COMMUNITY TREES: IDENTIFYING CODIVERSIFICATION IN THE PÁRAMO DIPTERAN COMMUNITY.

Running Title Estimating Community Divergence

Bryan Carstens[1,2], Michael Gruenstaeudl[3,4], Noah Reid[5,6].

[1]Department of Evolution, Ecology and Organismal Biology, 318 W 12th Ave., The Ohio State University, Columbus, OH, 43210, USA.

[1]carstens.12@osu.edu

[3] Institut für Biologie-Botanik, Dahlem Centre of Plant Sciences, Freie Universität Berlin, Altensteinstraße 6, 14195 Berlin, Germany.

[4] m.gruenstaeudl@fu-berlin.de

[5]Department of Environmental Toxicology, University of California-Davis, Davis, CA, USA.

[6]noah.reid@gmail.com

Abstract

Groups of codistributed species that responded in a concerted manner to environmental events are expected to share patterns of evolutionary diversification. However, the identification of such groups has largely been based on qualitative, *post hoc* analyses. We develop here two methods (PPS, K-F ANOVA) for the analysis of codistributed species that, given a group of species with a shared pattern of diversification, allow empiricists to identify those taxa that do not codiversify (i.e., "outlier" species). The identification of outlier species makes it possible to jointly estimate the evolutionary history of co-diversifying taxa. To evaluate the approaches presented here, we collected data from Páramo dipterans, identified outlier species, and estimated a "community tree" from species that are identified as having co-diversified. Our results demonstrate that dipteran communities from different Páramo habitats in the same mountain range are more closely related than communities in other ranges.

We also conduct simulation testing to evaluate this approach. Results suggest that our approach provides a useful addition to comparative phylogeographic methods, while identifying aspects of the analysis that require careful interpretation. In particular, both the PPS and K-F ANOVA perform acceptably when there are one or two outlier species, but less so as the number of outliers increase. This is likely a function of the corresponding degradation of the signal of community divergence; without a strong signal from a co-diversifying community, there is no dominant pattern from which to detect and outlier species. For this reason, both the magnitude of K-F distance distribution and outside

knowledge about the phylogeographic history of each putative member of the community should be considered when interpreting results.

KEYWORDS: comparative phylogeography, multispecies coalescent, Páramo, Diptera

One of the major objectives of historical biogeography is developing an understanding of how geographic ranges have evolved (Ree & Smith 2008). This process is initiated on the landscape scale, and thus phylogeographic studies offer important insights into the formation of broader biogeographic patterns (Avise et al. 1987). Investigations that compare phylogeographic patterns across species offer an important synthesis of patterns across species (Avise 2000), and serve as an initial step in the identification of cohorts of species that have codiversified (Smith et al. 2011). Most investigations have analyzed comparative data on a taxon-by-taxon basis, with a subsequent comparison of results. For example, many researchers have estimated gene trees from each species within a set of codistributed species and compared these estimates in a *post hoc* manner (e.g., Arbogast & Kenagy 2001). In this framework, evidence of similarity (e.g., genealogical congruence or similar values of parameter estimates) is interpreted as the result of a shared response to environmental events across lineages.

Comparative phylogeographic investigations that evaluate similarity in genealogical patterns or parameter estimates can be limited by the same factors that plague single species investigations. At neutral loci, the early stages of population divergence produce genealogical patterns that are subject to stochastic variance (Nordberg 1998), and thus the monophyly of alleles sampled from a particular geographic region is unlikely (Hudson & Coyne 2002). Although phylogeography has expanded beyond single locus data (Garrick et al. 2015), comparative investigations that rely on parameters estimated from neutral genetic loci remain subject to inference error. Parameters such as temporal divergence are difficult to estimate with a limited number of loci (Wakeley and Hey 1997; Edwards &

Beerli 2001), and parameter estimates often rely on insufficiently justified demographic models (Carstens et al. 2013). When inference in comparative phylogeography is driven by parameter estimates, the uncertainty in these estimates may be compounded across the constituent species.

