic Biology 55:73–88.

Jeffreys, H., 1939. Theory of Probability. 1st ed. Clarendon Press, Oxford, U.K.

Jukes, T. H. and C. R. Cantor, 1969. Evolution of protein molecules. chap. 24, Pp. 21–132, *in* H. N. Munro, ed. Mammalian Protein Metabolism, vol. III. Academic Press, New York.

Lindley, D. V., 1957. A statistical paradox. Biometrika 44:187–192.

Linkem, C. W., A. C. Diesmos, and R. M. Brown, 2011. Molecular systematics of the Philippine forest skinks (Squamata: Scincidae: *Sphenomorphus*): testing morphological hypotheses of interspecific relationships. Zoological Journal of the Linnean Society 163:1217–1243.

Linkem, C. W., K. M. Hesed, A. C. Diesmos, and R. M. Brown, 2010. Species boundaries and cryptic lineage diversity in a Philippine forest skink complex (Reptilia; Squamata; Scincidae: Lygosominae). Molecular Phylogenetics and Evolution 56:572–585.

Lomolino, M. V., B. R. Riddle, and J. H. Brown, 2010. Reconstructing the History of Lineages. chap. 11, Pp. 423–456, *in* Biogeography, 4th ed. Sinauer Associates, Sunderland, Massachusetts, USA.

Losos, J. B. and R. E. Ricklefs, 2009. Adaptation and diversification on islands. Nature 457:830–836.

Miller, K. G., M. A. Kominz, J. V. Browning, J. D. Wright, G. S. Mountain, M. E. Katz, P. J. Sugarman, B. S. Cramer, N. Christie-Blick, and S. F. Pekar, 2005. The Phanerozoic record of global sea-level change. Science 310:1293–1298.

Oaks, J. R., 2014. An improved approximate-bayesian model-choice method for estimating shared evolutionary history. BMC Evolutionary Biology 14:150.

———, 2018. Full Bayesian comparative phylogeography from genomic data. bioRxiv .

Oaks, J. R., K. A. Cobb, V. N. Minin, and A. D. Leaché, 2018. Marginal likelihoods in phylogenetics: a review of methods and applications. arXiv:1805.04072 [q-bio.PE] .

Oaks, J. R., C. W. Linkem, and J. Sukumaran, 2014. Implications of uniformly distributed, empirically informed priors for phylogeographical model selection: A reply to hickerson et al. Evolution 68:3607–3617.

Oaks, J. R., J. Sukumaran, J. A. Esselstyn, C. W. Linkem, C. D. Siler, M. T. Holder, and R. M. Brown, 2013. Evidence for climate-driven diversification? a caution for interpreting ABC inferences of simultaneous historical events. Evolution 67:991–1010.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond, 2014. Tracer version 1.6. http://tree.bio.ed.ac.uk/software/tracer/.

Roberts, T. E., 2006. Multiple levels of allopatric divergence in the endemic Philippine fruit bat *Haplonycteris fischeri* (Pteropodidae). Biological Journal of the Linnean Society 88:329–349.

Rohling, E. J., M. Fenton, F. J. Jorissen, P. Bertrand, G. Ganssen, and J. P. Caulet, 1998. Magnitudes of sea-level lowstands of the past 500,000 years. Nature 394:162–165.

Sathiamurthy, E. and H. K. Voris, 2006. Maps of Holocene sea level transgression and submerged lakes on the Sunda Shelf. Natural History Journal of Chulalongkorn University Supplement 2:1–43.

Siddall, M., E. J. Rohling, A. Almogi-Labin, C. Hemleben, D. Meischner, I. Schmelzer, and D. A. Smeed, 2003. Sea-level fluctuations during the last glacial cycle. Nature 423:853–858.

Siler, C. D., A. C. Diesmos, A. C. Alcala, and R. M. Brown, 2011. Phylogeny of Philippine slender skinks (Scincidae: *Brachymeles*) reveals underestimated species diversity, complex biogeographical relationships, and cryptic patterns of lineage diversification. Molecular Phylogenetics and Evolution 59:53–65.

Siler, C. D., J. R. Oaks, K. Cobb, O. Hidetoshi, and R. M. Brown, 2014. Critically endangered island endemic or peripheral population of a widespread species? conservation genetics of Kikuchi's gecko and the global challenge of protecting peripheral oceanic island endemic vertebrates. Diversity and Distributions 20:756–772.

Siler, C. D., J. R. Oaks, J. A. Esselstyn, A. C. Diesmos, and R. M. Brown, 2010. Phylogeny and biogeography of Philippine bent-toed geckos (Gekkonidae: *Cyrtodactylus*) contradict a prevailing model of Pleistocene diversification. Molecular Phylogenetics and Evolution 55:699–710.

Siler, C. D., J. R. Oaks, L. J. Welton, C. W. Linkem, J. Swab, A. C. Diesmos, and R. M. Brown, 2012. Did geckos ride the Palawan raft to the Philippines? Journal of Biogeography 39:1217–1234.

Vences, M., K. C. Wollenberg, D. R. Vieites, and D. C. Lees, 2009. Madagascar as a model region of species diversification. Trends in Ecology & Evolution 24:456–465.

Wallace, A. R., 1869. The Malay Archpelago: The Land of the Orang-utan, and the Bird of Paradise. Macmillan and Co., London.

Welton, L. J., C. D. Siler, A. C. Diesmos, and R. M. Brown, 2010a. Phylogeny-based species delimitation of southern Philippine bent-toed geckos and a new species of *Cyrtodactylus* (squamata:gekkonidae) from western mindanao and the sulu archipelago. Zootaxa 2390:49–68.

Welton, L. J., C. D. Siler, C. W. Linkem, A. C. Diesmos, and R. M. Brown, 2010b. Philippine bent-toed geckos of the *Cyrtodactylus agusanensis* complex: Multilocus phylogeny, morphological diversity, and descriptions of three new species. Herpetological Monographs 24:55–85.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wilke, C. O., 2018. ggridges version 0.4.1. https://CRAN.R-project.org/package=ggridges.

Supporting Information

Title: The comparative biogeography of geckos challenges predictions from a paradigm of climate-driven vicariant diversification across an island archipelago

Authors: Jamie R. Oaks Corresponding author: joaks@auburn.edu[1], Cameron D. Siler[2], and Rafe M. Brown[3]

[1]Department of Biological Sciences & Museum of Natural History, Auburn University, Auburn, Alabama 36849, USA
[2]Sam Noble Oklahoma Museum of Natural History and Department of Biology, University of Oklahoma, Norman, Oklahoma 73072
[3]Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045, USA

Table S1. The data for all samples included in the three RADseq libraries are included in a separate tab-delimited text file (https://raw.githubusercontent.com/phyletica/gekgo/master/tex/tables/msg-samples-edited.txt).

