The comparative biogeography of Philippine 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 73072-7029
[3]Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045, USA

April 26, 2019

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 "species-pump" model, in which repeated bouts of co-speciation are 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 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. To test this prediction, we collected comparative genomic data from 16 pairs of insular gecko populations. 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 speciation for gekkonid lizards on the Philippine Islands.

KEY WORDS: biogeography, diversification, Philippines, phylogeography

*Corresponding author: joaks@auburn.edu

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 archipelago represent such a model system, with more than 7,100 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, when, and by which mechanisms this diversity accumulated has piqued the interest of evolutionary biologists since the early development of the field of biogeography (Wallace, 1869; Huxley, 1868; Dickerson, 1928; Diamond and Gilpin, 1983; Brown, 2016; Lomolino et al., 2016).

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; Siler et al., 2010; Brown and Siler, 2014; Lomolino et al., 2016). 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; Spratt and Lisiecki, 2016), 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., 2010a; 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 (see Brown et al., 2013, for a review). 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 (Diamond and Gilpin, 1983; Brown and Guttman, 2002; Brown and Siler, 2014) and finer-scale isolating mechanisms that led to in situ diversification (Heaney et al., 2011; Linkem et al., 2011; Siler et al., 2011; Hosner et al., 2013). Nonetheless, the question remains: Was climate-driven fragmentation of the islands an important process of speciation?

A "species-pump" model of diversification (Jetz et al., 2004; Fjeldså and Rahbek, 2006; Kozak and Wiens, 2010; Sedano and Burns, 2010; Schoville et al., 2012; Papadopoulou and Knowles, 2015) via repeated vicariance predicts that divergences across taxa that occur on historically connected "island archipelagos" should be clustered around times of historical

isolating mechanisms. This model can be relevant to a diversity of structured environments, including deep-ocean (Ricklefs and Bermingham, 2008; Brown et al., 2013; Papadopoulou and Knowles, 2015) and coastal (Papadopoulou and Knowles, 2017; Senczuk et al., 2018) islands, and mountain tops (i.e., "sky" islands; Knowles, 2000, 2001; McCormack et al., 2008) Within the Philippines, this model predicts that divergences among taxa distributed across islands within the same PAIC should be associated with times of rising sea levels that 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) tested this prediction initially 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, and 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, providing evidence against predictions of the PAIC model of diversification, and underscoring the importance for caution against adhering to overly simplistic models of diversification 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 samples from pairs of localities that were inferred to be closely related, but not necessarily sister, from previous, independent genetic data (Siler et al., 2010, 2012, 2014; Welton et al., 2010a,b). 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. Furthermore, to avoid complications associated with intra-island population structure, we only used localities where previous genetic data were consistent with the samples being from a single population.

We also selected pairs of populations that are independent based on previous phylogenetic estimates (i.e., they do not share any branches in previously estimated phylogenies; Siler et al., 2010, 2012, 2014; Welton et al., 2010a,b).

We also sought to sample pairs that spanned islands connected during glacial periods, as well as islands that were never connected (Figure S1 & S2; Amante and Eakins, 2009; Brown et al., 2013; Spratt and Lisiecki, 2016). We included the latter as "controls." Because these islands were never connected, the distribution of closely-related populations inhabiting them can only be explained by inter-island dispersal. Thus, it seems reasonable to assume that divergence between these populations was either due to 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 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 vicariant 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; Tables 1 & S1).

The island pairs 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. To maximize the number of pairs and thus increase our chances of detecting shared divergences if they occurred, we include these populations in our analyses, but leave their connection status during glacial lowstands as ambiguous (Figure 1; Table 1).

Some of our pairs are currently recognized as different species, whereas others are not (Table 1). Recent taxonomic work on both of these genera of lizards suggests they comprise many more species than previously recognized, with more revisions necessary (Brown et al., 2009; Linkem et al., 2010b; Siler et al., 2010; Welton et al., 2010a,b; Brown et al., 2011; Grismer et al., 2018b,a). Over-water dispersal events are necessary to explain the existence of these populations on oceanic islands. However, these events are likely too random and rare to contribute meaningful levels of gene flow between islands. Therefore, given the insularity of these populations, we assume that none of them are experiencing migration during interglacial periods, like today. Thus, all of the populations we sampled are likely independent evolutionary lineages, regardless of whether they are currently recognized by taxonomists as distinct species.

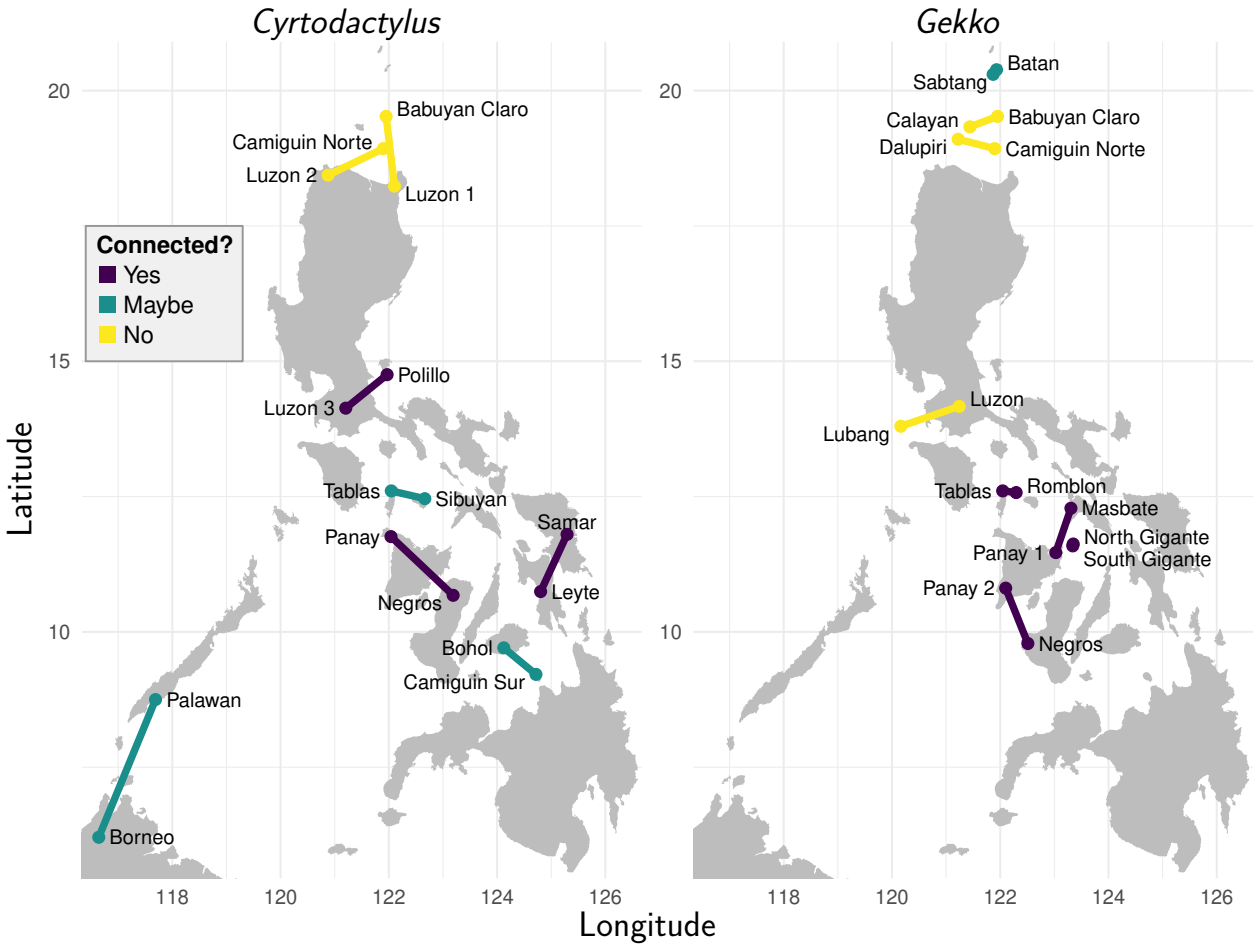

Figure 1. Philippine sampling localities for the eight pairs of *Cyrtodactylus* (left) and *Gekko* (right) populations. Localities for each pair are connected by a line and color-coded (see key) to indicate whether the islands were connected via terrestrial dry land bridges that formed during Pleistocene 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	2	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, 2019). Although 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, 2019). 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, after which there is no migration. This method can directly estimate the likelihood of values of these unknown parameters from orthologous biallelic characters by analytically integrating over all possible gene trees and mutational histories (Bryant et al., 2012; Oaks, 2019). 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 (Siler et al., 2010, 2012, 2014; Welton et al., 2010a,b) allowing us to make the simplifying assumption that all pairs *within* each genus share the same rate of mutation. Furthermore, we set the rate to one so that effective population sizes and time are scaled by the mutation rate, and thus time is in expected substitutions per site.

Based on previous data (Siler et al., 2010; Welton et al., 2010a,b) we assumed a prior on divergence times of $\tau \sim \text{Exponential}(\text{mean} = 0.005)$ 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 on divergence times of $\tau \sim \text{Exponential}(\text{mean} = 0.05)$. Based on previous data (Siler et al., 2012, 2014), we assumed a divergence-time prior of $\tau \sim \text{Exponential}(\text{mean} = 0.0005)$ 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 $\alpha \sim \text{Gamma}(1.1, 56.1)$ 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 $\alpha \sim \text{Gamma}(1.5, 3.13)$ and $\alpha \sim \text{Gamma}(0.5, 1.31)$. The former corresponds to 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 $\text{Gamma}(\text{shape} = 4.0, \text{mean} = 0.004)$ on the effective size of the populations scaled by the mutation rate ($N_e\mu$), and a prior distribution of $\text{Gamma}(\text{shape} = 100, \text{mean} = 1)$ 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). Data that satisfy

this assumption include single-nucleotide polymorphisms (SNPs) that are well-spaced across the genome. However, by analyzing simulated data, Oaks (2019) showed the method performs better when all linked sites are used than when data are excluded to avoid violating the assumption of unlinked sites. We simulate data sets based on our gekkonid data to confirm these results hold for our sampling design (see below). Furthermore, when analyzing the RADseq from three of the *Gekko* population pairs we are analyzing here, Oaks (2019) found biologically unrealistic estimates of divergence times and population sizes when only unlinked variable sites (i.e., SNPs) were analyzed. Using additional simulations, Oaks (2019) found these unrealistic estimates were likely due to data-acquisition biases, which are known to be common in alignments from reduced-representation genomic libraries (Harvey et al., 2015; Linck and Battey, 2019). Oaks (2019) found that using all of the sites, rather than only SNPs, greatly improved the robustness of these parameter estimates to such acquisition biases. Considering all these findings, we are confident in the inclusion of all sites of our RADseq loci in the `ecoevolity` analyses. Given that all sites were used, the likelihood computed in `ecoevolity` was not conditioned on only sampling variable characters (Oaks, 2019).

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, 2019).

For each analysis, we ran 10 independent Markov chain Monte Carlo (MCMC; Metropolis et al., 1953; Hastings, 1970) chains for 150,000 generations, sampling every 100th generation. We assessed convergence and mixing of the chains by inspecting the potential scale reduction factor (PSRF; the square root of Equation 1.1 in Brooks and Gelman, 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).

2.5 Vetting our sampling and methodology

2.5.1 Simulation-based assessment of `ecoevolity` conditional on our sampling

Oaks (2019) tested the method implemented in `ecoevolity` using simulated data. However, our gekkonid RADseq data differ from the simulation conditions used by Oaks (2019) in a number of ways. For example, we have fewer individuals sampled from most of our populations than the five simulated by Oaks (2019), and the number of loci and sites vary dramatically among our pairs of populations (Table 1).

To assess how well `ecoevolity` is able to infer shared divergences based on our sampling design, we implemented new simulation options in the `simcoevolity` tool within the `ecoevolity` software package. Our modifications allow us to simulate data sets that exactly

match the sampling scheme of our gekkonid data. Specifically, we simulate data sets that match our empirical data in terms of

1. the number of loci for each pair of populations,
2. the number of sites within each locus, and
3. the number of gene copies sampled for each site (i.e., the same patterns of missing data).

We assume all loci are effectively unlinked with no intralocus recombination (i.e., all the sites of a locus evolved along the same gene tree that is independent of the other loci, conditional on the population history). Our simulator allows us to sample all sites from each locus, or only a maximum of one variable site per locus. The former violates the assumption of the model implemented in `ecoevolity` that all sites are effectively unlinked, whereas the latter avoids this model violation at the cost of excluding data.

For each genus, we simulated 500 fully sampled data sets and 500 data sets with, at most, one SNP sampled per locus. For each simulation, the divergence model and all parameter values were drawn from the same distributions we used as priors in our empirical analyses described above. Specifically, the concentration parameter of the Dirichlet process was drawn from a gamma distribution of Gamma$(1.5, 3.13)$, and the time of each divergence event was distributed as Exponential(mean $= 0.005$) and Exponential(mean $= 0.0005$) for *Cyrtodactylus* and *Gekko*, respectively. When analyzing each simulated dataset with `ecoevolity`, we used these same distributions as priors and ran four independent MCMC chains for 150,000 generations, sampling every 100th generation. After ignoring the first 501 "burn-in" samples from each chain (including the initial state), we collected 4,000 MCMC samples for each analysis.

2.5.2 Data partitioning to evaluate performance of `ecoevolity`

With our simulations, we can assess how well `ecoevolity` infers shared divergences we know to be true. However, the simulated data sets are undoubtedly simpler than our empirical data, without any model violations (barring the linked sites within loci) introducing variation. Thus, we took another approach using our gekkonid RADseq data directly. We split the 21,426 loci randomly from our population pair with the largest number of sampled loci (*C. philippinicus* from the islands of Sibuyan and Tablas) into two subsets of 10,713 loci. We then reanalyzed the data with `ecoevolity` using the methods described above, but treating the two subsets of loci as separate population pairs. If `ecoevolity` can reliably detect a shared divergence event, it should infer that the two sets of loci from the same pair of populations did indeed co-diverge.

3 Results

3.1 Data collection and MCMC convergence

Table 1 summarizes 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 char-

acters. The nucleotide diversity within and between each pair of populations is provided in Table S4. The 10 independent MCMC chains of all our `ecoevolity` analyses appeared to have converged almost immediately. We conservatively removed the first 101 samples, leaving 1,400 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 2,000, and all PSRF values were less than 1.005.

3.2 Testing for shared divergences

3.2.1 *Cyrtodactylus* 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 S3 & S4). The support is also consistent across both priors on divergence times and whether polyallelic sites are recoded or removed (Figures S5 & S6). Estimates of effective population sizes are also very robust to priors on α and τ, and whether polyallelic sites are recoded or removed (Figures S7 & S8).