In order to avoid the complications presented by *post hoc* comparisons, some researchers have followed the suggestion of Edwards and Beerli (2000) and treated codiversification as a null hypothesis. Sullivan et al. (2000) investigated the phylogeographic pattern in two species by asking if the genealogical pattern in one species was significantly different from that of a second species, but tested this using a phylogenetic approach that did not account for coalescent stochasticity. Other researchers have tested common phylogeographic hypotheses across multiple species, on a species-by-species basis (e.g., DeChaine and Martin 2005; Carstens and Richards 2007). Perhaps the most sophisticated approach to date is the hierarchical Bayesian model developed by Hickerson et al. (2006). Their method (msBayes) calculates the posterior distribution of the number of divergence events across some number of codistributed species and is commonly used to test if a set of taxa simultaneously diversified across a single barrier (e.g., Bell et al. 2011; Oaks et al. 2012).

While phylogeographic questions can sometimes be discretized to vicariance across a single barrier, most involve three or more distinct geographic regions. Such questions often concern the topology and timing of divergence across different geographic regions – an aimed shared by studies on historical biogeography (e.g., Morrone and Crisci 1995) - and, particularly in comparative phylogeographic studies, whether this diversification is similar across species. In such cases, it is possible to extend the multispecies coalescent

model (Degnan & Rosenberg 2009) to estimate the pattern of diversification in multiple codistributed species. For example, if some number of codistributed species have responded to shared environmental events in a concerted manner, a model of this codivergence would resemble a phylogenetic tree and would have important similarities to a species tree, as both are simplified conceptual models of diversification that reflect shared evolutionary history. In the case of the species tree, data are sampled from unlinked genetic loci that evolve within the confines of a shared evolutionary history, and discord between these loci is attributed to the stochastic process of gene coalescence. Other evolutionary processes that can lead to discord between gene trees and species trees (e.g., gene duplication/loss, lateral transfer or migration) are not incorporated into the model, because the species tree model is a compelling characterization of the evolutionary process (the history of organismal diversification through time) and because we are not yet able to include all of these processes in a single analytical framework.

In a manner analogous to the species tree, a *'community tree'* that models codiversification represents a compelling framework for comparative phylogeography. If data from unlinked neutral genetic loci are sampled across multiple species and analyzed in a framework that assumes shared evolutionary history, the resulting community tree would represent the estimation of the pattern and timing of diversification within a community. Similar to the species tree, estimates of branch length and branching pattern would benefit from a multilocus perspective. While the term *'community'* has a number of meanings in the ecological literature (see Morin 2008), it is used here to describe a set of samples from different species that are distributed in the same geographic area and potentially share a history of codiversification. In the same manner that a species tree

analysis assumes that the genes are sampled from within an organismal lineage, a community tree requires that the species included in the analysis have responded to historical events in a concerted manner (i.e., congruent patterns and timing of diversification). As a practical matter, this requirement would appear to prevent researchers from inferring community trees directly from the data, as identifying codiversification – a property required by the analysis – is the primary goal of the analysis and thus not known *a priori*. The community tree framework necessarily includes methods for identifying those taxa that have not diversified in a concerted manner (hereafter "outlier" taxa).

Here, we develop two such methods. We (*i*) describe an approach that measures the cost of the community constraint (i.e., the cost of assuming a concerted response across species) by calculating gene tree distributions under this constraint and comparing them to unconstrained gene tree distributions, and (*ii*) extend an approach to assessing the fit of the multispecies coalescent model using posterior predictive simulation (Reid et al. 2014; Gruenstaeudl et al. 2015) to the level of the community. When the results of these analyses are considered together, they allow us to identify species with a concerted response to historical vicariance. Once outlier species are removed from the dataset, the timing and pattern of divergence can be directly estimated using a community tree framework. After simulation testing to assess the performance of the proposed community tree framework, we apply our methods to an empirical data set of Central American Parámo Dipterans.

Parámo dipterans

Central American Páramos occur in the Cordillera Volcánica of Costa Rica and the Cordillera de Talamanca in Costa Rica and Panamá (Fig. 1A). These mountain ranges are areas of high endemism (Olson et al., 2001) and Páramo sky islands occur above ~3200m. The Páramo is a neotropical alpine habitat characterized by high levels of insolation, wide daily temperature fluctuation with frequent freezes and little seasonality (Luteyn, 1999). While the Talamancas are known for montane forest endemism at middle elevations, these small patches of alpine Páramo also harbor a unique biota, including two endemic plant genera, a bird, and a now extinct frog (Sklenar et al., 2005, Savage & Bolaños, 2009, Stiles & Skutch, 1989). Lineage diversification is evident within and among these ranges in amphibians (García-París et al., 2000, Streicher et al., 2009) and at least one bird (Stiles & Skutch, 1989). As in other alpine systems (e.g., Moritz et al., 2008), this biota is threatened by climate change through the process of mountaintop extinction.