Table S2. The data for all samples included in the 16 pairs of populations analyzed in this study are included in a separate tab-delimited text file (https://raw.githubusercontent.com/phyletica/gekgo/master/tex/manuscripts/ecoevolity/tables/comparison-samples-edited.txt). This is a subset of the data in Table 1.

Table S3. Settings used for assembling loci for each pair of gekkonid populations.

ipyrad setting	**Value**
assembly_method	denovo
datatype	rad
restriction_overhang	TATG,
max_low_qual_bases	4
phred_Qscore_offset	33
mindepth_statistical	6
mindepth_majrule	6
maxdepth	10000
clust_threshold	0.85
max_barcode_mismatch	0
filter_adapters	1
filter_min_trim_len	35
max_alleles_consens	2
max_Ns_consens	4
max_Hs_consens	5
min_samples_locus	4
max_SNPs_locus	20
max_Indels_locus	8
max_shared_Hs_locus	0.5
trim_reads	0, 0, 0, 0
trim_loci	0, 0, 0, 0

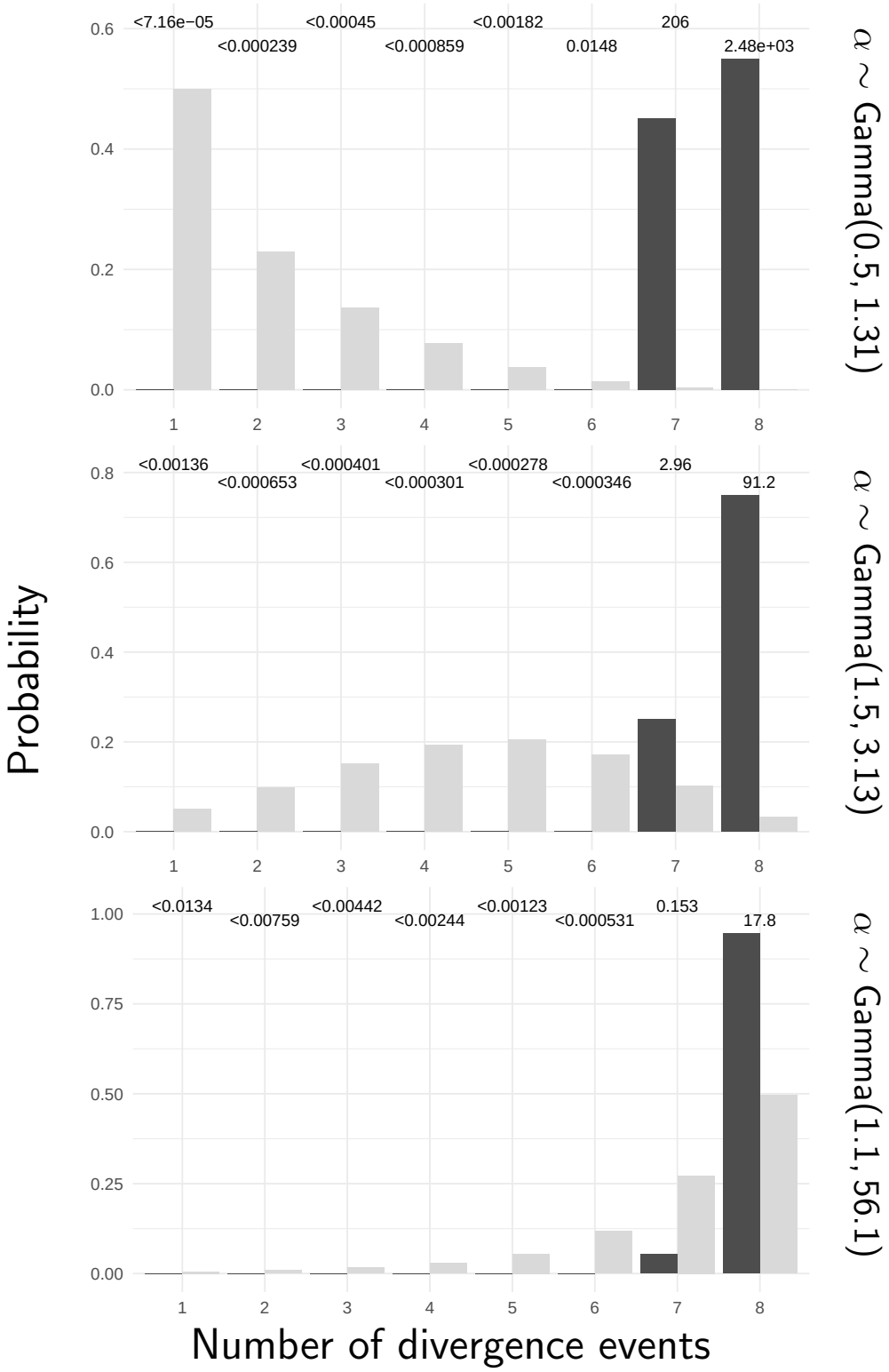

Figure S1. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

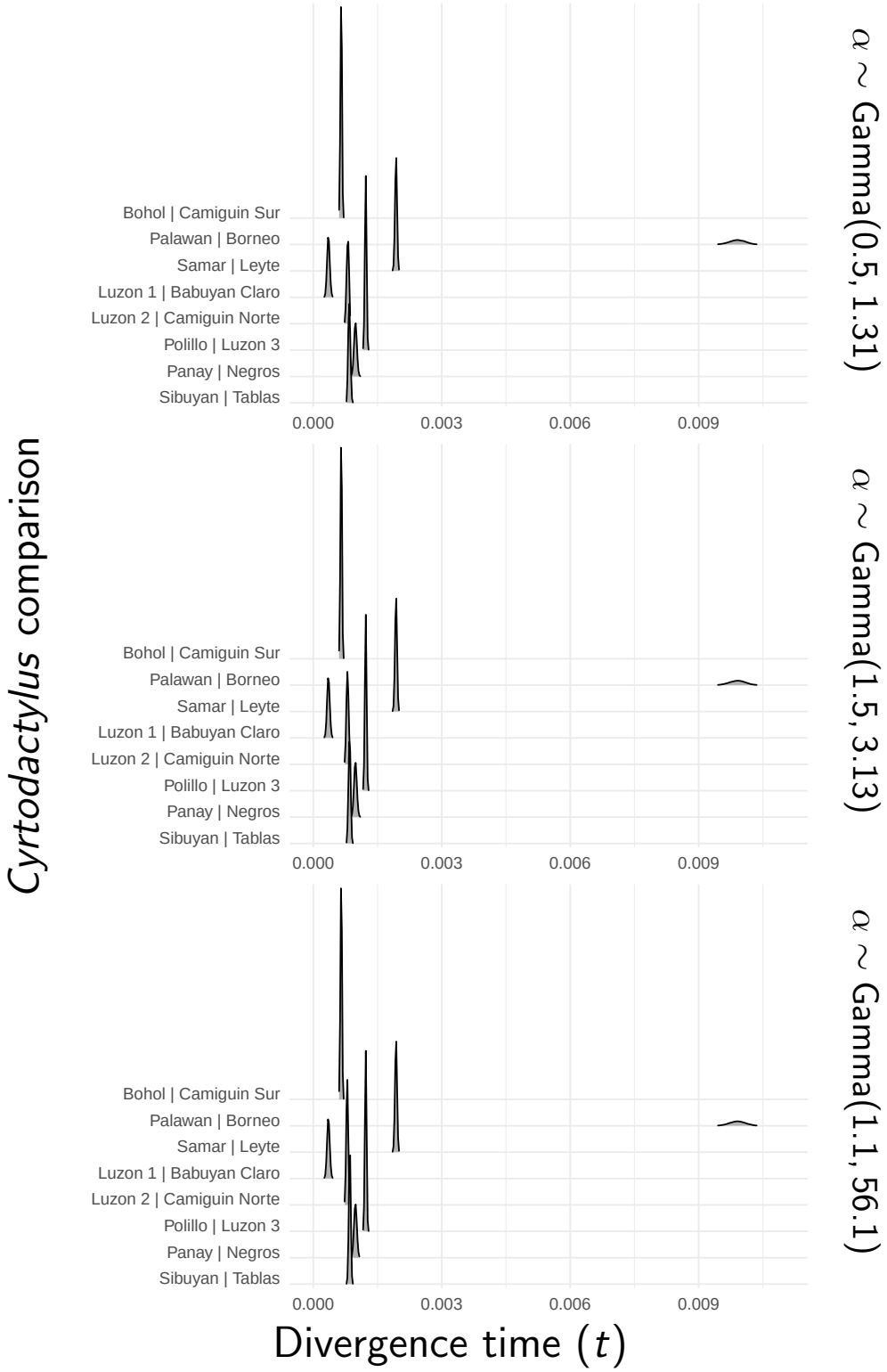

Figure S2. Approximate marginal posterior densities of divergence times for each pair of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