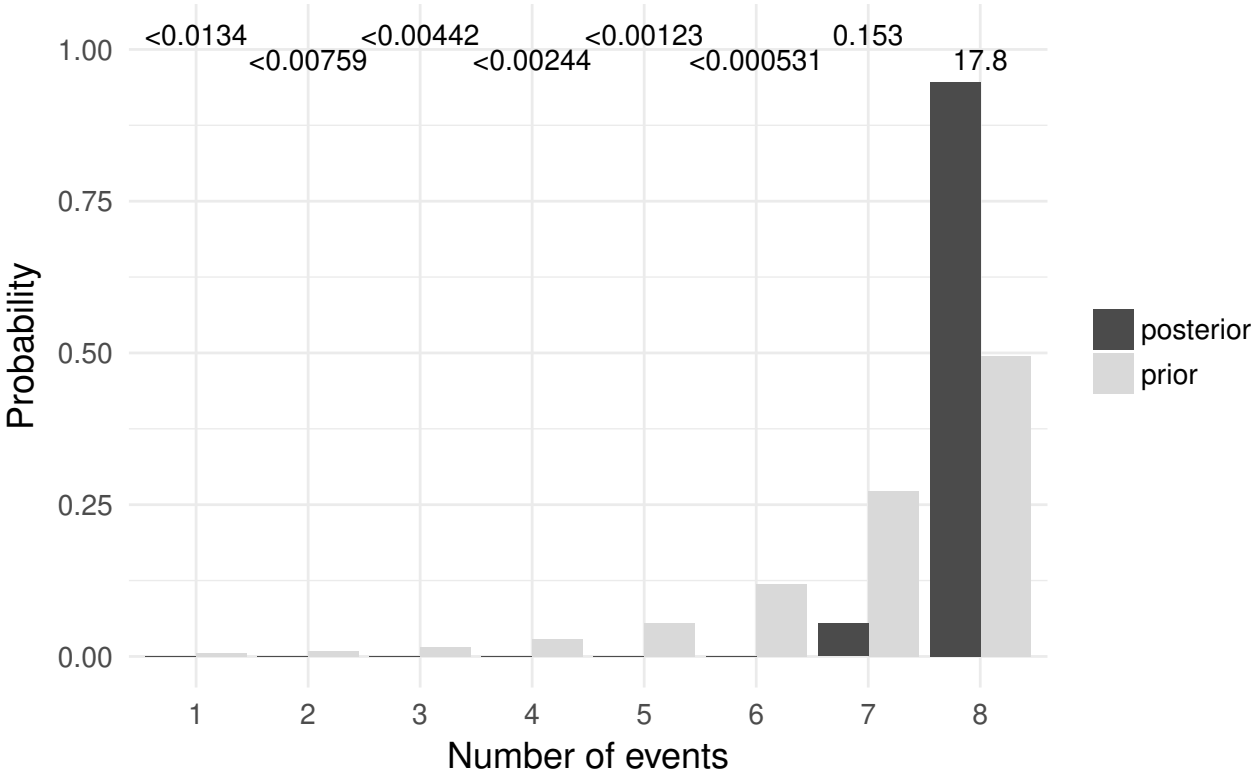

Figure 2. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of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

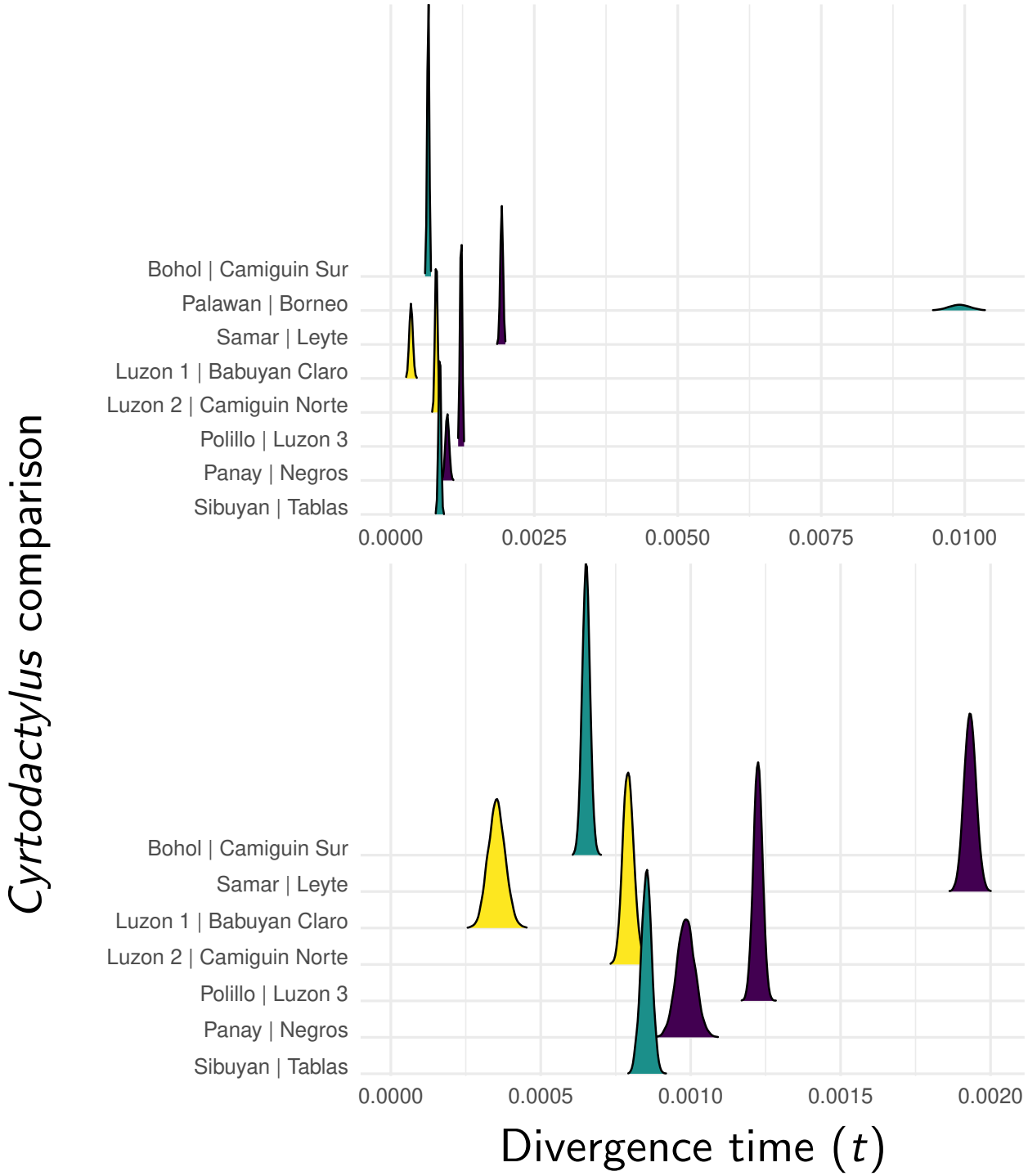

Figure 3. Approximate marginal posterior densities of divergence times (in expected substitutions per site) for each pair of *Cyrtodactylus* populations. The density plot of each pair is color-coded to indicate whether the islands were connected during glacial periods (Fig. 1). The top plot shows all eight pairs of populations, whereas the bottom plot excludes the pair of *C. redimiculus* and *C. baluensis* from Palawan and Borneo. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

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 the concentration parameter, support increases for this shared divergence (Figures S9 & S10). Under the prior that puts most of the probability on one shared event, posterior probabilities prefer six divergences (Figure S9) with another shared divergence between *G. crombota* and *G. rossi* on the islands of Babuyan Claro and Calayan and *G. romblon* on the islands of Romblon and Tablas (Figure S10); 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 S11 & S12).

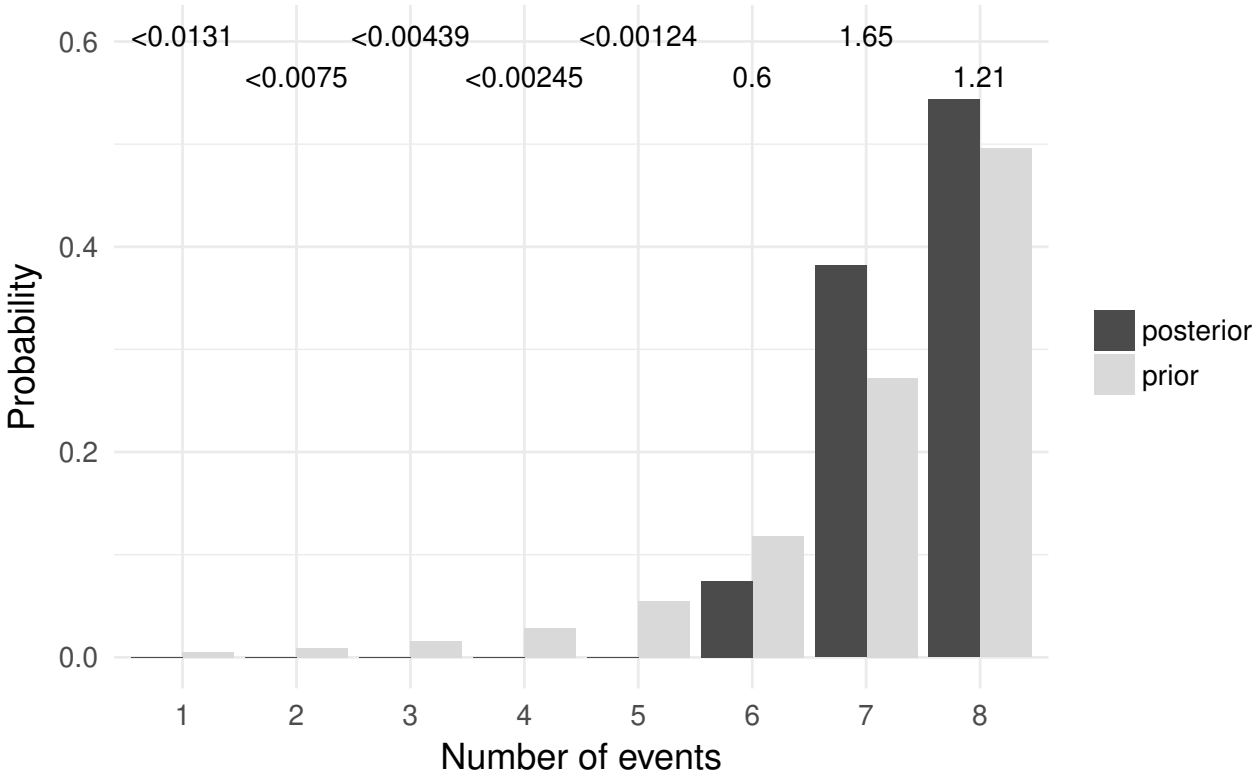

Figure 4. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations. Bayes factors for each number of divergence times is given above the corresponding bars. Each Bayes factor compares the corresponding number of events 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 S12), and population size estimates are robust to priors on α and τ, and whether polyallelic sites are recoded or removed (Figures S13 & S14).

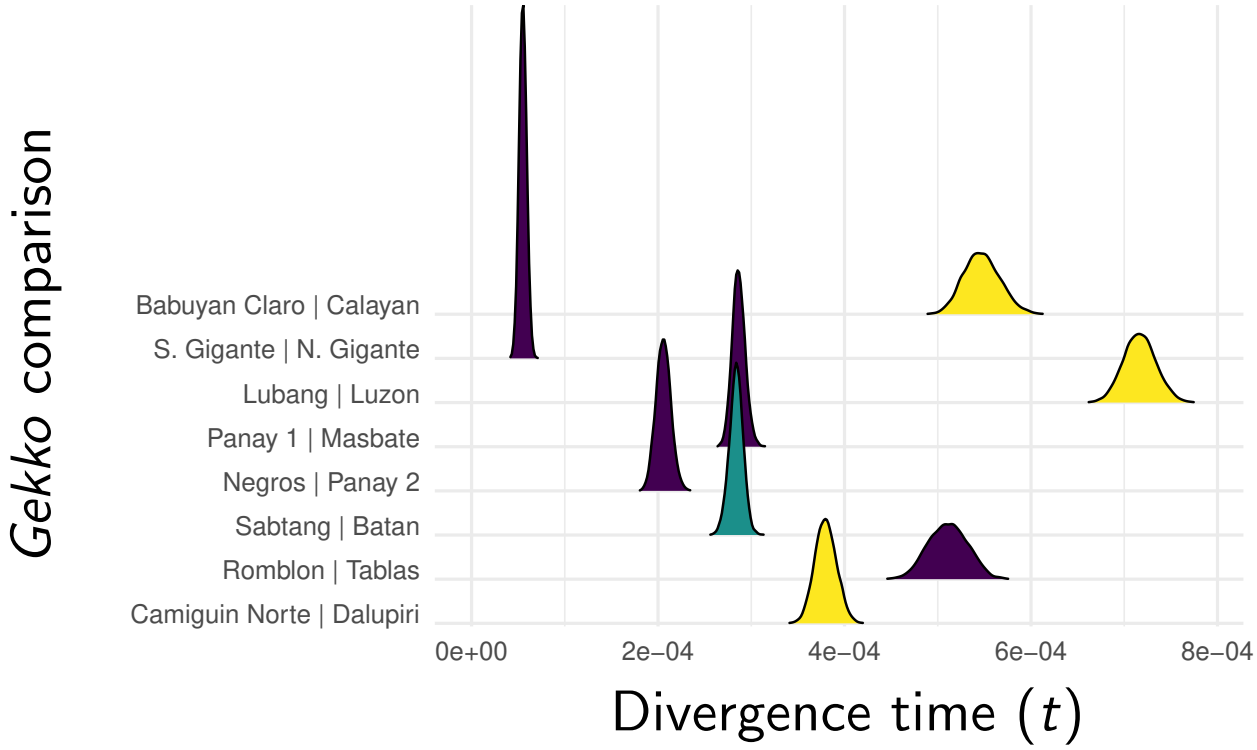

Figure 5. Approximate marginal posterior densities of divergence times (in expected substitutions per site) for each pair of *Gekko* populations. The density plot of each pair is color-coded to indicate whether the islands were connected during glacial periods (Fig. 1). Figure generated with ggridges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

3.3 Simulation results

The results from analyses of the simulated data sets show our gekkonid RADseq data are sufficient for `ecoevolity` to accurately estimate the timing (Figure 6) and number (Figure 7) of divergence events, and the effective sizes of the ancestral (Figure S15) and descendant populations (Figure S16). Consistent with Oaks (2019), we also find that estimation accuracy and precision are much better when all sites are analyzed rather than only unlinked SNPs.

3.4 Data partitioning to vet `ecoevolity`

When we reanalyzed our *Cyrtodactylus* data with the loci from the pair of *C. philippinicus* populations from the islands of Sibuyan and Tablas randomly split into two sets and treated as separate comparisons, the Bayes factors approximated by `ecoevolity` strongly support (Jeffreys, 1961) that the two subsets of loci codiverged. The posterior odds of them codiverging was 72.83 (posterior probability 0.963) and 144.81 (posterior prorbability 0.889) times greater than the prior odds when the hyperprior on the concentration parameter of the Dirichlet process was Gamma$(1.5, 3.13)$ and Gamma$(1.1, 56.1)$, respectively. We performed these calcuations with the `sumcoevolity` tool of the `ecoevolity` software package, using one million simulations under the Dirichlet process to approximate the prior odds.