The Páramo entomofauna is poorly known and not included in published accounts of several prominent Costa Rican biodiversity inventories (e.g., Colwell et al., 2008, Janzen et al., 2009), but is thought to be far less diverse than that of forested lower elevations (Janzen, 1973, Kappelle & Horn, 2005). Owing to an extreme climatic regime, it is likely to be highly endemic. We have focused on the Diptera of these habitats for several reasons. First, Diptera represent one of the most poorly studied of the hyperdiverse insect orders, particularly in the tropics (Brown, 1995). They may compose 10% of global animal diversity, but estimates suggest that only 8% of dipteran species are currently described (Brown, 1995). Second, Diptera encompasses a broad range of life histories (typical sampling yields detritivores, herbivores, parasitoids and predators), and a large diversity

of species from a given locality can be easily captured using sampling methods that are easily standardized (Brown, 2010).

METHODS

*BEAST (Heled & Drummond 2010) implements a multispecies coalescent model (MSCM) to estimate a species tree from sequence data. DNA sequence alignments are used to estimate a posterior distribution of gene trees with ultrametric branch lengths, along with a posterior distribution of species trees, which include divergence times and effective population sizes on each branch (Heled and Drummond 2010). We propose that this model can also be applied to data collected from a group of codistributed species that may have codiversified. A successful application of the MSCM to comparative phylogeographic systems would (i) enable outlier species to be detected and (ii) allow a community tree to be estimated from data collected from those species that are not identified as outliers. In this study, we apply the MSCM as implemented in *BEAST to mitochondrial data collected from codistributed Dipteran species. While independent species are expected to have different population sizes, we do not explicitly scale the data to reflect population size estimates for each species because we found that that such scaling did not dramatically influence our results (see simulation testing, below). We suspect that this results stems from some combination of phylogenetic conservatism in Páramo fly θ and the limited ability that one has to estimated $\theta = 4Ne\mu$ using single locus data (Felsenstein 2006). The decision to conduct a community analysis without scaling by population sizes is not without precedent (e.g., Dolman and Joseph 2012). Note that our method assumes that all shared polymorphism results from incomplete lineage sorting rather than gene flow.

*Posterior Predictive Simulation in *BEAST*

The first approach to identifying outlier species makes use of posterior predictive simulation (PPS), a method for assessing the statistical fit of a probabilistic model implemented in a Bayesian framework (Gelman et al. 2009). Reid et al. (2014) described an approach for testing the fit of the MSCM to the empirical data using PPS. Briefly, samples in PPS are taken from the joint posterior distribution of gene trees, and the probability of the data given the model is calculated following Rannala and Yang (2003). Data are then simulated under the model using corresponding draws from the posterior distribution of species trees, and the probabilities of these simulated data are calculated. Empirical data that fit the MSCM will thus be broadly similar (in terms of their probability) to data simulated under this model. Under a perfect fit between model and data we anticipate an overlap with zero of roughly 0.5 (i.e., approximately one-half of the values will be > 0.0). Empirical data that do not fit the MSCM are recognized when the difference between the posterior and posterior predictive distributions has a minimal overlap less than a predefined tolerance level, where this level functions in a manner analogous to a *P*-value in frequentist statistics (Gelman et al. 2009).

Gruenstaeudl et al. (2015) described an R package (P2C2M) that conducts the above analysis and present simulations that verify the efficacy of this approach. However, their simulation testing was limited to cases in which the MSCM was used to generate the data, or those where a MSCM with gene flow was the generating model. To be useful in the community tree framework, posterior predictive simulation should be able to identify cases of poor fit to the MSCM in which either the pattern or the timing of divergence among

taxa differ. In order to explore the performance of PPS in identifying outlier species in the community tree framework, we assessed P2C2M using simulation testing.