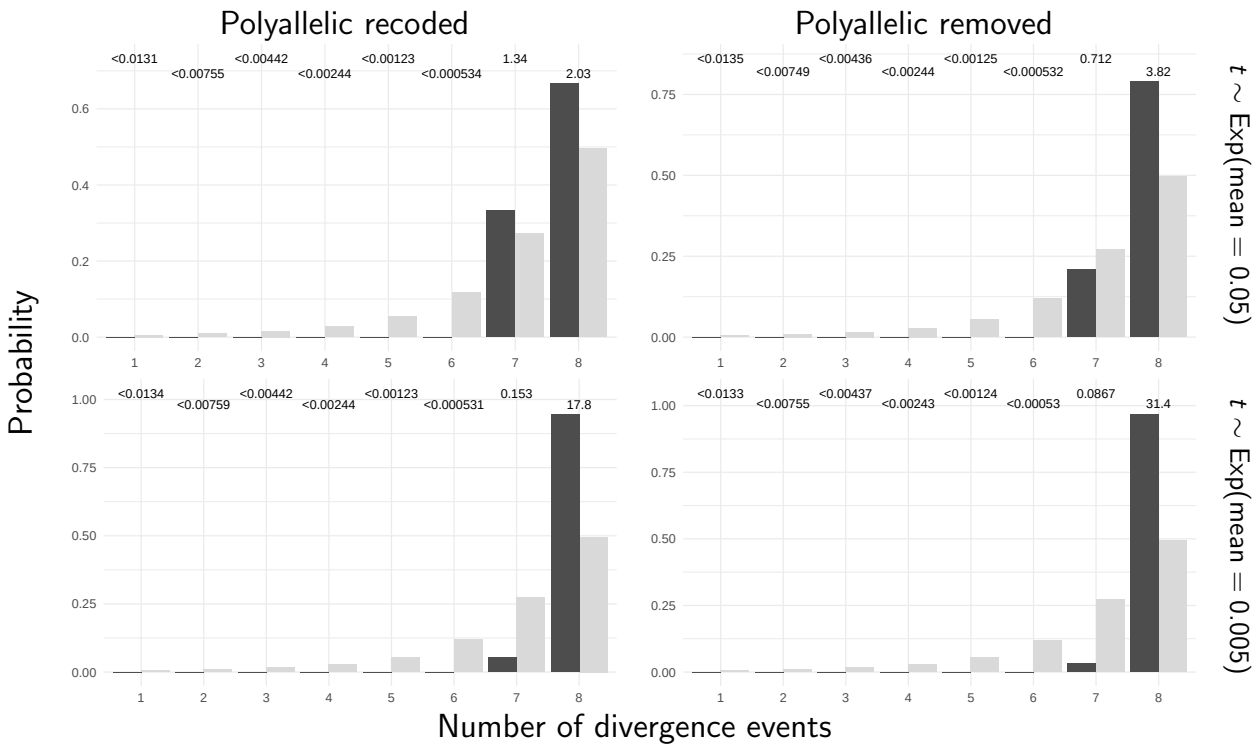

Figure S3. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

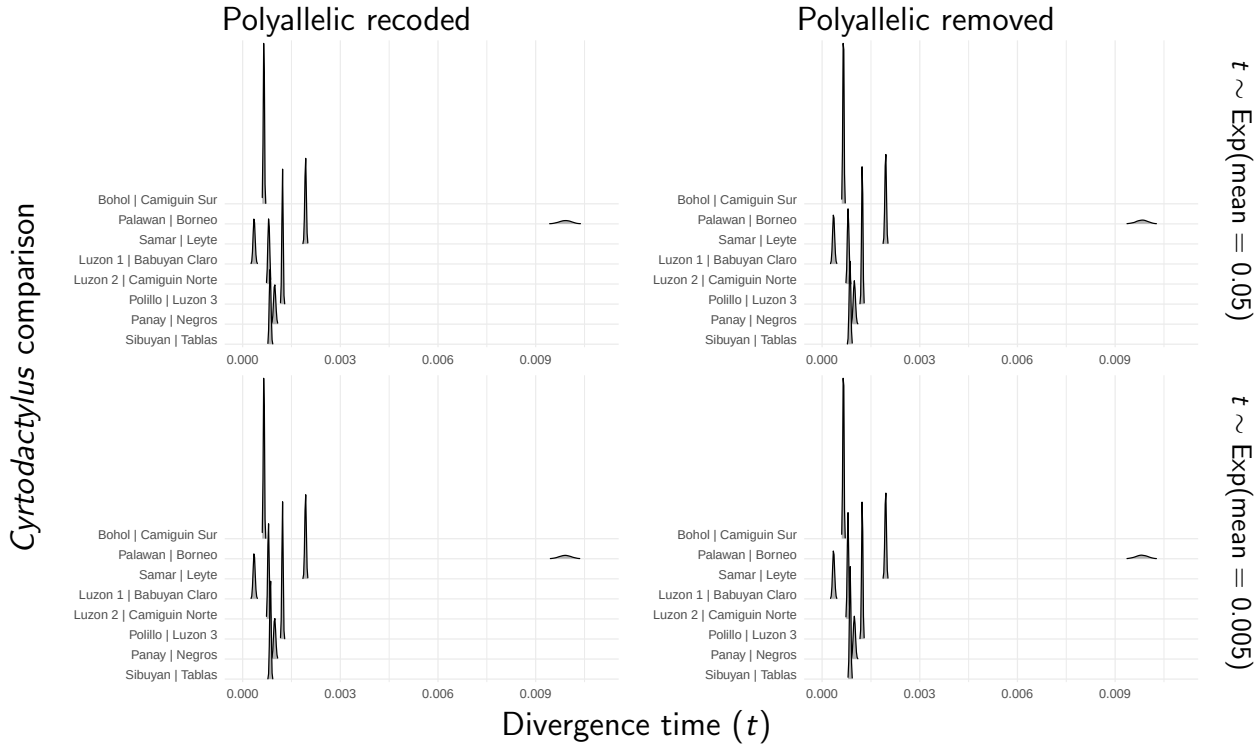

Figure S4. Approximate marginal posterior densities of divergence times for each pair