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. crombota* and *G. rossi* on the islands of Babuyan Claro and Calayan and *G. romblon* 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 S9, S10, S11, & S12). 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

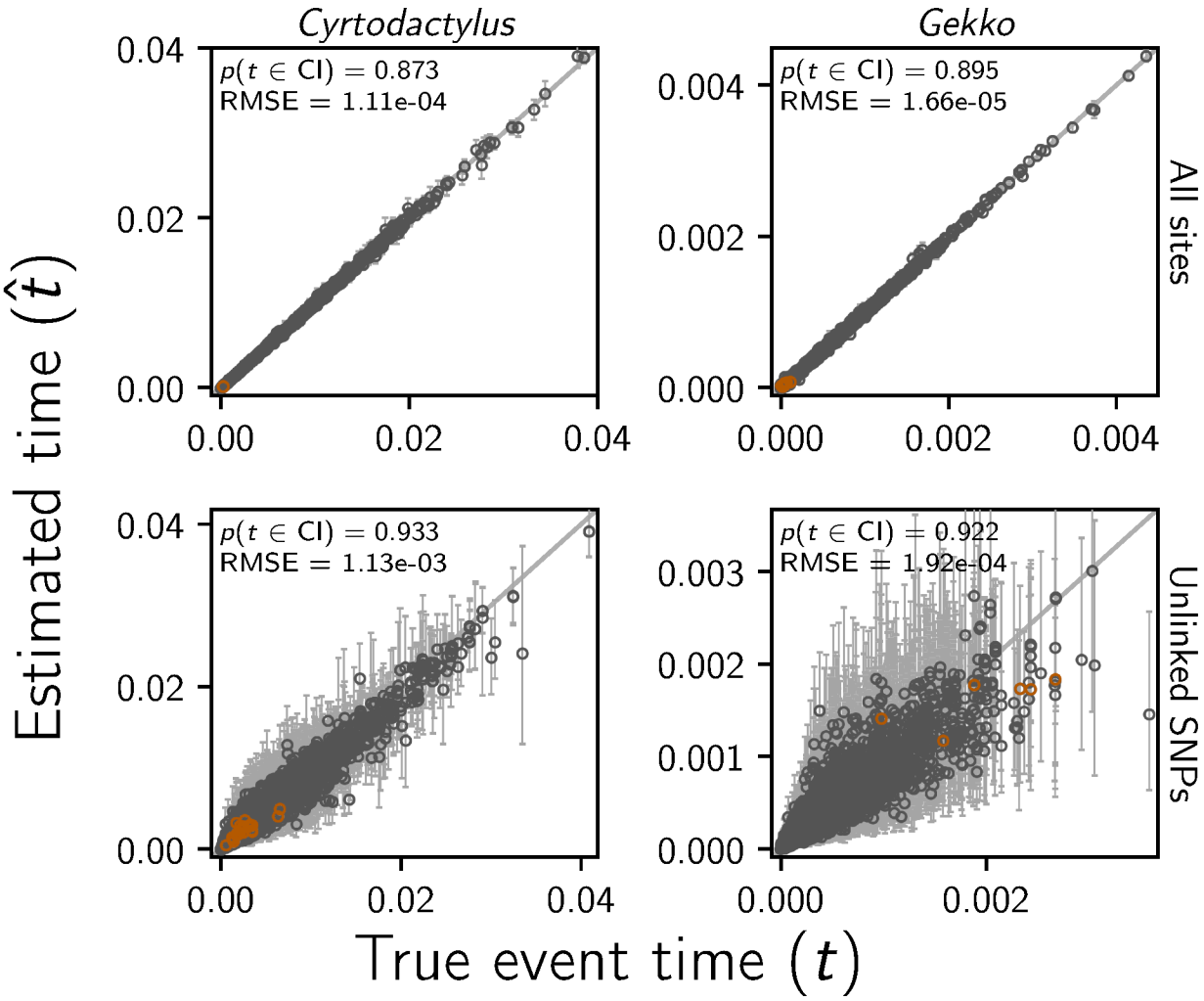

Figure 6. The accuracy and precision of `ecoevolity` divergence-time estimates (in units of expected subsitutions per site) when applied to data simulated to match our *Cyrtodactylus* (left) and *Gekko* (right) RADseq data sets with all sites (top) or only one SNP per locus (bottom). Each circle and associated error bars represents the posterior mean and 95% credible interval for the time that a pair of populations diverged. Estimates for which the potential-scale reduction factor was greater than 1.2 (Brooks and Gelman, 1998) are highlighted in orange. Each plot consists of 4,000 estimates—500 simulated data sets, each with eight pairs of populations. For each plot, the root-mean-square error (RMSE) and the proportion of estimates for which the 95% credible interval contained the true value—$p(t \in \mathrm{CI})$—is given. Figure generated with matplotlib Version 2.0.0 (Hunter, 2007).

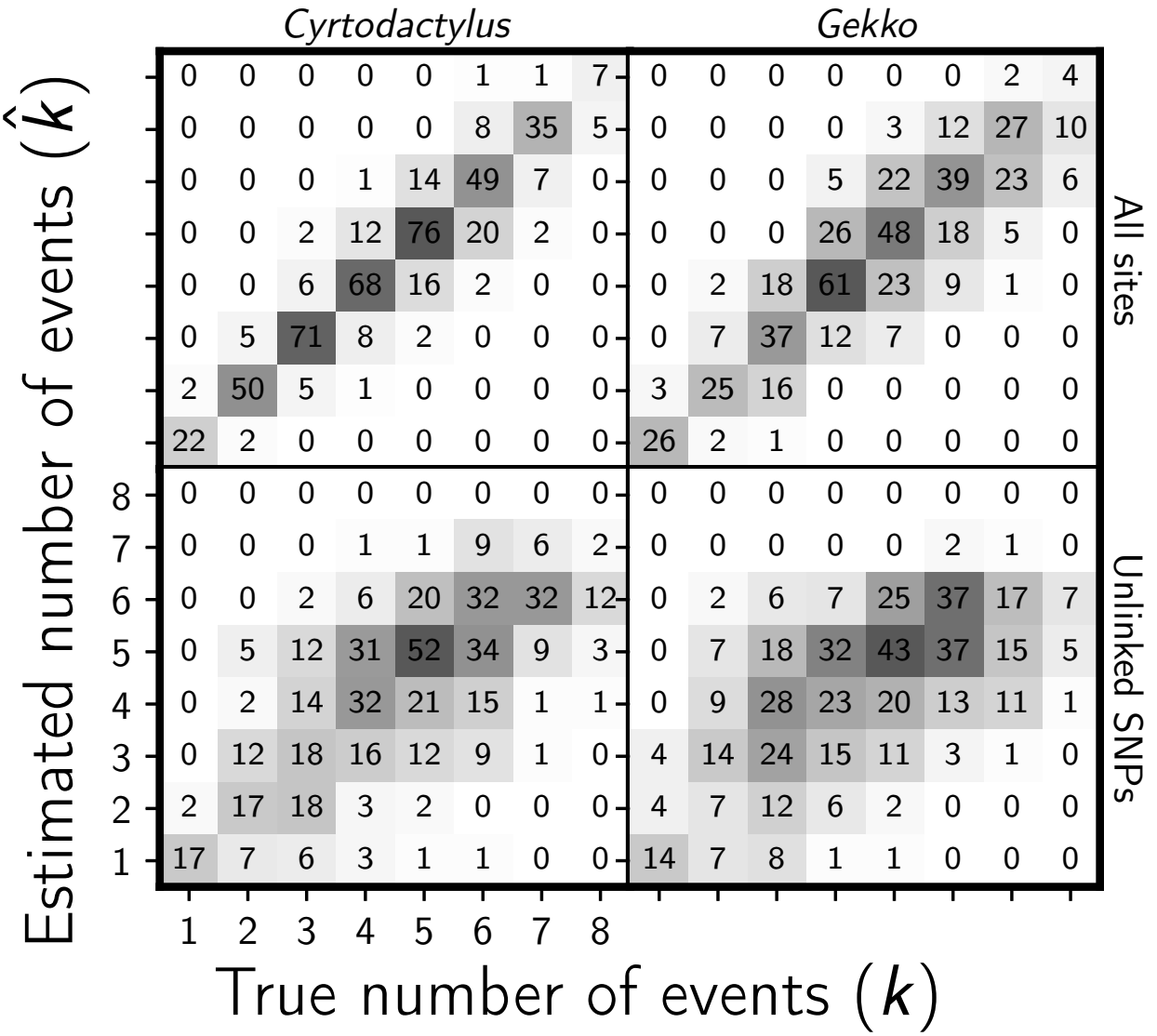

Figure 7. The accuracy of `ecoevolity` to estimate the number of divergence events when applied to data simulated to match our *Cyrtodactylus* (left) and *Gekko* (right) RADseq data sets with all sites (top) or only one SNP per locus (bottom). Each plot illustrates the results of the analyses of 500 simulated data sets, each with eight pairs of populations. The number of simulation replicates that fall within each possible cell of true versus estimated numbers of events is shown, and cells with more replicates are shaded darker. Figure generated with matplotlib Version 2.0.0 (Hunter, 2007).

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. Therefore, 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. Based on probability theory, we should favor the explanation that all eight pairs of *Gekko* populations diverged independently.

4.2 Caveats and implications for the PAIC diversification model

4.2.1 Limited numbers of comparisons

For each genus we sampled 5 or 6 pairs of populations that span two islands that were connected (or nearly so) by terrestrial dry land bridges during Pleistocene glacial periods (Brown and Diesmos, 2009; Brown et al., 2013). The connectivity between each pair of islands was likely fragmented by rising sea levels six or more times over the past three million years (Figures S1 & S2; Rohling et al., 1998; Siddall et al., 2003; Spratt and Lisiecki, 2016; Amante and Eakins, 2009). 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. Nonetheless, with seven pairs (11 including ambiguous island connections), we find no support for shared divergences, suggesting that, at the very least, climate-mediated vicariance is not the primary mode of population divergence in these insular gekkonids.

4.2.2 Variation in fragmentation times among island pairs

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 bathymetry data (Amante and Eakins, 2009), all of the previously connected pairs of islands we sampled (Figure 1) were connected when sea levels were 5–15m below current levels (Figure S1). Based on these bathymetry data and the sea level projections of Spratt and Lisiecki (2016), the timing of fragmentation among these pairs of islands would have differed by less than 3,000 years during the last two interglacial periods (Figure S2). 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 3.5×10^{-6} substitutions per site between two pairs that diverged during the same interglacial. This is likely an over-estimate, given that both the number of years between island separations and the mutation rate are toward the upper end of plausible.

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} substitutions per site. This is more than 2.7 times larger than our maximum expected divergence within an interglacial cycle of 3.5×10^{-6} substitutions per site, suggesting that `ecoevolity` would not have the temporal resolution given our data to distinguish the divergence times of two pairs that diverged during the same interglacial fragmentation event. Also, among all remaining pairs, we found support for distinct divergences with much larger differences than expected within an interglacial period. Thus, it seems unlikely that variation in island separation times within interglacial periods explains the variation in divergence times we see across the pairs of gekkonid populations.

Nonetheless, it would be ideal to sample pairs of populations that are co-distributed across the same pair of islands so that we know the fragmentation occurred at the same time. However, doing so comes with the inherent trade-off of having to compare more distantly related taxa. For example, we could sample pairs of *Cyrtodactylus* and *Gekko* populations that span the same islands, but not multiple pairs within each genus. This is important, because to compare divergence times among taxa, we need to make strong assumptions about their relative rates of mutation. Without information about mutation rates, we can assume equal rates across the comparisons, as we did here, but this assumption becomes much more questionable as the taxa we wish to compare are more distantly related from one another. Given the variation among island fragmentation times is small ($< 3{,}000$ years) relative to evolutionary timescales, we feel this trade-off is more desirable than making simplifying assumptions about relative mutation rates among distantly related taxa. However, if information about relative rates of mutation among taxa that span the same islands can be brought to bare, an analysis of such a system would provide a strong and complementry empirical test of the PAIC model.

4.2.3 Assumptions about mutation rates

As discussed above, we made the simplifying assumption that mutation rates were equal across pairs and constant through time. To minimize the impact of violations of this assumption, we analyzed the gekkonid genera separately, each of which comprise species that are closely related relative to other comparative phylogeographic studies that have made this assumption (Hickerson et al., 2006; Leaché et al., 2007; Plouviez et al., 2009; Voje et al., 2009; Barber and Klicka, 2010; Daza et al., 2010; Chan et al., 2011; Huang et al., 2011; Oaks et al., 2013; Stone et al., 2012; Smith et al., 2014). For example, all of the populations of *Gekko* we sampled were estimated to share a common ancestor less than 25 million years ago (Siler et al., 2012). Given that the pairs within each genus are closely related and have similar life histories, we do not expect substantive differences in mutation rates. Nonetheless, small differences in rates would affect the comparability of our divergence time estimates across our comparisons. Because we assumed a mutation rate of one for all comparisons, the estimated time of divergence for each pair of populations would still be accurate in units of expected

substitutions per site. However, the assumption that the relative estimates among pairs are proportional to absolute time would be violated. Thus, at least some of the variation in divergence times we estimated among taxa is due to variation in rates of mutation.

A strong assumption about relative rates of mutation must be made for any comparative phylogeographic method to compare the timing of events across taxa (Hickerson et al., 2006; Huang et al., 2011; Chan et al., 2014; Oaks, 2014, 2019). This is because there is no information in the data to distinguish differences in mutation rates among comparisons when the population history of each is modeled separately (i.e., they are modeled as disconnected "species" trees). To relax this assumption, fully phylogenetic approaches to the problem of estimating co-divergences are needed so that information from the data about relative mutation rates across the phylogeny can inform the model while jointly estimating co-divergences.

4.2.4 Assumptions about migration

We also assumed there was no migration between the populations of each pair after they diverged. One reason for this assumption is practical: Currenty, `ecoevolity` does not model migration. Methods for estimating shared divergences that allow migration are based on approximate likelihoods and cannot handle genomic data (Huang et al., 2011; Oaks, 2014). Even when there is no migration, these methods have been shown to be extremely sensitive to prior assumptions and biased toward estimating shared divergences (Oaks et al., 2013; Hickerson et al., 2014; Oaks et al., 2014; Oaks, 2014). A primary cause of this poor behavior is that the insufficient summary statistics used by these methods contain little information about the divergence times and population sizes. Adding additional migration parameters to these models is likely to make inference more challenging; i.e., trying to estimate additional parameters with insufficient statistics. Although ignoring migration is not ideal, it allows us to use genomic data with a full-likelihood method that exhibits much more desirable statistical behavior than approximate alternatives.

The second reason for assuming no migration is biological; given the insularity and natural history of these geckos, we do not expect contemporary migration between pairs of islands to be an important process. Undoubtedly, these geckos have dispersed among islands of the Philippines, but such over-water dispersal events are likely too rare to meaningfully contribute to contemporary gene flow. Nonetheless, gene flow among connected islands during glacial periods certainly could have been significant. A biologically inspired model of migration would thus require assumptions about (1) periods of time when island exposure was conducive for migration, (2) divergences pre-dating these periods to allow migration during them, and (3) *absolute* rates of mutation to ensure the molecular evolution of both genera is on the same timescale as the divergence and subsequent periods of potential migration. As a result, modeling the timing and magnitude of repeated bouts of migration would be challenging and would necessarily ignore a lot of uncertainty, especially regarding absolute mutation rates. Also, the data would likely lack information about processes back beyond the most recent bout of migration. Instead, we can more simply estimate the last time each pair of populations experienced significant gene flow (i.e., the divergence time modeled by `ecoevolity`). This is a simpler inference problem, and the climate-driven PAIC model of vicariant divergence would still predict clustering among the most recent time the population

pairs diverged. Furthermore, it seems unlikely that migration during glacial periods would bias our approach toward recovering independent divergences.

4.2.5 Assumptions about population sizes

Although we allowed the ancestral and descendant populations to have different effective sizes, we assumed these sizes were constant through time. This assumption was likely violated, especially if many of the divergences between the pairs of populations we sampled were due to over-water dispersal. Such events would likely have a strong founder effect on the effective size of one of the descendant populations. These demographic changes could affect our estimates of divergence times, however, based on coalescent theory, there are two reasons these effects should not be very large. First, the effective size of (the rate of coalescence within) the ancestral population is the most influential on the divergence time, because it determines the lag between the population divergence and the final gene coalescences, the latter of which is what the genetic data directly inform. Thus, as long as we have a reasonable estimate of the effective size of the ancestral population before the divergence, we should be able to accurately estimate the time of divergence. Second, the error in divergence-time estimation caused by demography should be limited to a magnitude on the order of $\approx 2N_e$ (or $2N_e\mu$ in the current study), because this is the expected disparity between gene coalescence and population divergence. Therefore, the additional variation in the data that is explained by incorporating more demographic realism may be offset by the error introduced by the additional model complexity.

4.3 Performance of `ecoevolity` with empirical RADseq data?

Given the caveats discussed above, and those associated with any model-based inference, it is important to evaluate how well the method implemented in `ecoevolity` can estimate divergence times conditional on the RADseq data we sampled from the gekkonid populations. Perhaps, with so many genomic data, `ecoevolity` fails to accurately estimate uncertainty in divergence times, and thus is biased toward finding differences between comparisons that do not exist. However, when we included two comparisons that represented random subsets of the loci from the same pair of populations in the analysis, `ecoevolity` strongly supported that they co-diverged. Thus, given real RADseq data from two comparisons that co-diverged, `ecoevolity` can confidently place them together.

Furthermore, from analyzing 2,000 data sets that were simulated to match the dimensions of our gekkonid RADseq data, we found that `ecoevolity` was able to accurately and precisely estimate the timing (Figure 6) and number (Figure 7) of divergence events. Also, these results confirm the findings of Oaks (2019) that the method performs better when analyzing all sites rather than only unlinked SNPs. This is important, because it shows this behavior generalizes to data sets simulated to match the linkage and missing data patterns of empirical RADseq data.