Measuring the cost of constraint

A second approach to identifying outlier species is to measure the cost associated with enforcing a community constraint on the data collected from each species. If a group of species has actually codiversified, the cost of such a constraint is expected to be minimal because the species trees within each species will share the same topology. If, by contrast, a particular species has diversified in a manner idiosyncratic to that of the community, we anticipate that the fit of the data under the (then incorrect) community constraint should be greatly diminished. In order to measure the cost of the community constraint, gene tree distributions are estimated from all loci within each species, both under a model of community diversification (i.e., using *BEAST) and without such a model (i.e., using BEAST on data from each species independently; Drummond et al. 2012). We then calculate the tree distances between the gene trees of the unconstrained and constrained posterior distributions using the Kuhner-Felsenstein distance metric (K-F distance; Kuhner & Felsenstein 1994). The K-F distance considers both topological pattern and branch lengths in the calculation of tree distances. The resulting K-F distance distributions are then compared using a set of n ANOVAs under mixed effects models, whereby n is the number of species s under study. Specifically, for each species s_i of the community tested, an ANOVA is conducted using a model in which all species other than s_i are grouped into a block (i.e., assumption of a fixed effect), but all species-specific effects are modeled as uncorrelated with the tree distances (i.e., assumption of a random effect). Each effects model, therefore,

tests if a set of species s_{n-i} collectively displays a lower cost of constraint than the full community species set s_n. Our model can be formalized as: (K-F distances) ~ (shared community history) + ε(idiosyncratic species variation), where epsilon represents all components of a species' variation that are not caused by the shared community history. KF-distances and ANOVAs were calculated as part of an automated pipeline available through the R package “CommT", with the ANOVA calculated using the function "aov" in the default "stats" package of R (R Development Core Team).

Simulation testing

To test our proposed framework for assessing the cost of a community constraint, we simulated data sets under models of community diversification as well as data where some of the constituent species were outliers (i.e., one or more species did not have the same evolutionary history). In each treatment, we simulated data for species trees with 5 tips, with *k* samples selected from a discrete uniform distribution (5, 15) in each species. This range was informed by the sampling in the empirical data. Phylogenetic trees were simulated under a Yule model of tree evolution (Yule 1924) at a total tree depth of 20N generations, a birth rate = 1.0 and death rate = 0.5, using the R package "ape" (Paradis et al. 2004). These species trees were used to simulate population level data that correspond to the data collected from individual species. The same species tree was used for species that were modeled as having a shared evolutionary history, while different species trees (chosen at random) were used for species that were modeled as outliers. For each simulation, ten data sets were simulated as follows: ms (Hudson 2002) was used to simulate coalescent genealogies with $\theta = 4Ne\mu = 1.0$, Seq-Gen (Rambaut and Grassley

1997) was used to simulate 658 bp of sequence data (HKY model; transition/transversion ratio = 3.0; base frequencies = 0.3, 0.2, 0.2, 0.3; substitution rate = 0.004) and a number of segregating sites with a mean and a variance consistent with the empirical data described below (Parámo fly COI mtDNA).

The data were formatted for community analysis using a Python script (available as supplemental material). A community tree from these data was estimated using *BEAST (Heled & Drummond 2010), and gene trees from each species were independently estimated using BEAST (Drummond et al., 2012). These gene tree distributions (i.e., one constrained to the community tree and estimated in *BEAST; a second unconstrained and estimated in BEAST) were subsequently compared by calculating distributions of K-F distances. In each case, analyses were conducted for 50 million MCMC generations, with the first 10 million discarded as burn-in, and all remaining generations subsampled to generate a set of 1000 genealogies. Upon the estimation of the posterior distributions, each data set was evaluated using posterior predictive simulation using the '*lcwt*' summary statistic (Gruenstaeudl et al. 2015), and the cost of constraint was measured by computing K-F distances (Treatment A). We then repeated the above process, but swapped a single data set across simulations to represent a case where 9 out of 10 species in a community had codiversified (Treatment B). Next, we swapped 2 of the 10 species (Treatment C) and then 4 of the 10 species (Treatment D) to represent cases where 8/10 and 6/10 of the species codiversified, respectively. Finally, we conducted simulations where each of the 10 species has an idiosyncratic evolutionary history (Treatment E), in order to understand the results produced by an inappropriate modeling of species with distinct evolutionary histories as a single community tree.

It is important to understand the rates of different types of error in identifying outlier species, under both the PPS and the K-F ANOVA methods to detect outlier species. Three indicators of performance are of interest: (i) the rate of accurate detection of species with an evolutionary history distinct from that of the community, (ii) the rate of failed detection of species with an evolutionary history distinct from that of the community, and (iii) the false identification of species as outliers from a co-diversifying community when they in fact share a history of codiversification. Hereafter, these are referred to as (i) accurate detection, (ii) failed detection, and (iii) false identification. Treatments B-D (above) are designed to allow us to estimate these rates given differing levels of codiversification, while Treatment A can only be used to measure the false identification rate. Since we need to understand of the effects of these types of errors on the distributions of gene trees under the community constraint, we also examined how the K-F distance distributions are affected as the number of codiversifying species in the modeled community changes.