of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

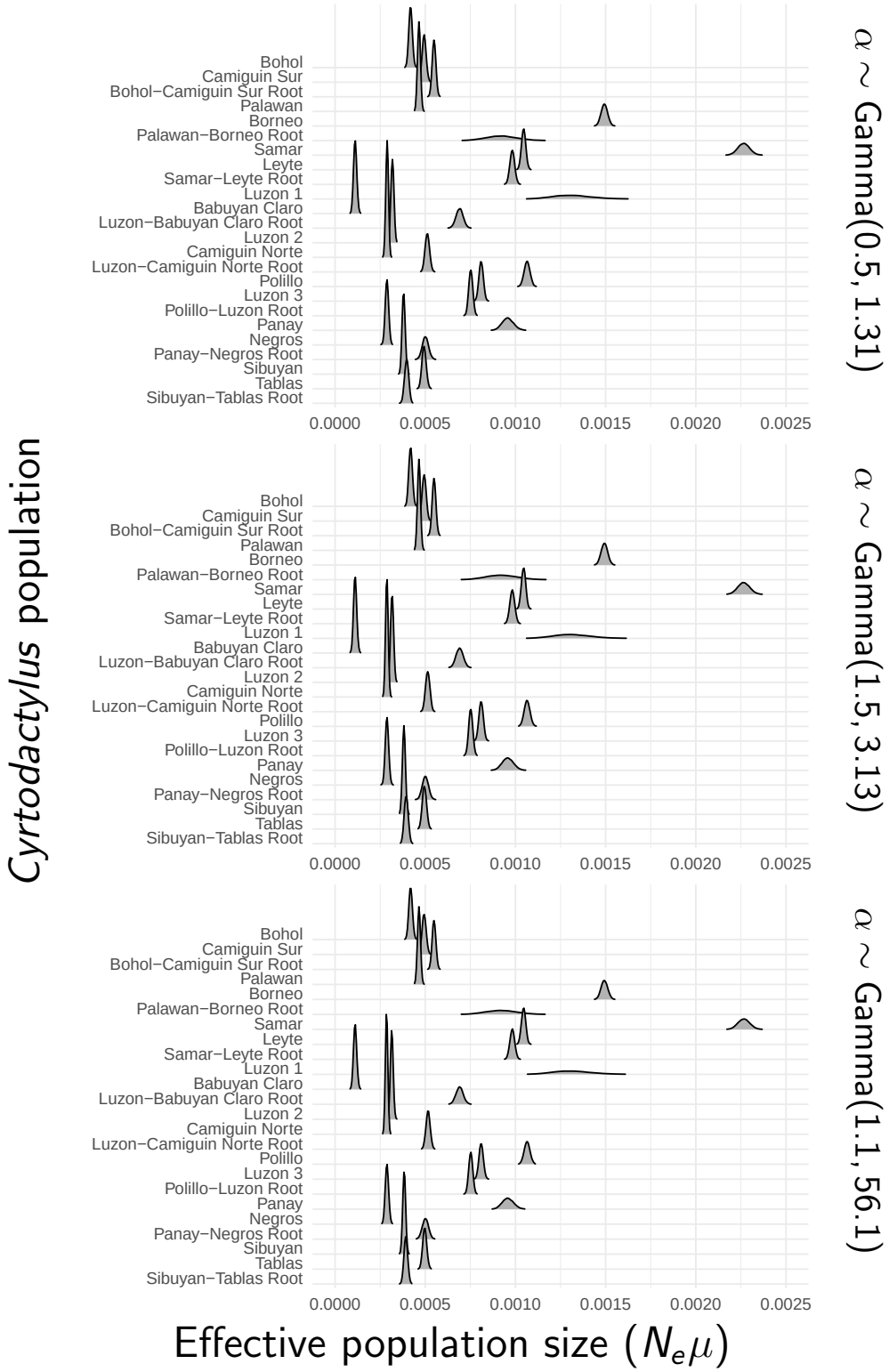

Figure S5. Approximate marginal posterior densities of population sizes for each pair of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