It is not surprising that `ecoevolity` performs better when using all of the data, despite the linkage among sites within loci violating the model. This behavior matches theoretical expectations that the parameters in the model should not be biased by the linked sites, because information about linkage among sites is not used by the model. Linkage among

sites does not change the expected site patterns under the model, it only reduces the variance of those patterns. Thus, the accuracy of estimates of divergence times and effective sizes of populations should not be affected by linked sites, as demonstrated here and by Oaks (2019). Furthermore, removing all but (at most) one variable site per locus is a rather draconian measure to avoid violating the linkage assumption, because it discards a substantial amount of informative data. Our results are also consistent with those of Chifman and Kubatko (2014), who found quartet inference of species trees from SNP data was also robust to the violation of unlinked characters. Nonetheless, our results should not be generalized to other methods that assume unlinked characters, especially methods that use information about site linkage patterns.

Our simulation results also show that `ecoevolity` is robust to large disparities in the number of sampled individuals and loci (Table 1; Figures 6 & 7). This is not surprising given Oaks (2019) found the benefit of collecting more characters begins to plateau quickly, even with as few as 200 loci. For example, we can compare the estimation of divergence times between the pairs that were simulated to match our RADseq data from *C. philippinicus* populations from the islands of Luzon and Babuyan Claro versus Polillo and Luzon. The former pair consists of only two samples per population and 3,855 loci, whereas the latter has five samples per population and 19,561 loci (Table 1). Despite these large disparities in sampling, the accuracy and precision of divergence-time estimates are very similar, especially when all sites are included in the analyses (Figure S17).

Perhaps most importantly, our simulation results also allow us to better interpret our empirical findings. There are two patterns worth highlighting in this regard. (1) When applied to datasets that were simulated with all eight pairs diverging independently (the rightmost column of the plots in Figure 7), `ecoevolity` has only moderate success in preferring a model with eight divergence events. (2) When the true number of divergences is less than eight, `ecoevolity` almost never estimates eight divergences (only 3 out of almost 2,000 simulations; Figure 7). Taken together, these observations demonstrate that `ecoevolity` is unlikely to spuriously support the model where all pairs of populations diverge independently. Thus, the empirical support we found for all eight pairs of *Cyrtodactylus* and *Gekko* populations diverging independently is likely robust.

4.4 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 S5 & S11). Although 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 it 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, which are the result of the prior probabilities being updated by the marginal likelihoods of the divergence models. Because the marginal likelihoods are averaged with respect to the priors on all the parameters of the model, they can be sensitive to those priors regardless of the informativeness the data (Oaks et al., 2019).

4.5 Gekkonid diversification in the Philippines

Philippine *Gekko* and *Cyrtodactylus* species are nocturnal insectivores that inhabit a variety of geological substrates, forest types, and variable local atmospheric conditions (e.g., prevailing temperatures and precipitation) throughout many Philippine landmasses where they are codistributed. The Philippine *Gekko* populations studied here exhibit a more restricted microhabitat preference for rocky substrates, and appear more patchily distributed in the vicinity of exposed rock, caves, and karst formations. *Cyrtodactylus* are also found in some of these same habitats, but also utilize forest interior microhabitats, where they perch additionally on tree trunks and understory vegetation. Until recently, widespread species were recognized in both genera (e.g., *Cyrtodactylus philippinicus* and *Gekko mindorensis*), suggesting their microhabitat preferences do not limit their vagility. However, subsequent investigations of widespread taxa have shown Philippine gekkonid species diversity to be underestimated greatly and represented by a larger number of range-restricted lineages (Brown et al., 2009, 2011; Linkem et al., 2010b; Welton et al., 2010a,b).

Our findings are consistent with what we know about Philippine gekkonid natural history (RMB and CDS pers. obs.) and on-going revisions of their species boundaries. The spatial and temporal variation in connectivity among pockets of these lizards' preferred structural microbabitats is likely a key predictor of past and present disributions of gekkonid populations across the Philippines. Environmental heterogeneity within islands is likely important for isolating populations, as evidenced by previous findings of multiple divergent lineages inhabiting the same island, such as the northern island of Luzon (Siler et al., 2010, 2012, 2014; Welton et al., 2010a,b). Furthermore, we consider it likely that the ephemeral, low-elevation habitat on the land bridges exposed during glacial periods was unsuitable for these forest species (Esselstyn and Brown, 2009; Hosner et al., 2014). In fact, rare, long-distance dispersal events among islands might actually be more likely to occur via rafting on vegetation across marine barriers following typhoons than movement across exposed land bridges during glacial periods (Linkem et al., 2013; Brown, 2016). Both intra-island processes of isolation associated with spatial and temporal environmental heterogeneity (Brown et al., 2013) and inter-island rafting (Brown, 2016) would predict our results of idiosyncratic divergence times across inter-island pairs of populations.

4.6 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. Although 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, and was archived on zenodo (https://dx.doi.org/10.5281/zenodo.2642461). Supplemental Tables 1 and 2 and the animation from Supplemental Figure 2 are also available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.n34d4m7.

6 Acknowledgments

We thank the members of the Phyletica Lab (the phyleticians) for helpful feedback on multiple drafts of this paper. We also thank editors Mohamed Noor and David Weisrock and four anonymous reviewers for constructive feedback that greatly improved this work. We are grateful to Patrick Monnahan and John Kelly for their help with the MSG library protocol. The computational work was made possible by the Auburn University (AU) Hopper Cluster supported by the AU Office of Information Technology and a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority. This work was possible thanks to funding provided to JRO from the National Science Foundation (DBI 1308885 and DEB 1656004). Philippines research has been supported by NSF DEB 0073199, 0743491, 0640737, 1418895, and 1654388 to RMB; NSF IOS 1353683 and DEB 1657648 and 0804115 to CDS; and the University of Kansas Biodiversity Institute. This paper is contribution number 891 of the Auburn University Museum of Natural History.

7 Author contributions

J.R.O., C.D.S., and R.M.B. conceived the study. R.M.B. and C.D.S. conducted fieldwork. J.R.O. collected, assembled, and analyzed the sequence data. J.R.O. led the writing of the manuscript, with contributions from C.D.S. and R.M.B.

References

Amante, C. and B. W. Eakins, 2009. ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis. Tech. rep., National Geophysical Data Center, Marine Geology

and Geophysics Division, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA.

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.

Barber, B. R. and J. Klicka, 2010. Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. Proceedings Of The Royal Society B-Biological Sciences 277:2675–2681.

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., 2016. Biogeography of land vertebrates. Pp. 211–220, *in* R. M. Kilman, ed. Encyclopdia of Evolutionary Biology, vol. 1, 1st ed. Academic Press, Oxford, UK.

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., C. Oliveros, C. D. Siler, and A. C. Diesmos, 2009. Phylogeny of *Gekko* from the Northern Philippines, and description of a new species from Calayan Island. Journal of Herpetology 43:620–636.

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, A. C. Diesmos, and A. C. Alcala, 2011. A new *Gekko* from Sibuyan Island, Central Philippines. Herpetologica 67:460–476.

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.

Chan, L. M., J. L. Brown, and A. D. Yoder, 2011. Integrating statistical genetic and geospatial methods brings new power to phylogeography. Molecular Phylogenetics and Evolution 59:523–537.

Chan, Y. L., D. Schanzenbach, and M. J. Hickerson, 2014. Detecting concerted demographic response across community assemblages using hierarchical approximate Bayesian computation. Molecular Biology and Evolution 31:2501–2515.

Chifman, J. and L. Kubatko, 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324.

Daza, J. M., T. A. Castoe, and C. L. Parkinson, 2010. Using regional comparative phylogeographic data from snake lineages to infer historical processes in Middle America. Ecography 33:343–354.

Diamond, J. M. and M. E. Gilpin, 1983. Biogeographic umbilici and the origin of the Philippine avifauna. Oikos 41:307–321.

Dickerson, R. E., 1928. Distribution of life in the Philippines. Philippine Bureau of Science, Manila, 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.

Fjeldså, J. and C. Rahbek, 2006. Diversification of tanagers, a species rich bird group, from lowlands to montane regions of South America. Integrative and Comparative Biology 46:72–81.

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.

Grismer, L. L., P. L. Wood, Jr., E. S. H. Quah, M. L. Murdoch, M. S. Grismer, M. W. Herr, R. E. Espinoza, R. M. Brown, and A. Lin, 2018a. A phylogenetic taxonomy of the *Cyrtodactylus peguensis* group (Reptilia: Squamata: Gekkonidae) with descriptions of two new species from Myanmar. PeerJ 6:e5575.

Grismer, L. L., P. L. Wood, Jr., M. K. Thura, T. Zin, E. S. H. Quah, M. L. Murdoch, M. S. Grismer, A. Lin, H. Kyaw, and N. Lwin, 2018b. Twelve new species of *Cyrtodactylus* Gray (Squamata: Gekkonidae) from isolated limestone habitats in east-central and southern Myanmar demonstrate high localized diversity and unprecedented microendemism. Zoological Journal of the Linnean Society 182:862–959.

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

Harvey, M. G., C. D. Judy, G. F. Seeholzer, J. M. Maley, G. R. Graves, and R. T. Brumfield, 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3:e895.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

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., E. A. Stahl, and H. A. Lessios, 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435–2453.

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.

Hosner, P. A., A. S. Nyári, and R. G. Moyle, 2013. Water barriers and intra-island isolation contribute to diversification in the insular *Aethopyga* sunbirds (Aves: Nectariniidae). Journal of Biogeography 40:1094–1106.

Hosner, P. A., L. A. Sánchez-González, A. T. Peterson, and R. G. Moyle, 2014. Climate-driven diversification and Pleistocene refugia in philippine birds: Evidence from phylogeographic structure and paleoenvironmental niche modeling. Evolution 68:2658–2674.

Huang, W., N. Takebayashi, Y. Qi, and M. J. Hickerson, 2011. MTML-msBayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC Bioinformatics 12:1.

Hunter, J. D., 2007. Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9:90–95.

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.

———, 1961. Theory of Probability. 3rd ed. Oxford University Press, Oxford, U.K.

Jetz, W., C. Rahbek, and R. K. Colwell, 2004. The coincidence of rarity and richness and the potential signature of history in centres of endemism. Ecology Letters 7:1180–1191.

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.

Knowles, L. L., 2000. Tests of Pleistocene speciation in montane grasshoppers (genus textitmelanoplus) from the sky islands of western North America. Evolution 54:1337–1348.

———, 2001. Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. Proceedings of the Royal Society of London B: Biological Sciences 268:319–324.

Kozak, K. and J. Wiens, 2010. Niche conservatism drives elevational diversity patterns in Appalachian salamanders. The American Naturalist 176:40–54.

Leaché, A. D., S. C. Crews, and M. J. Hickerson, 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biology Letters 3:646–650.

Linck, E. B. and C. J. Battey, 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. Molecular Ecology Resources DOI: 10.1111/1755-0998.12995.

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

Linkem, C. W., R. M. Brown, C. D. Siler, B. J. Evans, C. C. Austin, D. T. Iskandar, A. C. Diesmos, J. Supriatna, N. Andayani, and J. A. McGuire, 2013. Stochastic faunal exchanges drive diversification in widespread Wallacean and Pacific island lizards (Squamata: Scincidae: *Lamprolepis smaragdina*). Journal of Biogeography 40:507–520.

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, 2010a. Species boundaries and cryptic lineage diversity in a Philippine forest skink complex (Reptilia; Squamata; Scincidae: Lygosominae). Molecular Phylogenetics and Evolution 56:572–585.

Linkem, C. W., C. D. Siler, A. C. Diesmos, and R. M. Brown, 2010b. A new species of *Gekko* (Squamata: Gekkonidae) from central Luzon Island, Philippines. zootaxa. Zootaxa 2396:37–94.

Lomolino, M. V., B. R. Riddle, and J. H. Brown, 2016. Biogeography. 5th ed. Sinauer Associates, Sunderland, Massachusetts, USA.

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

McCormack, J. E., B. S. Bowen, and T. B. Smith, 2008. Integrating paleoecology and genetics of bird populations in two sky island archipelagos. BMC Biology 6:28.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21:1087–1092.

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.

———, 2019. Full Bayesian comparative phylogeography from genomic data. Systematic Biology 68:371–395.

Oaks, J. R., K. A. Cobb, V. N. Minin, and A. D. Leaché, 2019. Marginal likelihoods in phylogenetics: a review of methods and applications. Systematic Biology DOI: 10.1093/sysbio/syz003.

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.

Pante, E. and B. Simon-Bouhet, 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS ONE 8:e73051. Doi:10.1371/journal.pone.0073051.

Papadopoulou, A. and L. L. Knowles, 2015. Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. Evolution 69:1501–1517.

———, 2017. Linking micro- and macroevolutionary perspectives to evaluate the role of Quaternary sea-level oscillations in island diversification. Evolution 71:2901–2917.

Plouviez, S., T. M. Shank, B. Faure, C. Daguin-Thiebaut, F. Viard, F. H. Lallier, and D. Jollivet, 2009. Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. Molecular Ecology 18:3903–3917.

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

Ricklefs, R. and E. Bermingham, 2008. The West Indies as a laboratory of biogeography and evolution. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363:2393–2413.

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.

Schoville, S. D., G. K. Roderick, and D. H. Kavanaugh, 2012. Testing the 'Pleistocene species pump' in alpine habitats: lineage diversification of flightless ground beetles (Coleoptera: Carabidae: *Nebria*) in relation to altitudinal zonation. Biological Journal of the Linnean Society 107:95–111.

Sedano, R. E. and K. J. Burns, 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). Journal of Biogeography 37:325–343.

Senczuk, G., K. Havenstein, V. Milana, C. Ripa, E. De Simone, R. Tiedemann, and R. Castiglia, 2018. Spotlight on islands: on the origin and diversification of an ancient lineage of the Italian wall lizard *Podarcis siculus* in the western Pontine Islands. Scientific Reports 8:15111.

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.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield, 2014. The drivers of tropical speciation. Nature 515:406–409.

Spratt, R. M. and L. E. Lisiecki, 2016. A Late Pleistocene sea level stack. Climate of the Past 12:1079–1092.

Stone, G. N., K. Lohse, J. A. Nicholls, P. Fuentes-Utrilla, F. Sinclair, K. Schönrogge, G. Csóka, G. Melika, J.-L. Nieves-Aldrey, J. Pujade-Villar, M. Tavakoli, R. R. Askew, and M. J. Hickerson, 2012. Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. Current Biology 22:532–537.

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.

Voje, K. L., C. Hemp, Ø. Flagstad, G.-P. Saetre, and N. C. Stenseth, 2009. Climatic change as an engine for speciation in flightless Orthoptera species inhabiting African mountains. Molecular Ecology 18:93–108.

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 Philippine 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 73072-7029
[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 available from the Dryad Digital Repository (https://datadryad.org/resource/doi:10.5061/dryad.n34d4m7/3).

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 available from the Dryad Digital Repository (https://datadryad.org/resource/doi:10.5061/dryad.n34d4m7/4). This is a subset of the data in Table S1.

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

Table S4. Per-site nucleotide diversity within (π_1 and π_2) and between ($\pi_{between}$) pairs of *Cyrtodactylus* and *Gekko* populations, calculated from the RADseq data using the SeqSift Python package (https://github.com/joaks1/SeqSift), which relies on Biopython (https://biopython.org/). The within-population nucleotide diversity (π_1 and π_2) is followed by the number of individuals sampled from the population in parentheses.