Field Sampling of Páramo dipterans

We sampled three habitat islands spanning ~60% of the distribution of Páramo islands in Costa Rica (Fig. 1A). The three sampling sites are located between 3300m and 3400m at Volcán Irazú (VIR) in the Cordillera Volcánica, and at Cerro de la Muerte (CLM) and Cerro Chirripó (CHI) in the Cordillera de Talamanca. The sites are arrayed from northwest to southeast (VIR – 46km – CLM – 28km – CHI). VIR is separated from CLM and CHI by the central valley, and the pass between them is ~1500m, while the pass between CHI and CLM is ~2700m (Fig. 1). Sampling methods at each site include malaise traps, flight intercept

traps, pan traps and sweep transects. Samples were retrieved daily between July 25th and August 15th, 2009, and stored in lots by trap and day in 90% ethanol until processing.

Specimen Processing, DNA Extraction and Sequencing

Individual samples were identified to morphospecies, and classified to the family level (Brown, 2010). For DNA extraction we chose 5-10 specimens at random from each morphospecies within each lot. Whole genomic DNA was extracted from entire individuals using either Qiagen DNeasy kits or the protocol of Ivanova et al. (2006), both of which are based on binding DNA to silica membranes in 96 well plates. Exoskeletons remain intact during this process and were retained as voucher specimens. A 648bp region of Cytochrome Oxidase I was amplified using standard primers (Vrijenhoek, 1994), and PCR products were sent to Beckman Coulter Genomics for Sanger sequencing. We edited chromatograms individually in Geneious version 5.04 (Drummond et al., 2011) and then aligned and verified each polymorphism by eye. We identified and discarded potential nuclear pseudogenes by checking each sequence for indel mutations that were not in groups of three nucleotides and stop codons. We discarded any sequences with multiple peaks causing ambiguity in base calls. Finally, we aligned all retained sequences using MUSCLE (Edgar, 2004) and constructed a neighbor-joining tree. We verified the identification of any specimen with a sequence that did not cluster with other specimens from its sampling locality or family. We generally were able to resolve these as cases of misidentification, mis-sorting, or contamination from an adjacent well in a 96-well plate. We also retrieved the closest BLAST hit in GenBank for each sequence using BioPython

(Cock et al., 2009) to look for irregularities and to determine if any close relatives have been sequenced.

Delimitiation of Páramo flies

DNA sequences were clustered into operational taxonomic units (referred to hereafter as species) using ABGD (Automated Barcode Gap Detection) (Puillandre et al., 2012), a non-parametric method based on the idea that pairwise genetic distances among individuals within species should be smaller than those among species. It searches for such "gaps" in the distribution of pairwise distances using the slope of ranked pairwise distances. It does this iteratively by first partitioning the entire dataset using the largest detected gap and then searching each cluster for gaps. Because these gaps are expected to appear by chance in intraspecific coalescent trees, the algorithm requires the user to provide a data-set wide estimate of maximum prior intraspecific divergence, P, that is used to calculate a value of $\theta=4Ne\mu$. This value of θ is then used to calculate the maximum expected intraspecific divergence in order to cluster sequences. There is no objective means to select the "best" P *a priori*, and different P produce different, but reasonable results. This is consistent with the expectation that there should be real uncertainty about assigning species limits from single-locus DNA sequence data, particularly in cases when genetic divergence among populations is high. Therefore, we first considered varying values before settling on a conservative value that we expect will error on the side of *lumping* (i.e., samples from closely related species may be included in a single taxon) rather than *splitting* (separating into separate species samples from what are actually the same species).

After species were identified using the ABGD, we adopted an iterative approach to infer a community tree. We first conducted analyses using data from all the delimited species. We then identified outlier species using the PPS and the ANOVA with the K-F distributions, removed the outliers, and reran the community tree analyses. We stopped this process when we no longer detected an outlier species.