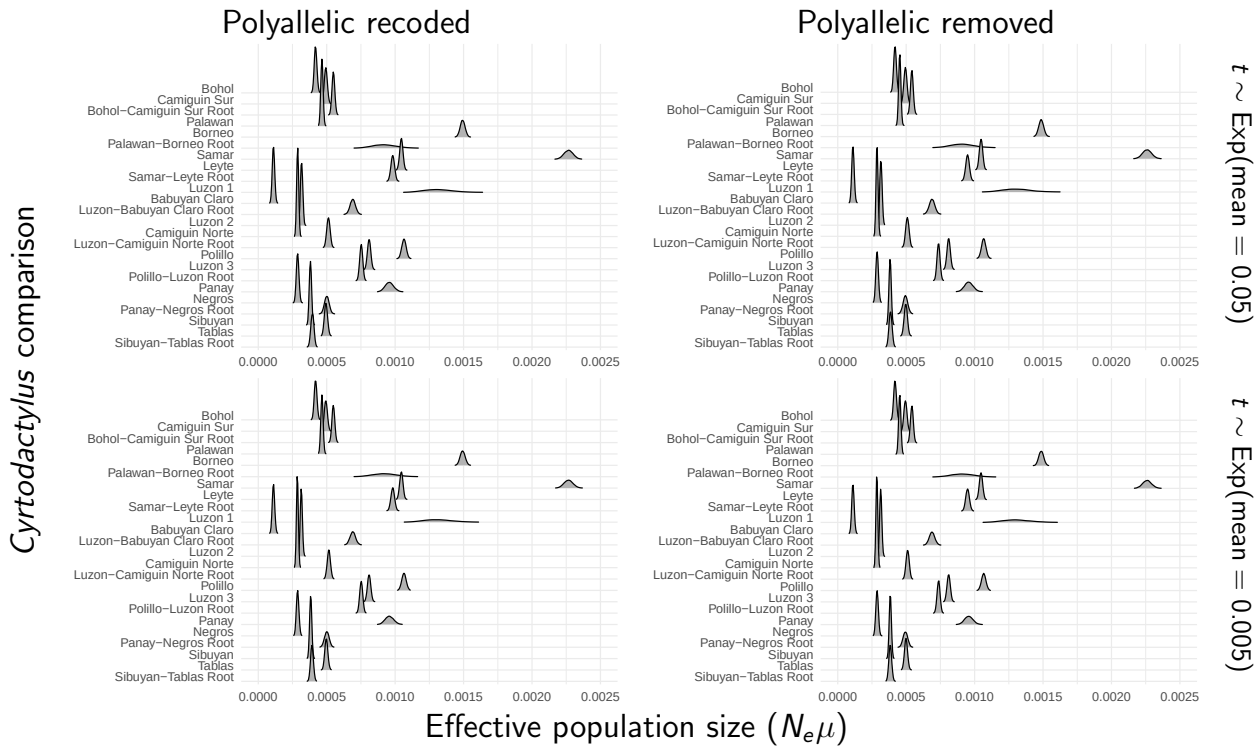

Figure S6. Approximate marginal posterior densities of population sizes for each pair of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