Species	Island 1	Island 2	π_1	π_2	$\pi_{between}$
C. annulatus	Bohol	Camiguin Sur	0.00118 (4)	0.00136 (4)	0.00303
C. baluensis-redimiculus	Palawan	Kinabalu	0.00135 (4)	0.00432 (3)	0.02475
C. gubaot-sumuroi	Samar	Leyte	0.00573 (5)	0.00223 (5)	0.00653
C. philippinicus	Luzon	Babuyan Claro	0.00131 (2)	0.00028 (2)	0.00226
C. philippinicus	Luzon	Camiguin Norte	0.00102 (3)	0.00068 (4)	0.00349
C. philippinicus	Polillo	Luzon	0.00226 (5)	0.00211 (5)	0.00429
C. philippinicus	Panay	Negros	0.00177 (3)	0.00095 (2)	0.00335
C. philippinicus	Sibuyan	Tablas	0.00136 (3)	0.00209 (3)	0.00319
G. crombota-rossi	Babuyan Claro	Calayan	0.00101 (5)	0.00111 (5)	0.00340
G. gigante	South Gigante	North Gigante	0.00167 (3)	0.00127 (4)	0.00149
G. mindorensis	Lubang	Luzon	0.00183 (5)	0.00119 (4)	0.00851
G. mindorensis	Panay	Masbate	0.00137 (4)	0.00130 (3)	0.00407
G. mindorensis	Negros	Panay	0.00103 (3)	0.00177 (5)	0.00259
G. porosus	Sabtang	Batan	0.00092 (4)	0.00115 (4)	0.00172
G. romblon	Romblon	Tablas	0.00178 (5)	0.00218 (2)	0.00349
G. sp. a-sp. b	Camiguin Norte	Dalupiri	0.00096 (5)	0.00111 (5)	0.00262

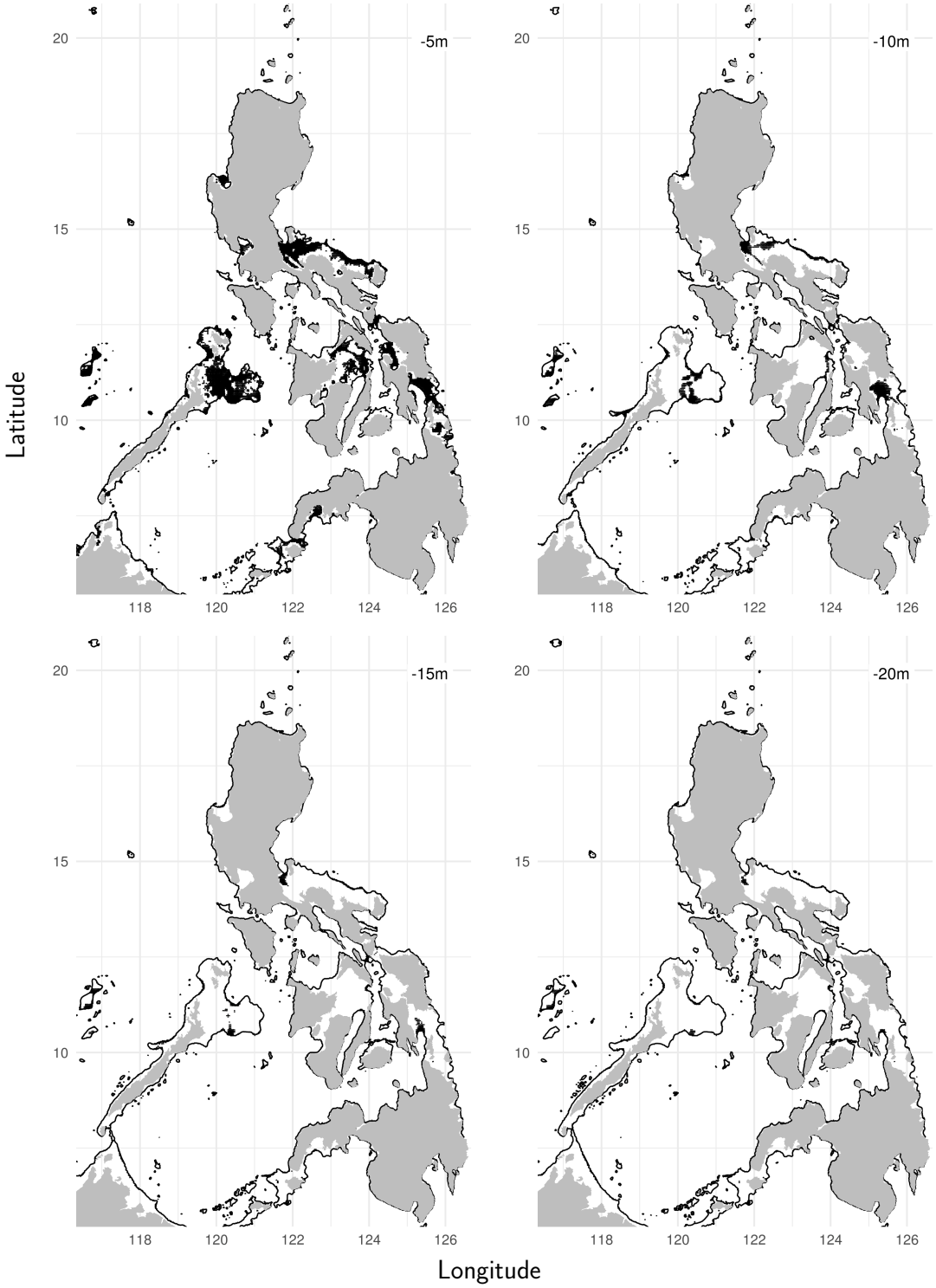

Figure S1. Bathymetry contours around the Philippine Islands at varying depths. Black lines show the contour associated with the depth indicated in the upper right of each plot. Contours are based on data from the ETOPO1 1-arc-minute global relief model (Amante and Eakins, 2009). Figure generated with marmap version 1.0.2 (Pante and Simon-Bouhet, 2013) and ggplot2 Version 2.2.1 (Wickham, 2009).

Figure S2. Animation of approximate sea-level changes in the Philippine Islands over the last 430,000 years. Sea-level estimates are from the projection of Spratt and Lisiecki (2016) based on seven reconstructions. Bathymetry data are from the ETOPO1 1-arc-minute global relief model (Amante and Eakins, 2009). Animation generated using marmap version 1.0.2 (Pante and Simon-Bouhet, 2013), ggplot2 Version 2.2.1 (Wickham, 2009), ImageMagick Version 6.9.10-8 Q16 x86_64 20180723, and FFmpeg Version 4.0.2-2. The source code for generating the plot is available at https://github.com/phyletica/animating-sea-level-change. This animation can also be viewed at https://youtu.be/NjGdCezUvw8 or downloaded from the Dryad Digital Repository: https://doi.org/10.5061/dryad.n34d4m7/5.

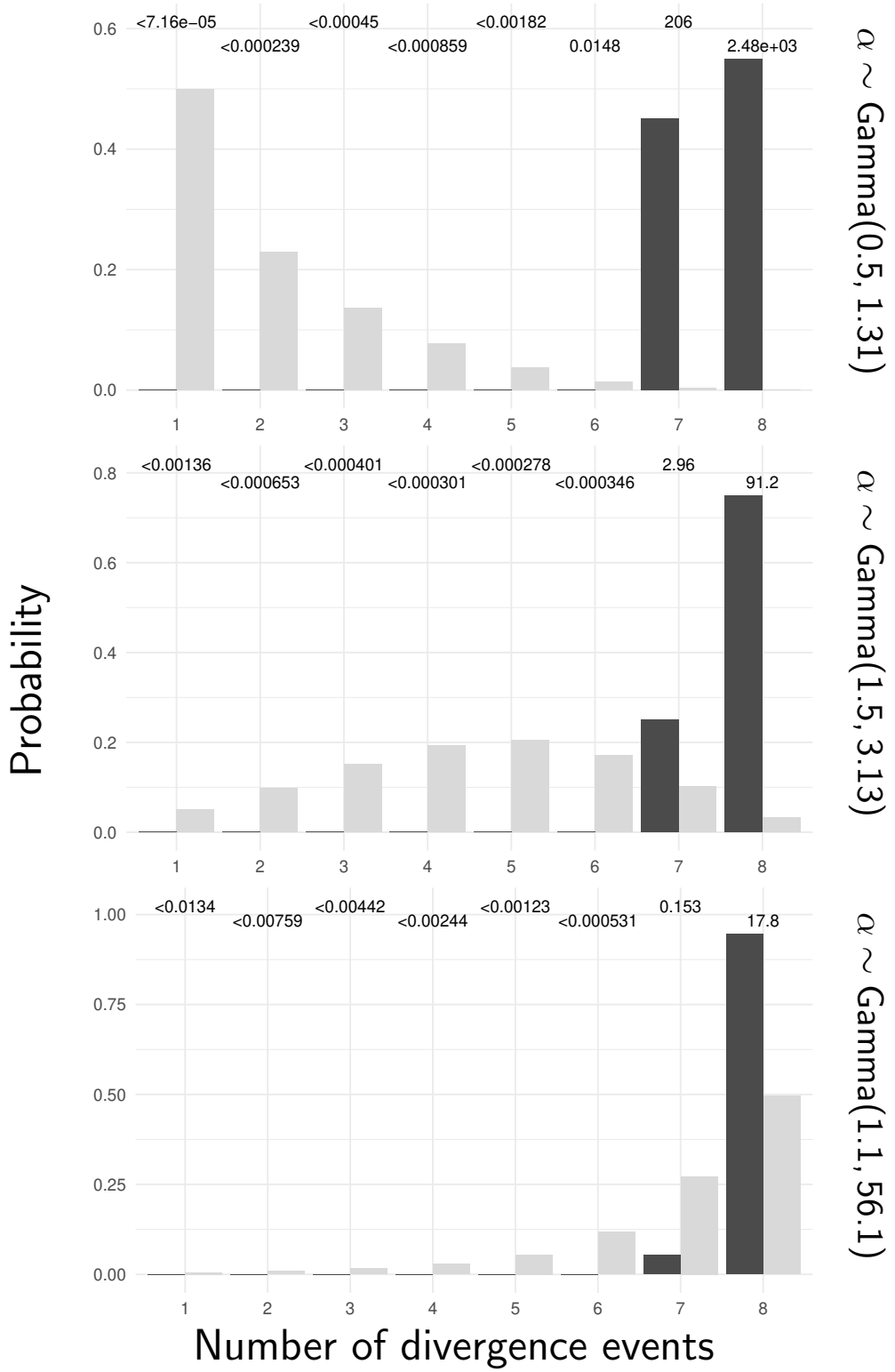

Figure S3. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Bayes factors for each number of divergence times is given 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).

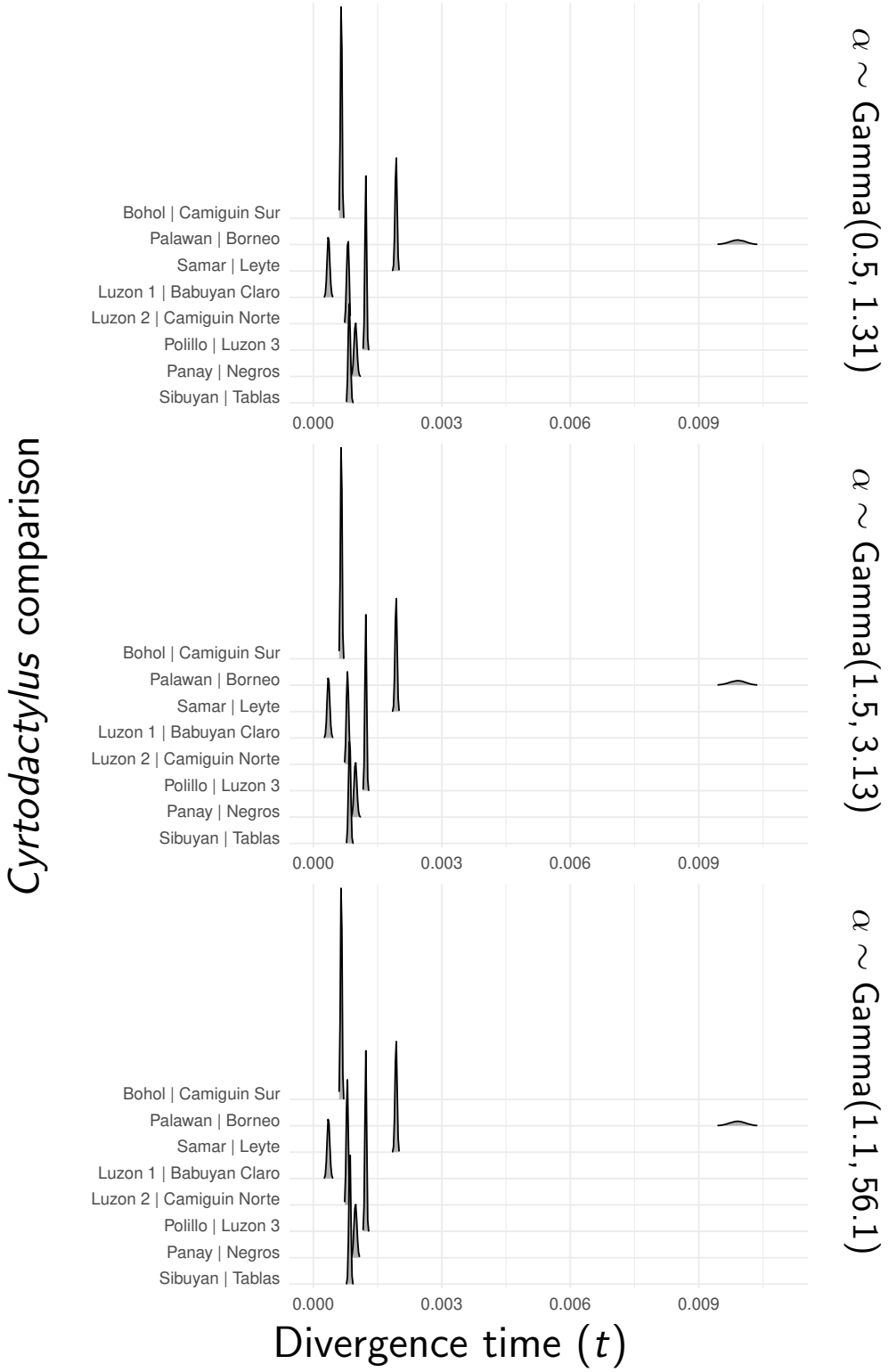

Figure S4. Approximate marginal posterior densities of divergence times for each pair of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

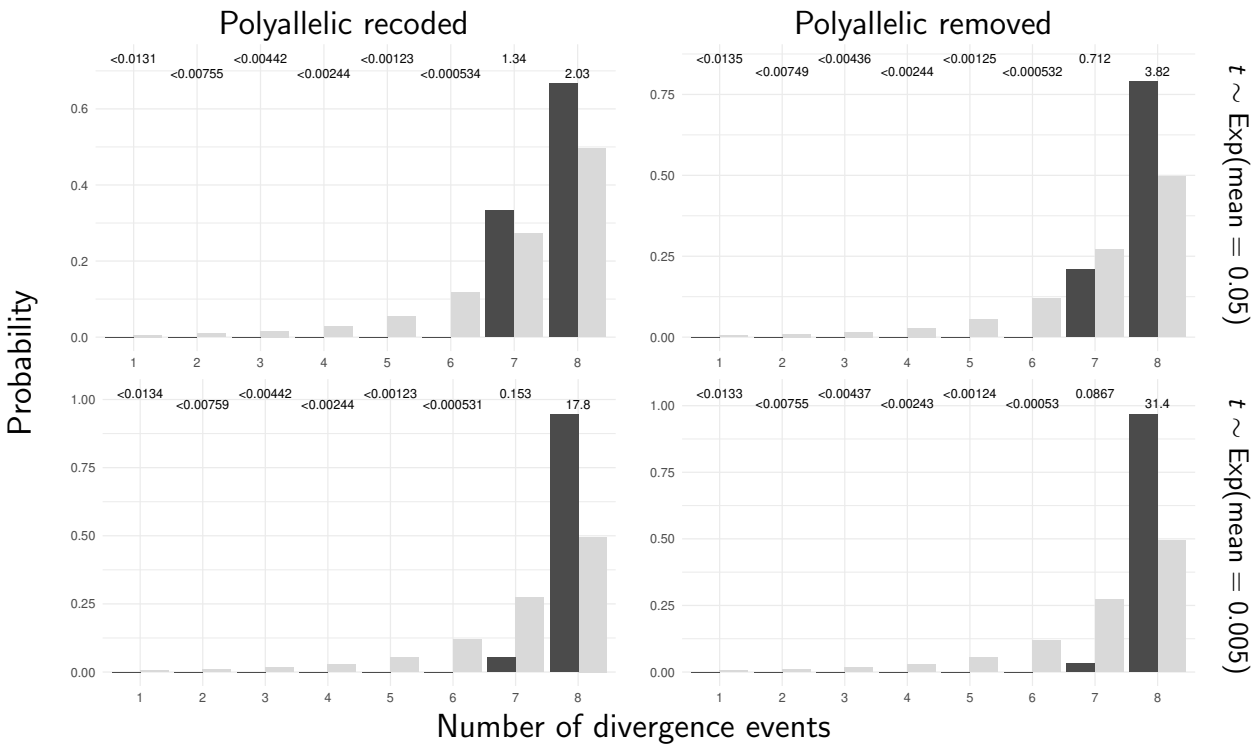

Figure S5. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Bayes factors 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).