RESULTS

Simulation testing

Simulation testing was conducted to evaluate how effective posterior predictive simulation and the ANOVA on K-F distances were at identifying species that do not share a common history with the community (Fig. 2). False identification rates were generally low in both the PPS and the ANOVA on K-F distances, except in the case of treatment A in the latter, when 0.65 of the treatment replicate analyses suggested that there was an outlier species when in fact none were present (Table 1). The ANOVA on K-F distances was much more effective in accurate detection of actual outlier species, particularly in cases where there were 1-2 such species were simulated (Table 1). While PPS was not as effective at accurate detection of outliers in these treatments, it was more effective when the number of outlier species increased to 4. In both measures, rates of failed detection of actual outliers increased with the number of outlier species in the analysis; in treatments C and D, both were usually able to detect one of the possible outliers, but not all of the possible outliers (Table 1). To summarize, in cases where no species (treatment A) or a single species (treatment B) are outliers of the dominant pattern of diversification, both measures have low false identification rates, and the ANOVA on K-F distances performs well. In the case

where no pattern of community diversification exists (Treatment E), neither measure performed well. In this case, however, the inappropriateness of the community model is easily recognized via examination of the K-F distributions, which greatly exceed 0.2 in these cases (Fig. 3; Supplemental Fig. 1). Also note that we observed no cases where a false positive result was found for the same species using both the PPS and K-F difference distribution.

Data collection from Parámo dipterans

Over 2700 dipteran samples were collected from three Páramo sites (Table 2). For a subset of samples, the mitochondrial gene COI was sequenced, resulting in 1728 high-quality sequences. These sequences represent 30 families with Muscidae, Tachinidae, and Sciaridae being the most diverse (see Reid 2012 for more details).

Delimitiation of Páramo flies

We tested a range of prior intraspecific divergences in ABGD, from P=0.001 to P=0.36. At each extreme, results seemed unrealistic, resulting in 740 and one species, respectively. In the case of P=0.001, species in our estimated trees were not monophyletic and most morphospecies were broken up. Between P=0.0017 and P=0.0215, the algorithm estimated between 334 and 283 species. We took this long area of relative stability to be an indication that this range of values was reasonable. These P values result in estimates of θ from 0.0006 to 0.0033, which are small, but comparable to values published from across animals (Bazin et al., 2006, Wares, 2010). Because there is no straightforward way to assess which value in this range is optimal, we chose to conduct all analyses on two partitions, a split

partition with 334 species at θ=0.0017 and a lumped partition with 283 species at θ=0.0215. We then evaluated the membership of these putative species, and identified 10 that contained samples from each of the three Páramo regions. Note that a large proportion of the ~280 species identified in this analysis were sampled from a single locality, and thus did not contribute to the community tree analysis.

Community tree analyses

We identified 10 species that were present in each of the three sampling sites, and used these species for the community tree analysis. Results from PPS (Table 3) indicated that species #193 does not fit the community diversification model. Results based on the K-F ANOVA are consistent with the PPS; they also suggest that including this species in a model of community diversification is unwarranted. After species #193 was removed from the analysis, we re-estimated the community tree and repeated the above analyses. No species were found to be a poor fit to the community tree model using PPS (Table 2), but a single species was found to have a gene tree distribution that was substantially changed under the K-F ANOVA. Given the high rate of false identification in the simulation testing when using the ANOVA on K-F distance distributions, we retained this species when inferring the community tree (Fig 4).

DISCUSSION

Community phylogeography

In this investigation, phylogeographic analysis of COI data collected from a community of Páramo flies was conducted under the assumption of codiversification. We first estimated a

community tree given the data from ten species, and then assessed the fit of this model to the data using PPS and by measuring the cost of constraint using an ANOVA on K-F distance distributions. Our results indicate that one species (#193) is a clear outlier, in inference that is not clear based on a visual examination of the gene trees (Supplemental Figure 2). Before we assess the biological implications of our results, it is important to consider the methodological assumptions of the community trees approach proposed here.

Comparative phylogeography typically builds on and expands single-species investigations conducted in well-studied regions, such as Europe (e.g., Hewitt 2004), the warm deserts in North America (Riddle et al. 2000), or the Australian wet forest (Hugall et al. 2002). Even as comparative studies began to incorporate coalescent models (e.g., Carstens et al. 2005; Hickerson & Cunningham 2005) and to explicitly focus on ecological communities (Garrick et al. 2008; Carnaval et al. 2009; Smith et al. 2011; Espindola et al. 2014), most comparative phylogeographic investigations proceed by analyzing data from each species independently before conducting *post hoc* comparisons across species. While it has been argued that the power of comparative phylogeographic investigation comes from the identification of common patterns across such *post hoc* comparisons (e.g., Bowen et al. 2014), little attention has been devoted to assessing the limits of comparative phylogeographic inference, in particular how similar the results from multiple species should be before concordance is inferred. So long as such inferences are based on qualitative inference from the data, overinterpretation is likely (Knowles and Maddison 2002).