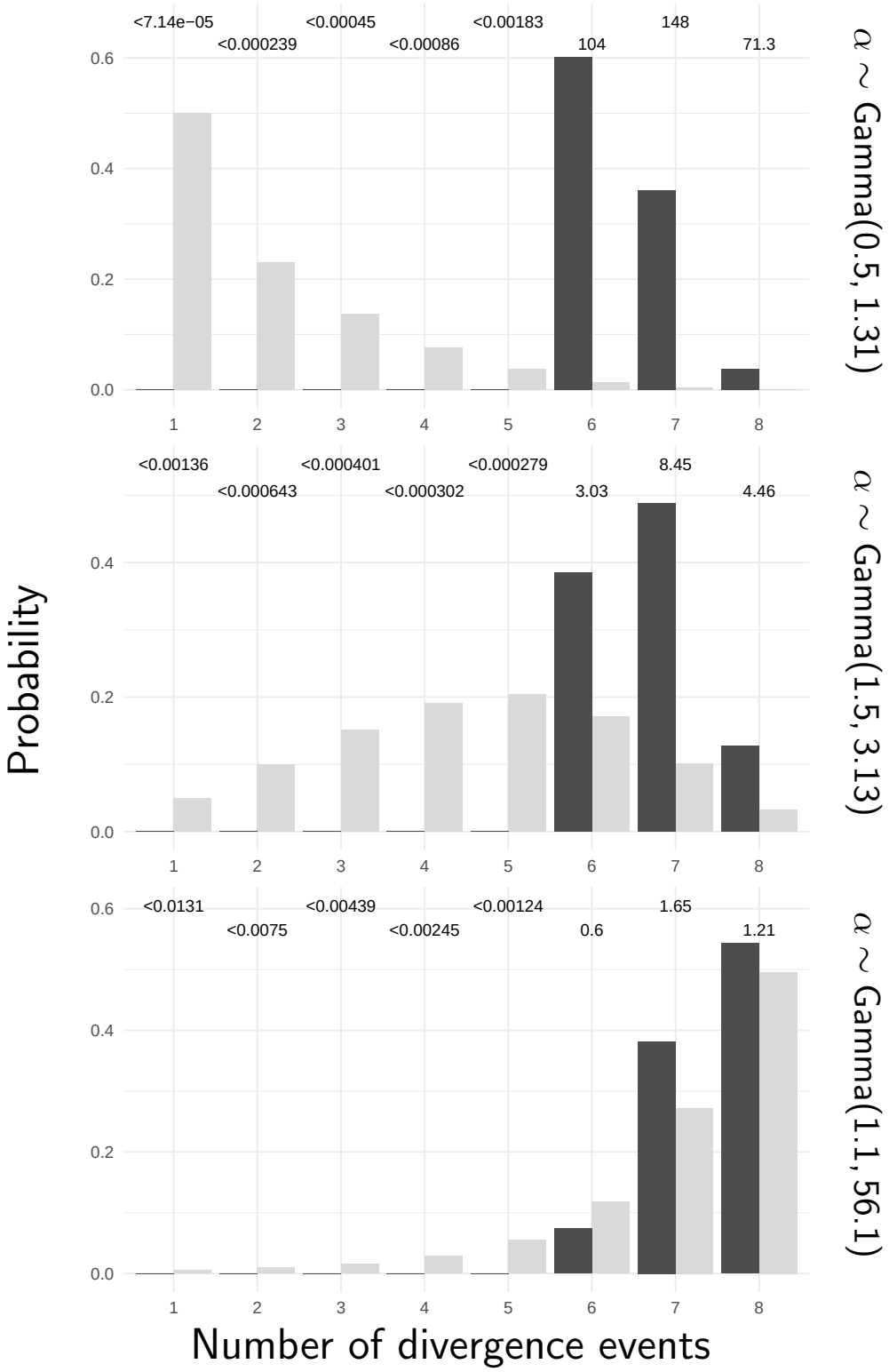

Figure S7. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

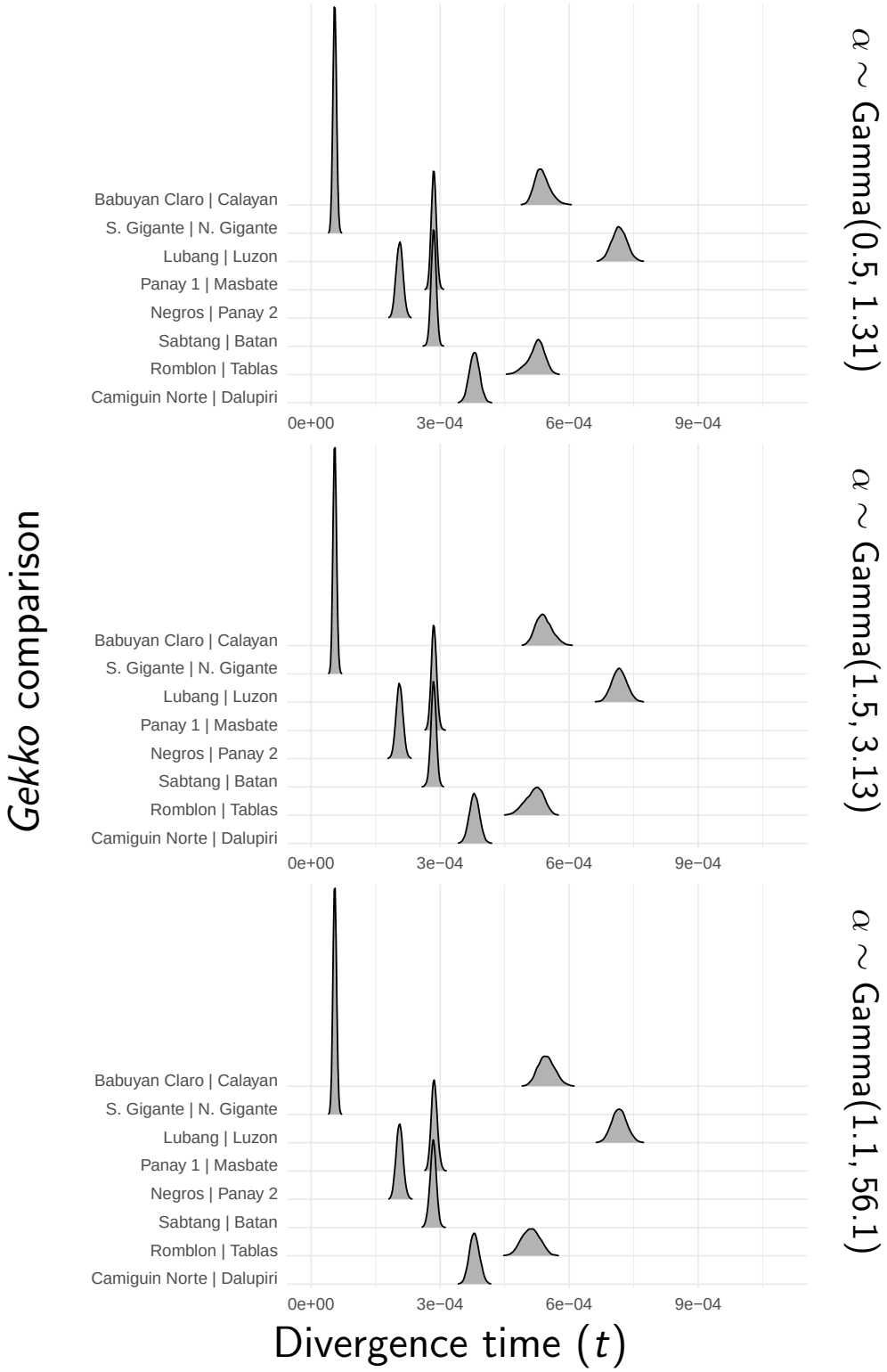

Figure S8. Approximate marginal posterior densities of divergence times for each pair of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