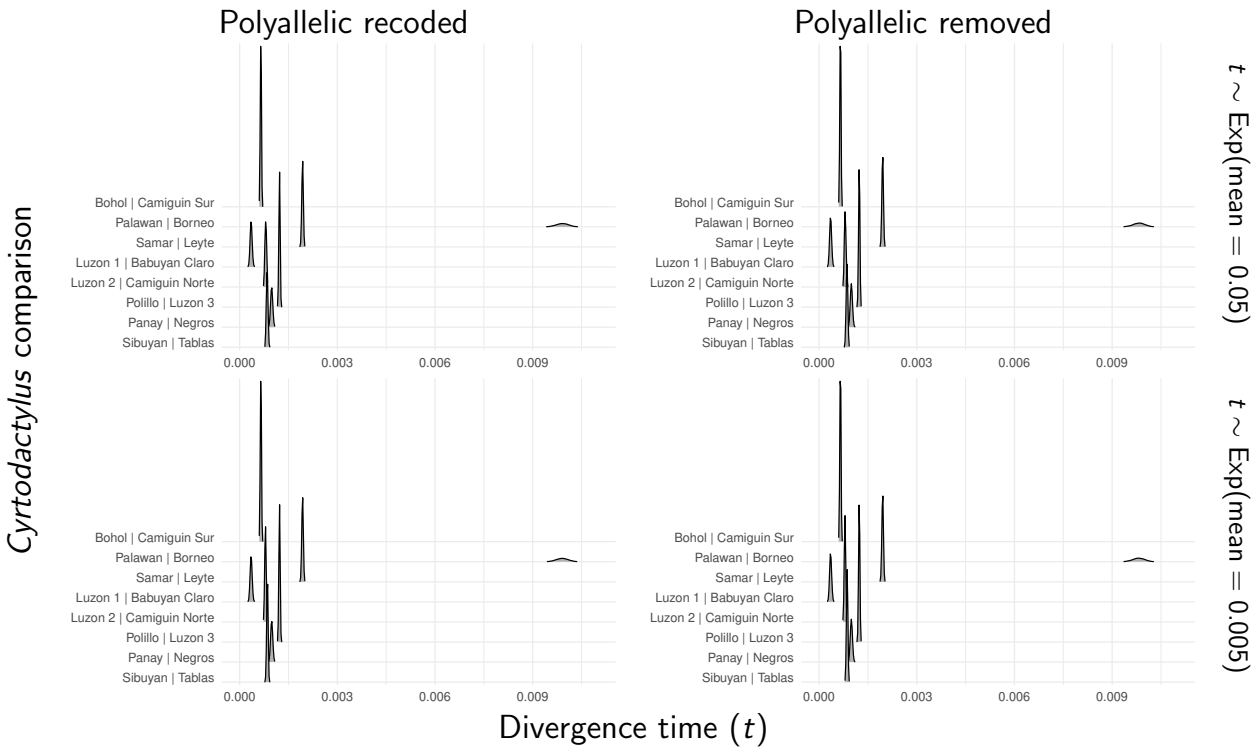

Figure S6. Approximate marginal posterior densities of divergence times for each pair of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018)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

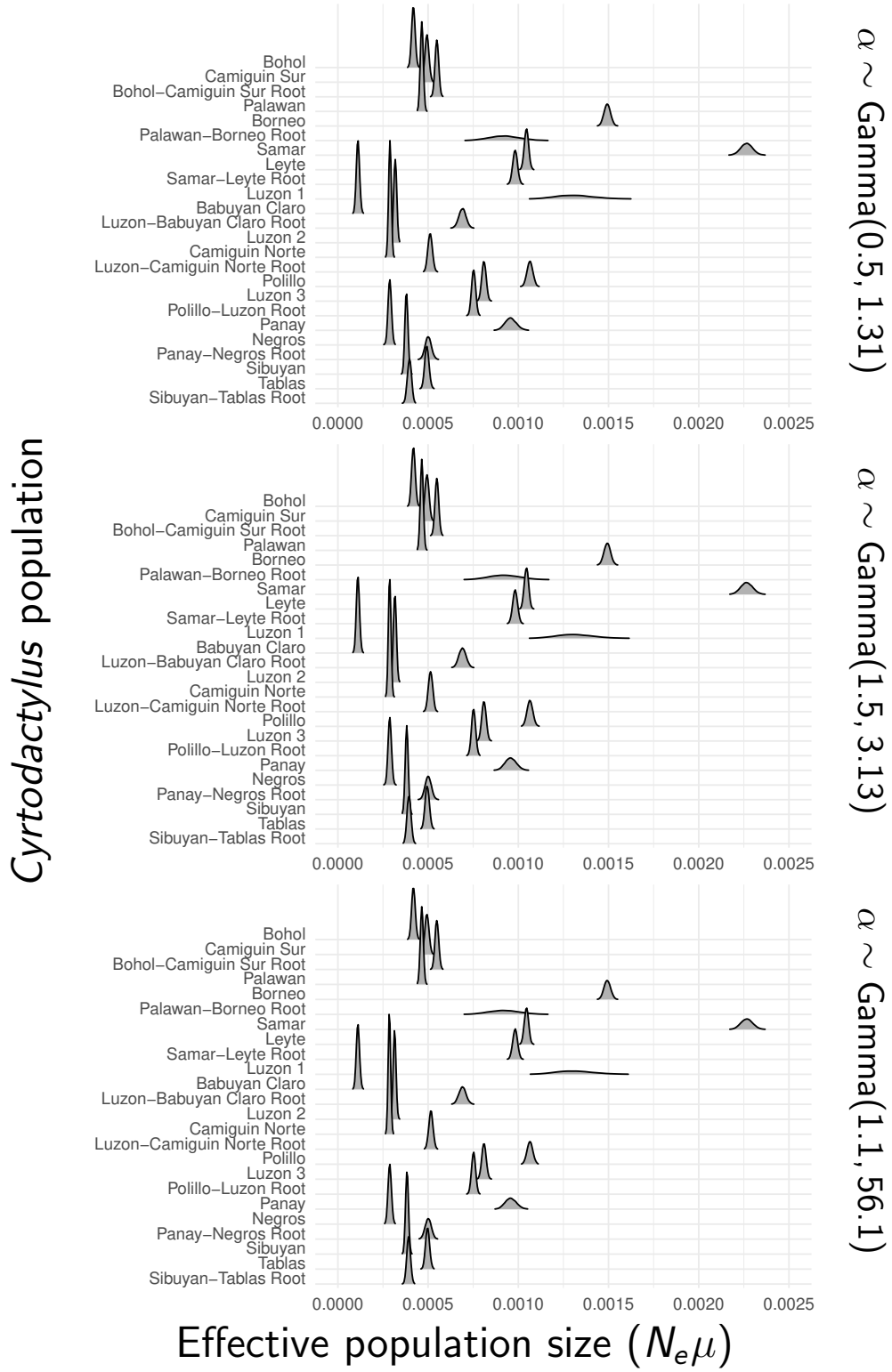

Figure S7. Approximate marginal posterior densities of population sizes for each pair of *Cyrtodactylus* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

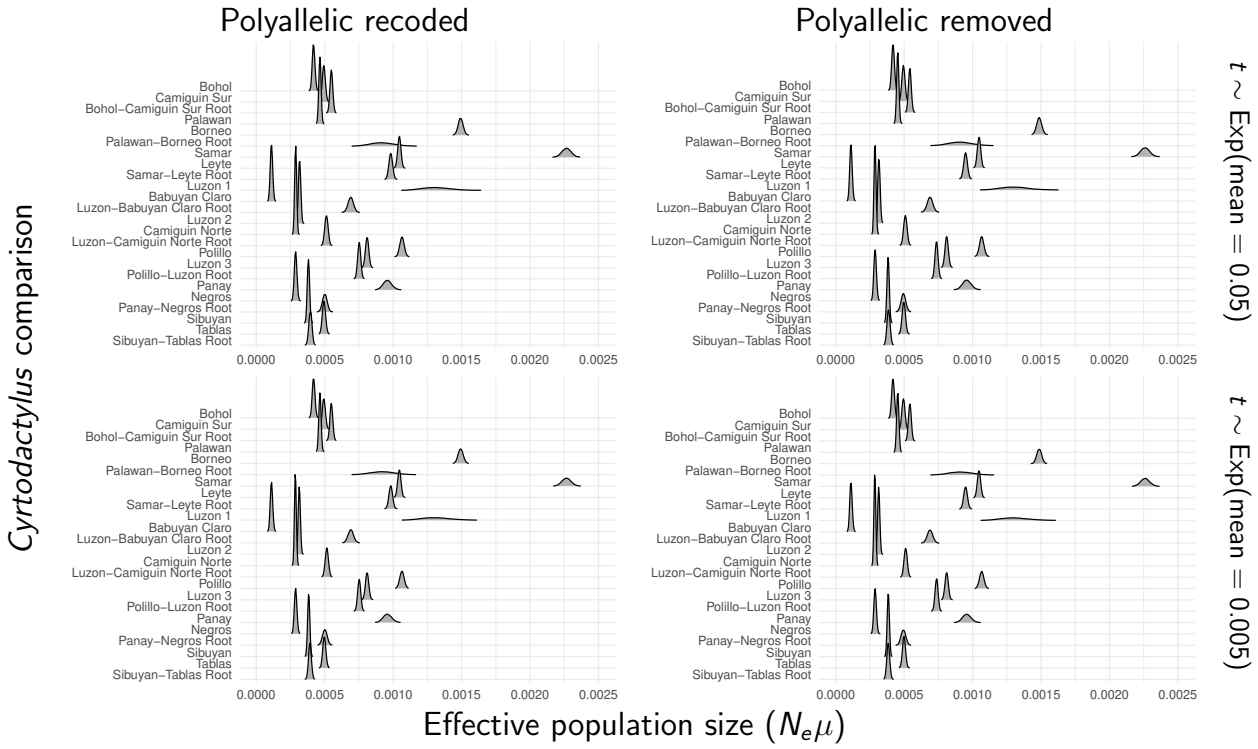

Figure S8. Approximate marginal posterior densities of population sizes for each pair of *Cyrtodactylus* populations under four different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

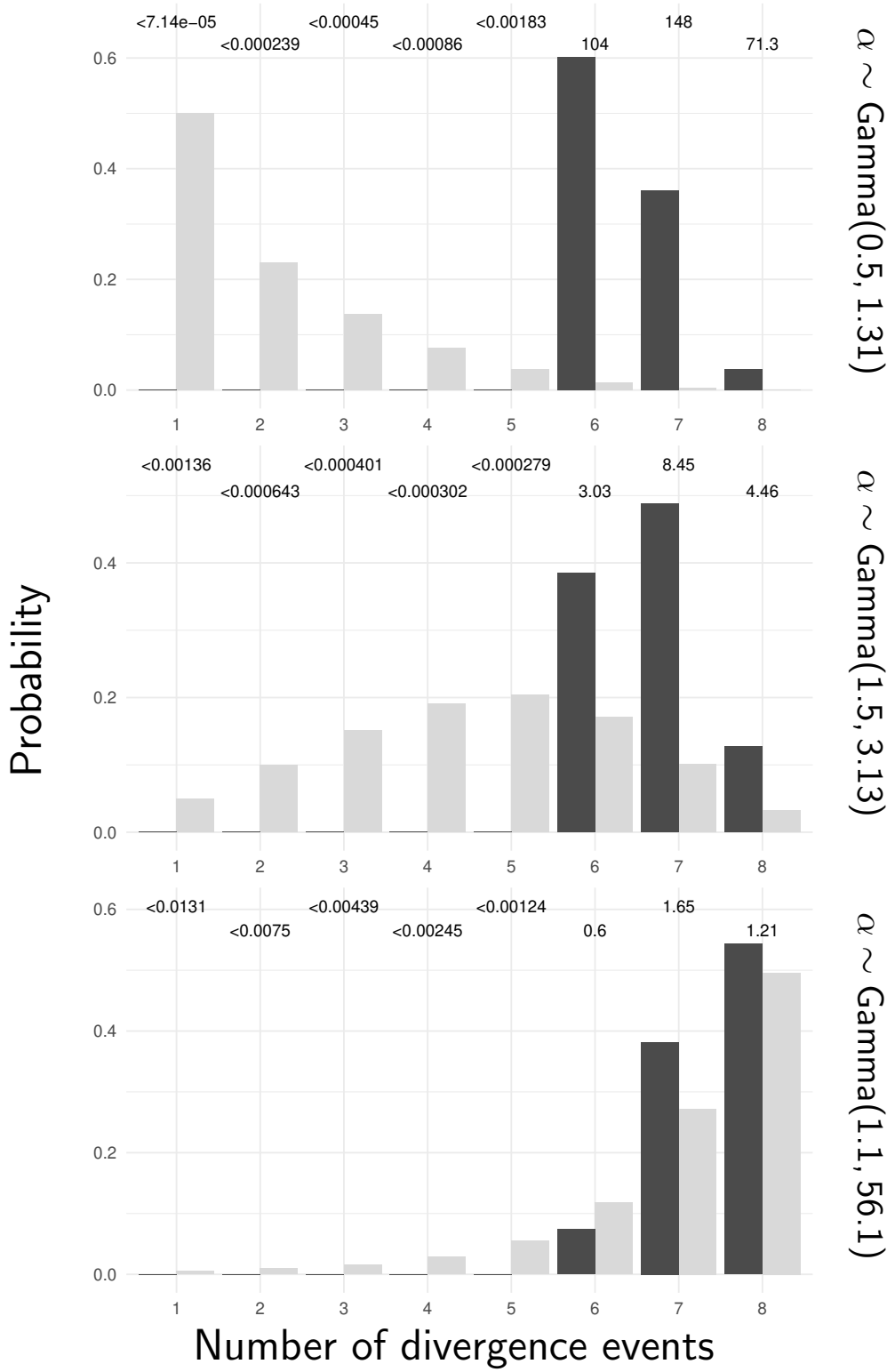

Figure S9. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Bayes factors for each number of divergence times is given above 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).