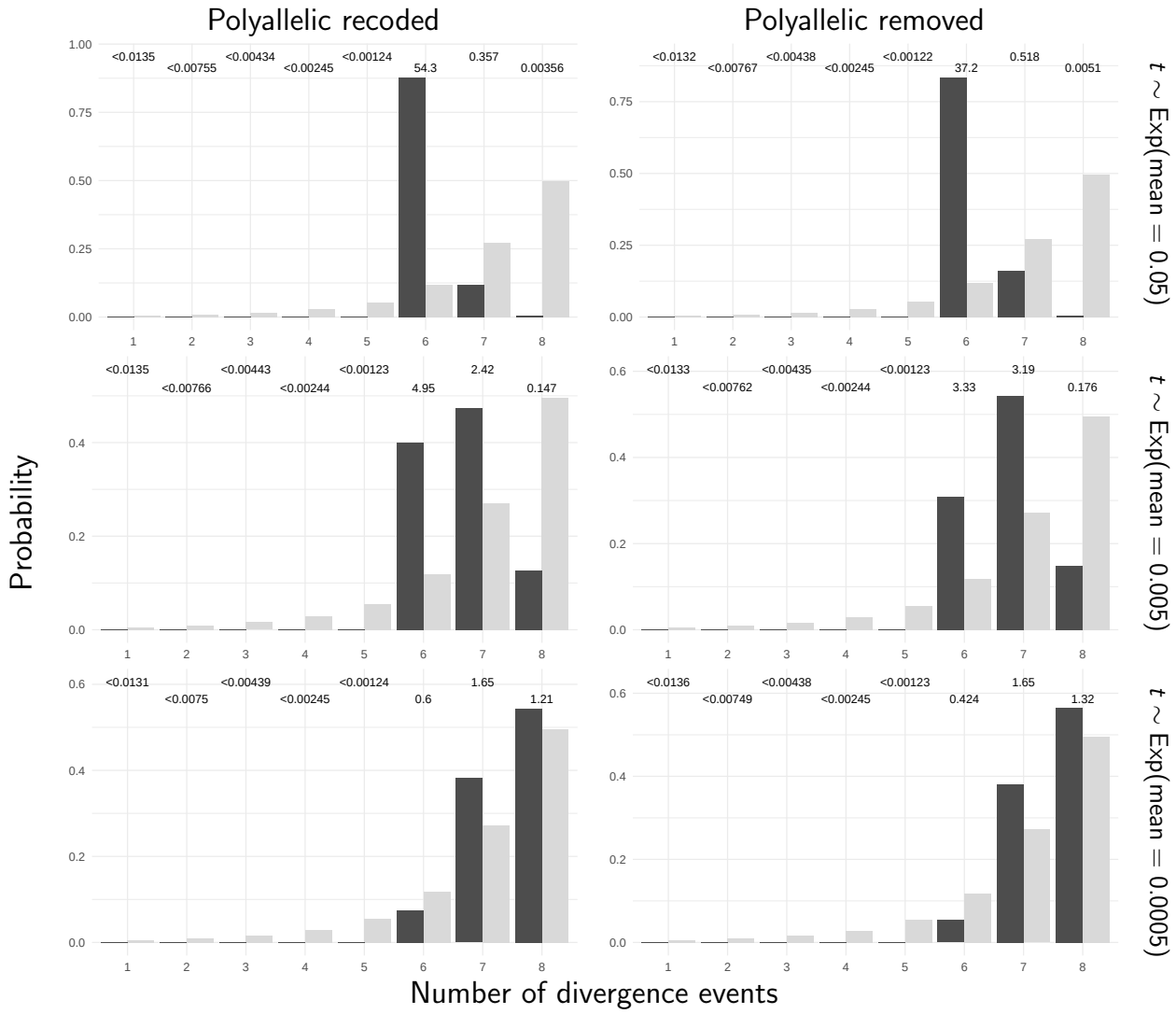

Figure S9. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events to all other possible numbers of divergence events. Figure generated with ggplot2 Version 2.2.1 (Wickham, 2009).

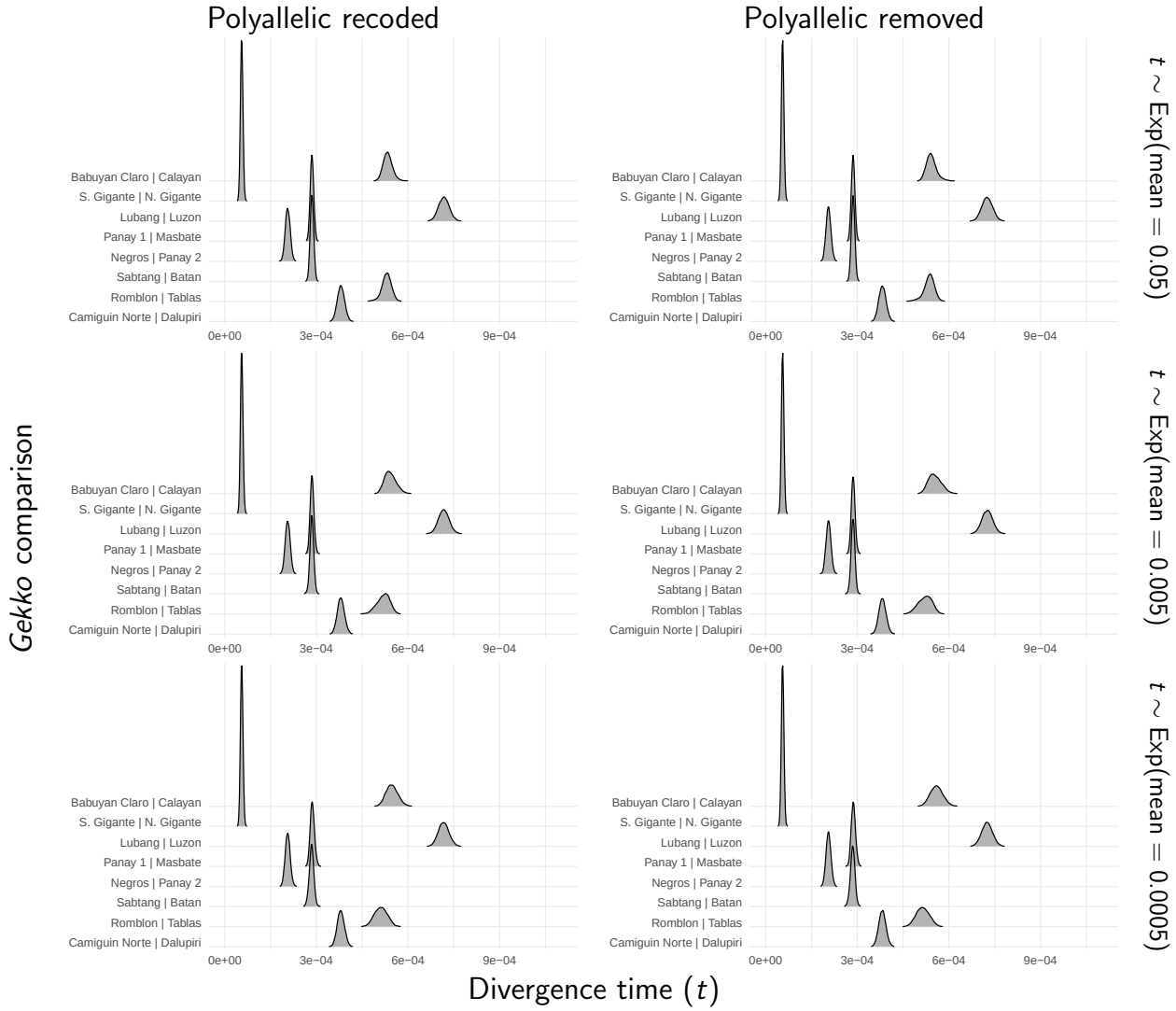

Figure S10. Approximate marginal posterior densities of divergence times for each pair of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