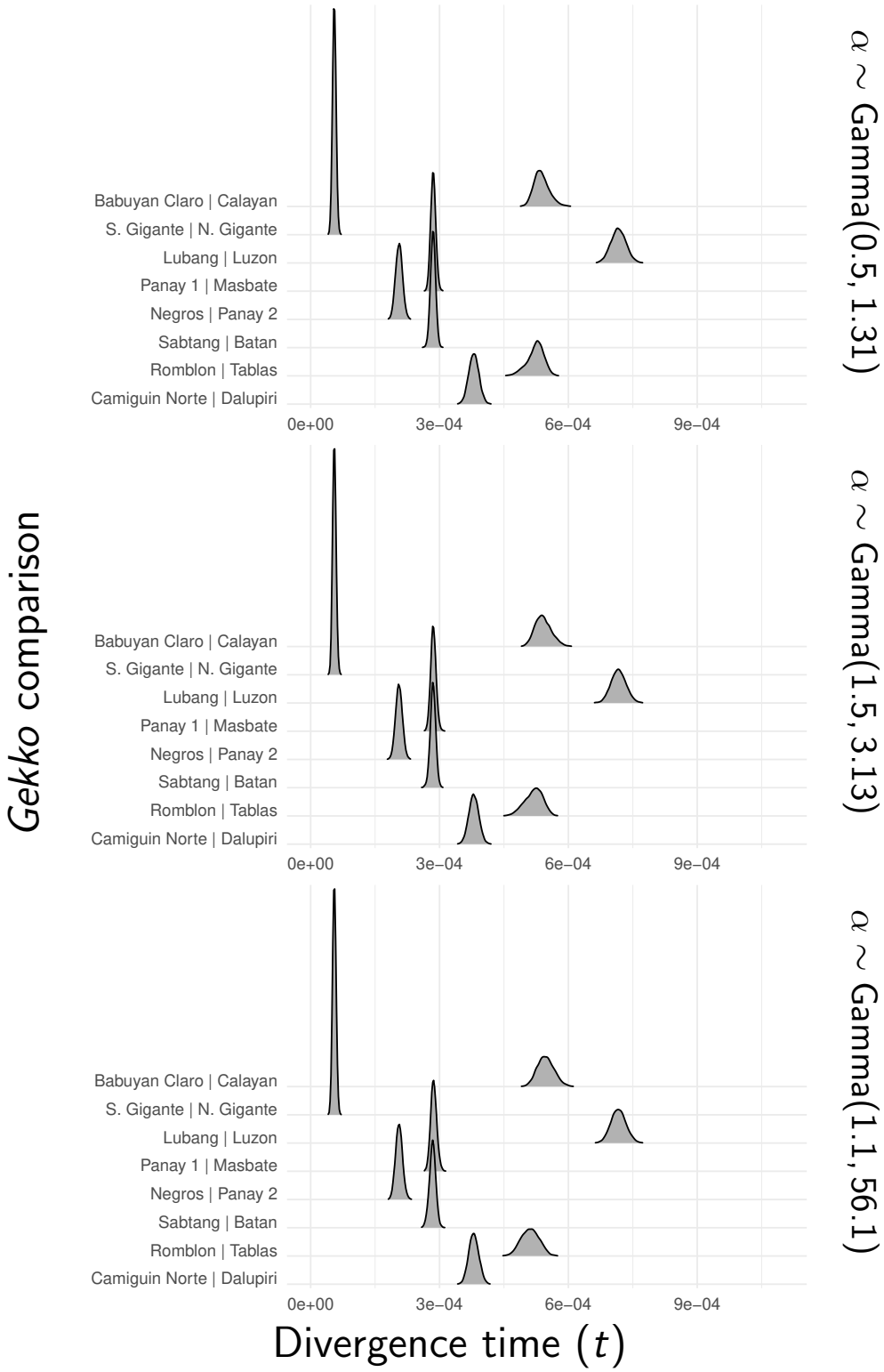

Figure S10. Approximate marginal posterior densities of divergence times for each pair of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

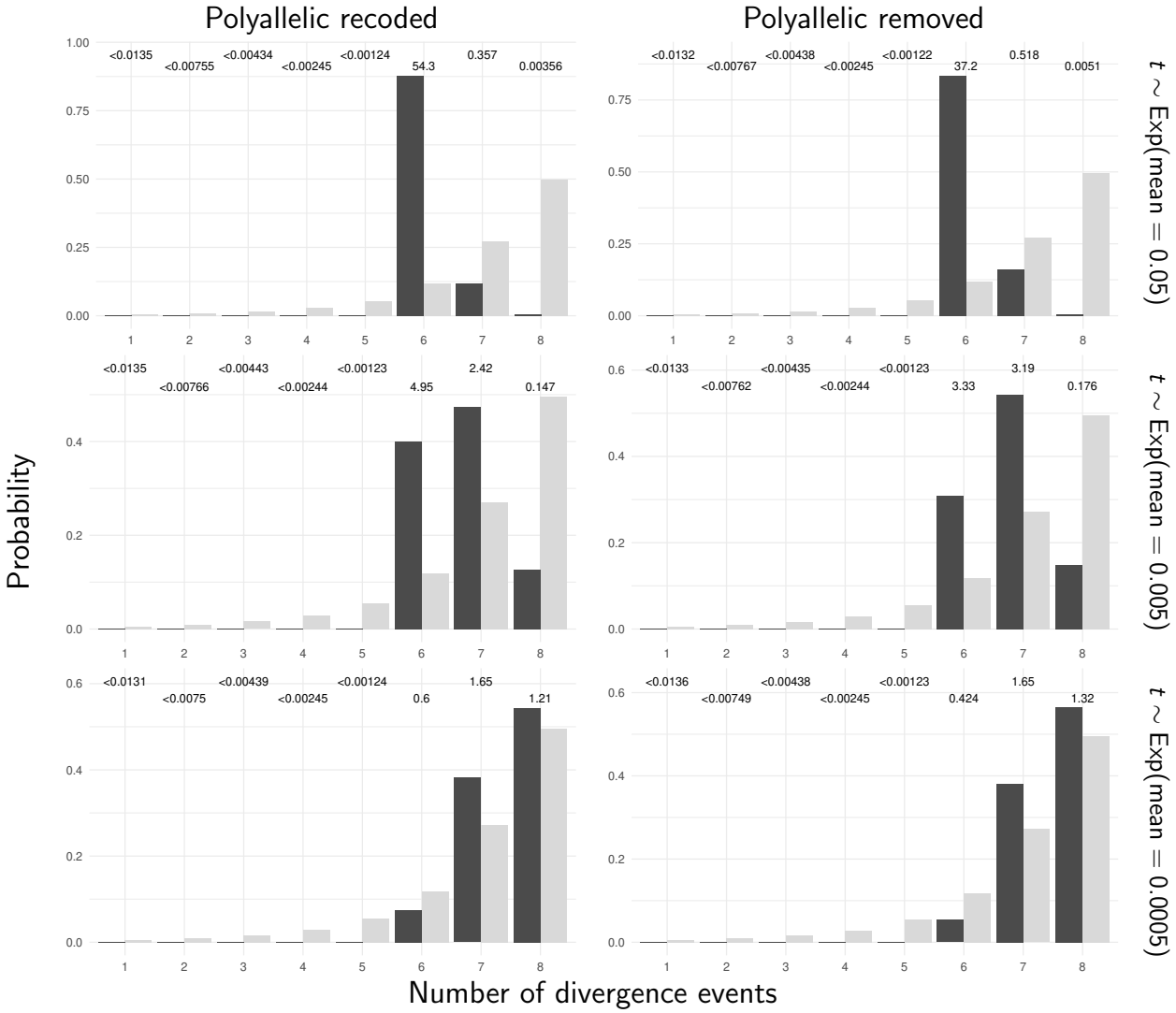

Figure S11. Approximate prior (light bars) and posterior (dark bars) probabilities of numbers of divergence events across pairs of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Bayes factors for each 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).

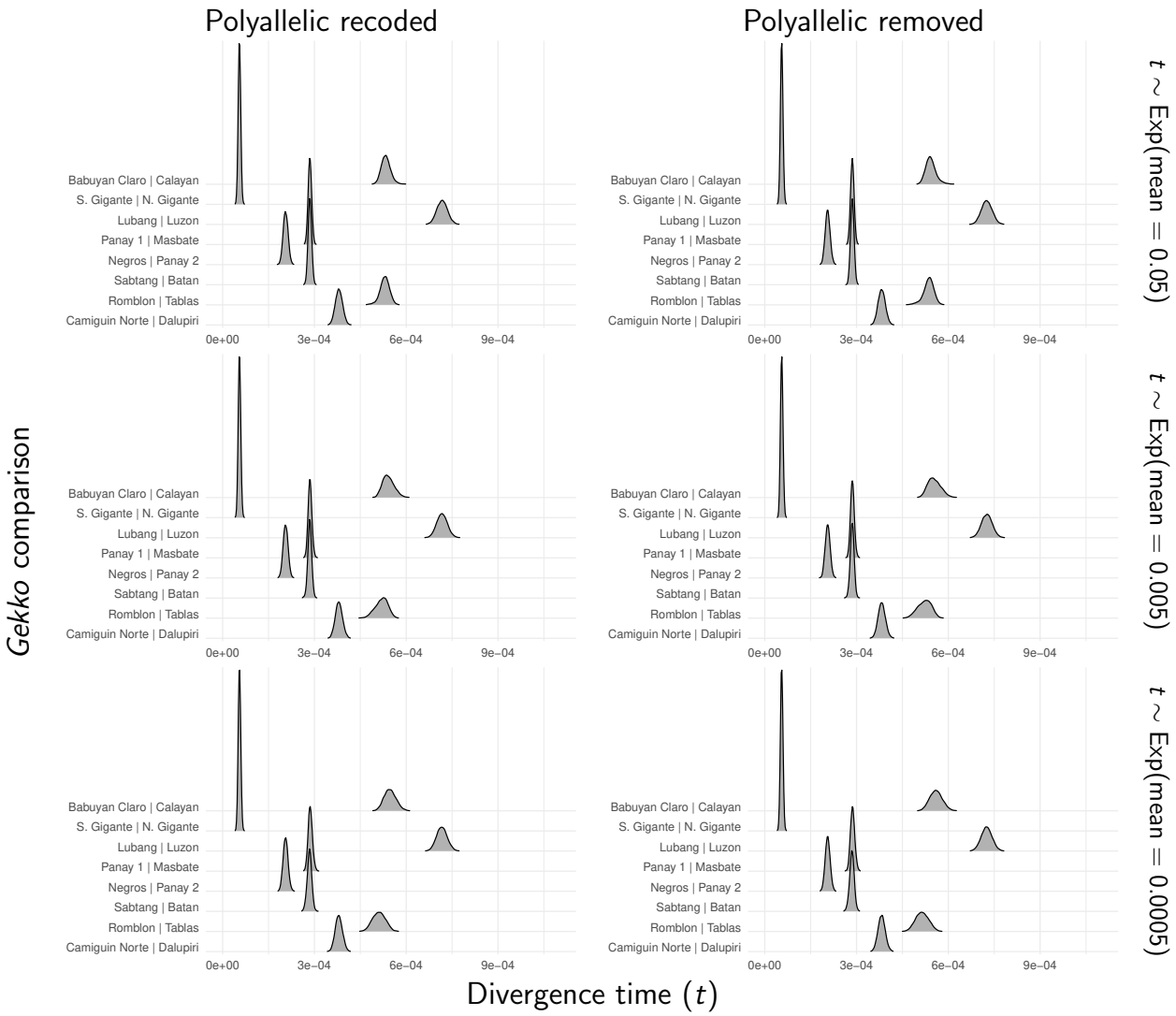

Figure S12. Approximate marginal posterior densities of divergence times for each pair of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and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

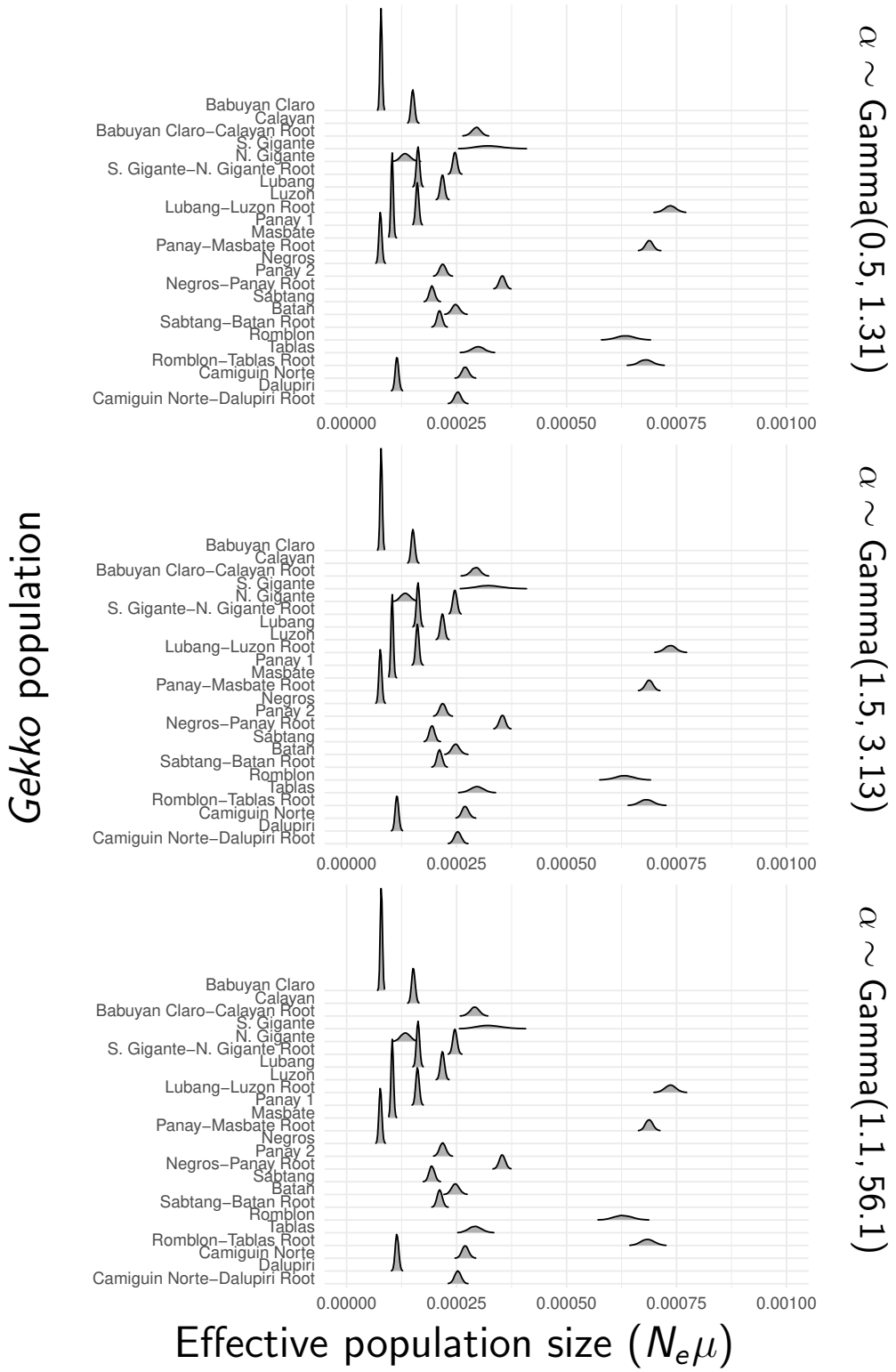

Figure S13. Approximate marginal posterior densities of population sizes for each pair of *Gekko* populations under three different priors on the concentration parameter of the Dirichlet process. Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

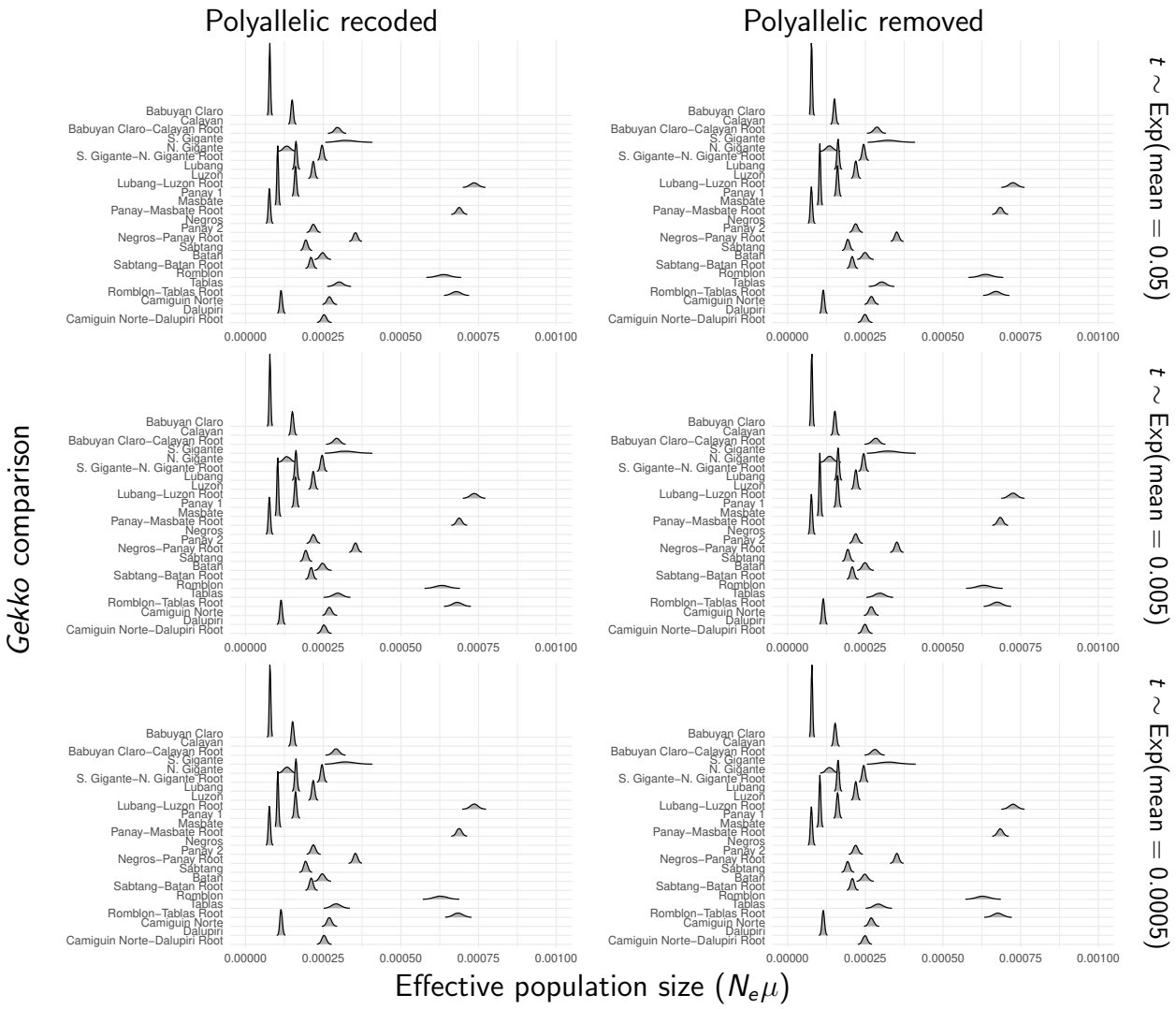

Figure S14. Approximate marginal posterior densities of population sizes for each pair of *Gekko* populations under six different combinations of prior on divergence times (rows) and recoding or removing polyallelic characters (columns). Figure generated with ggridges Version 0.4.1 (Wilke, 2018) and ggplot2 Version 2.2.1 (Wickham, 2009).

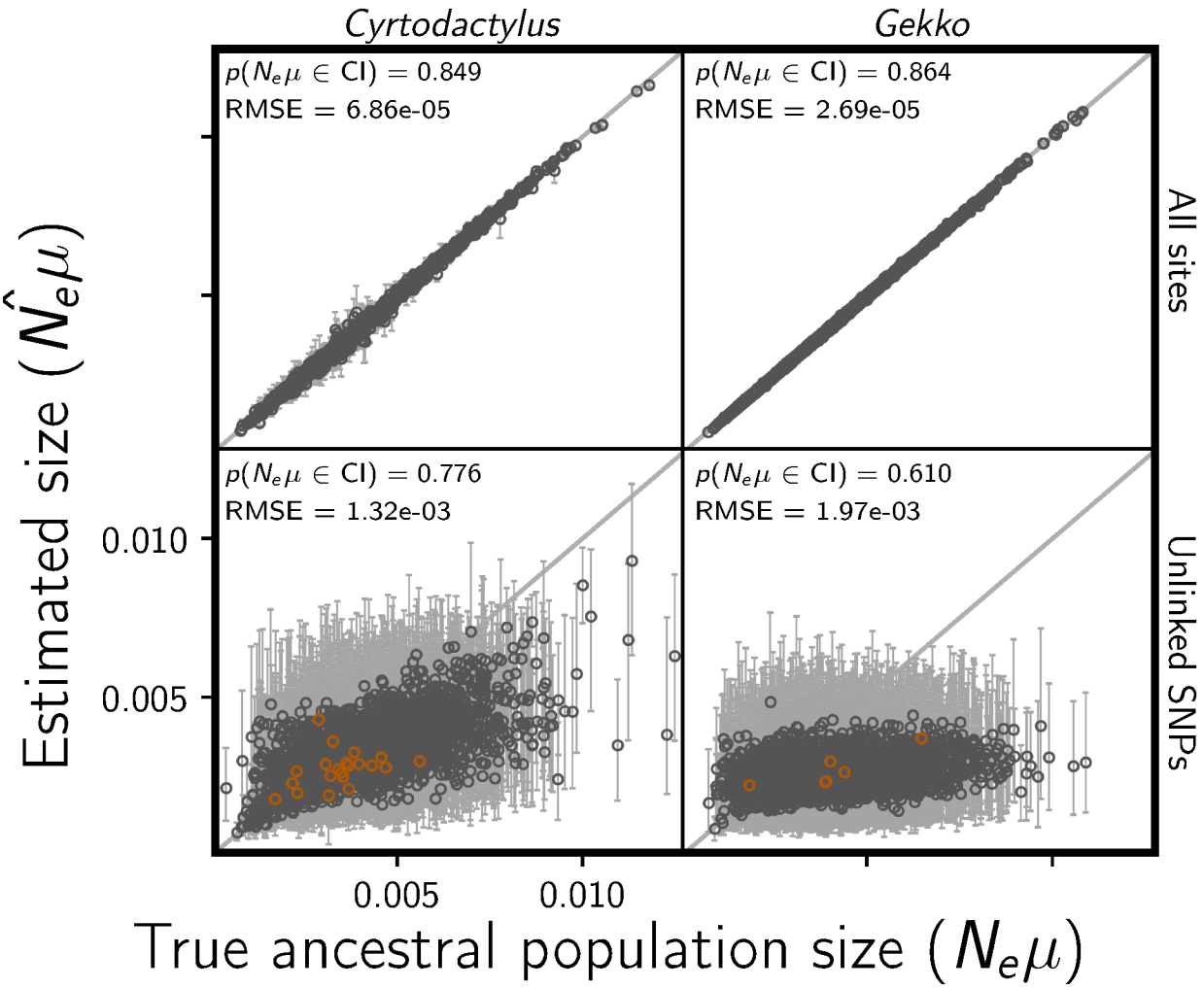

Figure S15. The accuracy and precision of `ecoevolity` estimates of the ancestral population size (scaled by the mutation rate) when applied to data simulated to match our *Cyrtodactylus* (left) and *Gekko* (right) RADseq data sets with all sites (top) or only one SNP per locus (bottom). Each circle and associated error bars represent the posterior mean and 95% credible interval. Estimates for which the potential-scale reduction factor was greater than 1.2 (Brooks and Gelman, 1998) are highlighted in orange. Each plot consists of 4000 estimates—500 simulated data sets, each with eight pairs of populations. For each plot, the root-mean-square error (RMSE) and the proportion of estimates for which the 95% credible interval contained the true value—$p(N_e\mu \in \text{CI})$—is given. Figure generated with matplotlib Version 2.0.0 (Hunter, 2007).

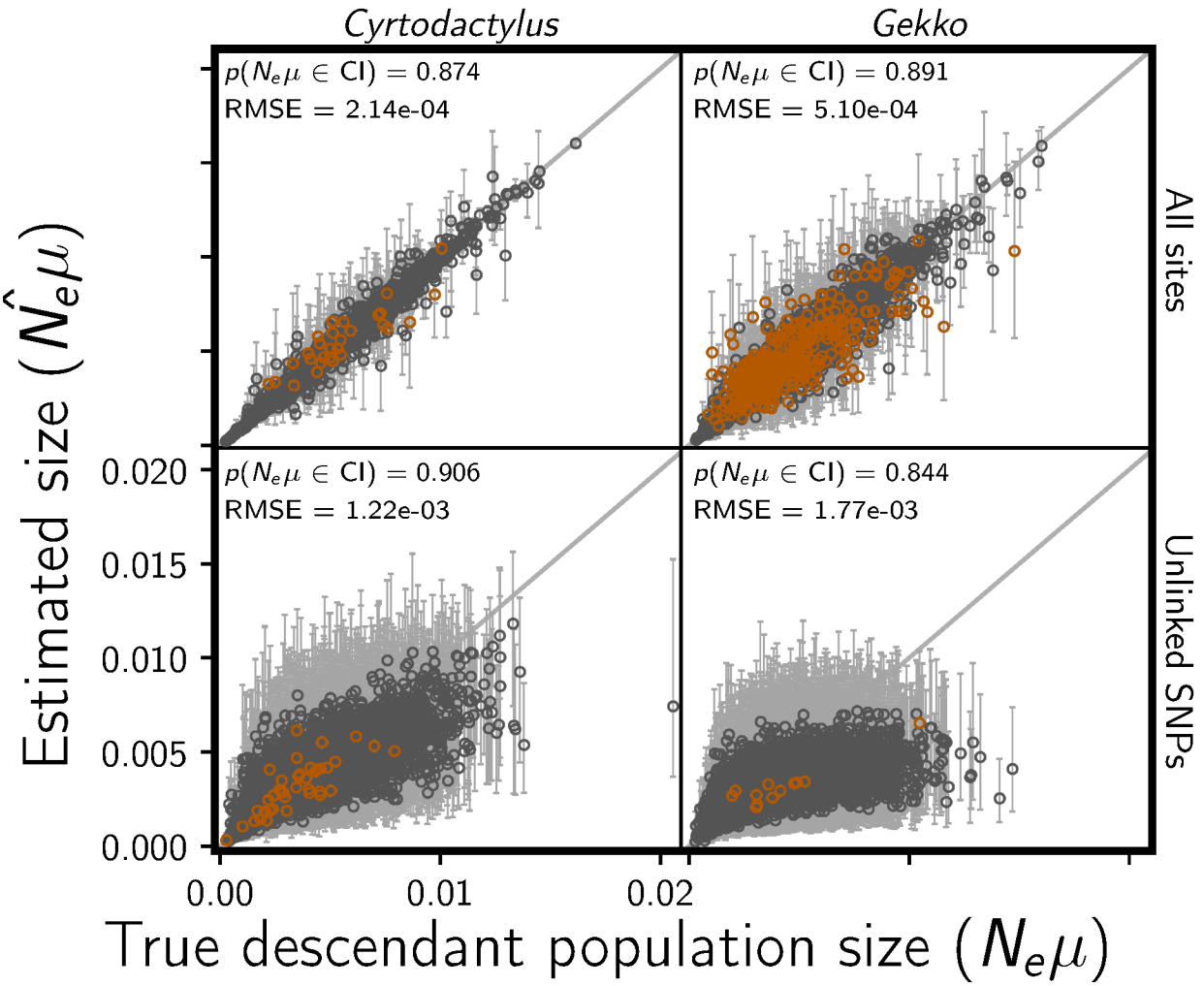

Figure S16. Accuracy and precision of `ecoevolity` estimates of the descendant population sizes (scaled by the mutation rate) when applied to data simulated to match empirical *Cyrtodactylus* (left) and *Gekko* (right) RADseq data sets with all sites (top) or only one SNP per locus (bottom). Each circle and associated error bars represents the posterior mean and 95% credible interval. Estimates for which the potential-scale reduction factor was greater than 1.2 (Brooks and Gelman, 1998) are highlighted in orange. Each plot consists of 8000 estimates—500 simulated data sets, each with eight pairs of populations. For each plot, the root-mean-square error (RMSE) and the proportion of estimates for which the 95% credible interval contained the true value—$p(N_e\mu \in \text{CI})$—is given. Figure generated with matplotlib Version 2.0.0 (Hunter, 2007).

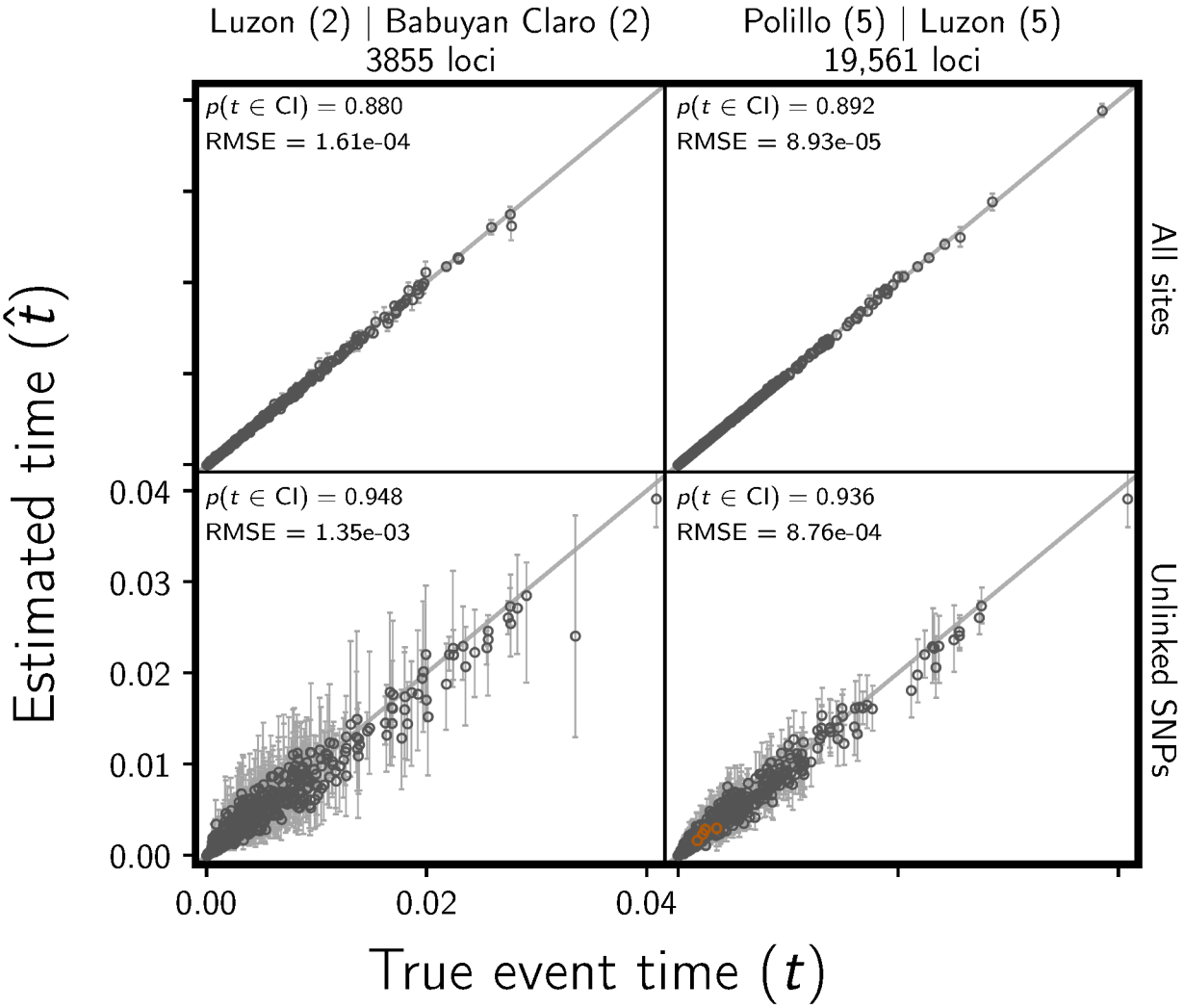

Figure S17. Accuracy and precision of `ecoevolity` divergence-time estimates (in units of expected subsitutions per site) when applied to data simulated to match empirical RADseq data sets sampled from the pairs of *Cyrtodactylus philippinicus* populations from the islands of (left) Luzon and Babuyan Claro and (right) Polillo and Luzon (a subset of the results plotted in Figure 6). Results are shown for `ecoevolity` analyses of data sets that contain all sites (top), or only one SNP per locus (bottom). The number of individuals sampled from each island population is indicated in parentheses at the top of each column of plots. Results for these two pairs of populations are plotted separately here to compare divergence-times estimated from data sets with large differences in the number of sampled individuals and loci. Each circle and associated error bars represents the posterior mean and 95% credible interval for the time that a pair of populations diverged. Estimates for which the potential-scale reduction factor was greater than 1.2 (Brooks and Gelman, 1998) are highlighted in orange. Each plot consists of 500 estimates—one from each of the 500 simulated data sets. For each plot, the root-mean-square error (RMSE) and the proportion of estimates for which the 95% credible interval contained the true value—$p(t \in \text{CI})$—is given. Figure generated with matplotlib Version 2.0.0 (Hunter, 2007).