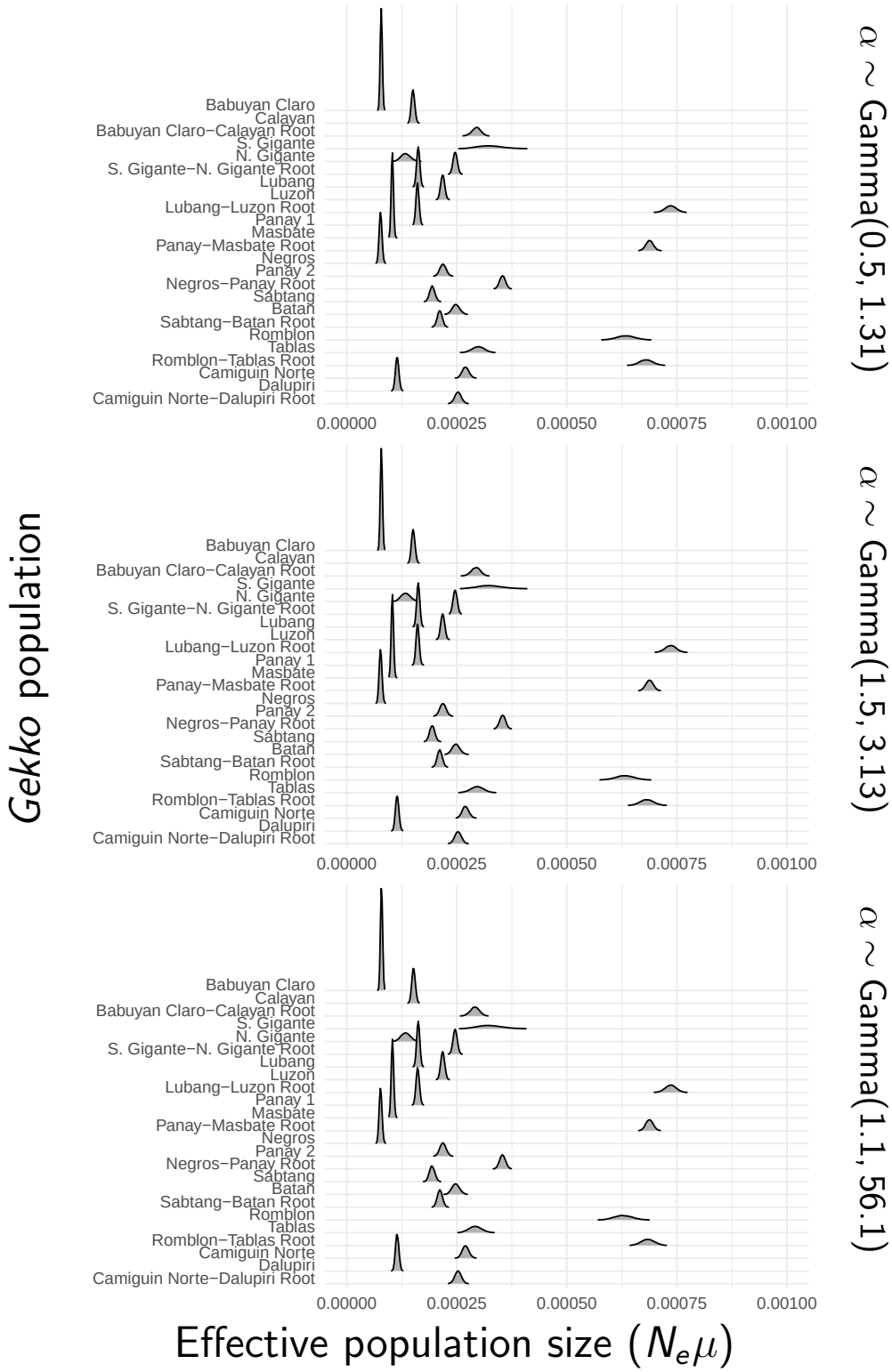

Figure S11. Approximate marginal posterior densities of population sizes for each pair of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

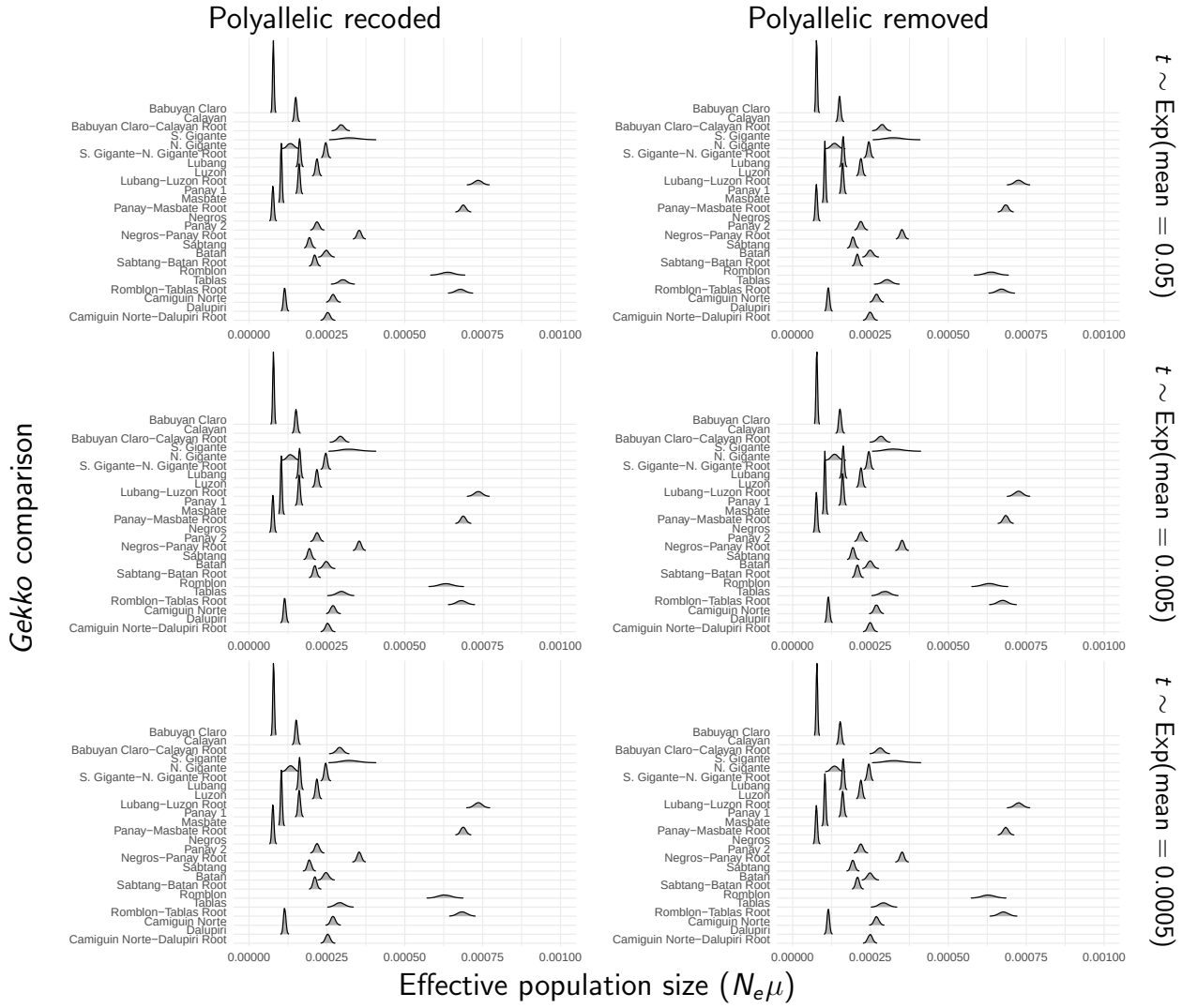

Figure S12. Approximate marginal posterior densities of population sizes for each pair of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).