Our goal here is to explicitly model codivergence, and attempt to identify species that do not fit this model. From the perspective of community phylogeography, the

diversification of a community is modeled (by default) in a manner analogous to the species tree. The community tree assumes that the included species have diversified in a concerted manner in response to shared historical events, and that the timing and pattern of diversification in the resulting species trees are congruent. Unlike the species tree, where the presence of loci that exist in the same genome is assumed to be sufficient justification for their inclusion in a species tree analysis (but see Reid et al. 2014), it is important to evaluate the appropriateness of the set of species that are included in the community tree analysis.

Assessing the fit of the community tree model to empirical data

We introduce two statistical tools to assess the fit of the community tree framework to our empirical data: PPS and an ANOVA on K-F distance distributions. The former measures the deviation of the observed posterior distributions to the posterior predictive distributions expected under the community tree model, while the latter measures the significance of the effect on gene tree topology imposed by the community-specific topological constraints. Posterior predictive simulation, which was introduced to evolutionary genetics by Huelsenbeck et al. (2001), has experienced a recent surge in its application to evolutionary questions (e.g., Joly 2012; Brown et al. 2014; Lewis et al. 2014; Reid et al. 2014), and the variety of these applications suggest that PPS is a flexible technique for assessing model fit. We modified the R package P2C2M to conduct the PPS on community data. Our simulation testing results suggest that PPS, by itself, is only moderately successful in identifying species that do not fit the community tree model. It performs fairly well in cases where the majority of species share a common pattern, with only one or two species deviating from

the dominant history in the community, but less so as the number of outlier species increases. This is likely a function of the corresponding degradation of the signal of community divergence; without a strong signal of some diversifying community, there is no dominant pattern from which to detect and outlier. This is also likely why we cannot, with these statistical tests, detect outliers in the case where there is no common pattern (e.g., treatment E). The degree of topological similarity and internal branch lengths of the individual species trees input into the community tree analysis likely influence the ability of these methods to detect outlier species. An examination of the species histories used to generate the data of treatments B and C indicates that, in cases where PPS was not able to detect outlier species, these outliers had a species tree topology similar to that of the modeled community tree topology. For example, in several cases we observed that the outlier species had topologies that differed from the dominant topology of the community at nodes with relatively short internal branches. In such cases, it appears that the resulting data lacked a sufficient number of synapomorphic substitutions to correctly identify missing model fit. This is perhaps not surprising, as the MCMC machinery in *BEAST is designed to optimize the probability of the data given a single evolutionary history, and in these cases it is likely that the branch lengths of the community tree would be affected by an outlier, but not to an extent where PPS could easily identify this species.

In the context of phylogeographic investigations, PPS alone is not sufficient for identifying outlier species. Our results suggest that a quantification of the cost of enforcing the community tree on the gene tree distributions estimated across species provides vital information, that in combination with the PPS enable outlier species to be recognized. The approach that we develop here is to (i) calculate gene tree distributions from the same

species under the community constraint and independent of this constraint, (ii) compare these distributions by calculation K-F tree distances, and (iii) identify distributions that represent a significant outlier using an ANOVA. In addition to the statistical test that enables us to identify the most significant outlier species (if present), the magnitude of the K-F distance distributions also contains useful information. In our simulation testing, K-F distance distributions have median values < 0.2 when all species diversify as part of a community, and > 0.4 when there is no common pattern to the diversification (Fig. 3; Suppl. Fig 1). In the absence of a common pattern of diversification, the community tree estimated from the data approaches a star-like phylogeny, making it difficult to detect outlier species with PPS because such a phylogeny offers a marginal fit to data from nearly any species. However, these cases are characterized by very large average values of K-F distance distributions (Supplemental Fig. 1E), presumably because the gene tree distributions are distorted by the community constraint and thus are very different from the gene tree distributions estimated from the same data absent this constraint. This characteristic makes if very unlikely that researchers will infer a community when the constituent species lack a common pattern. We encourage researchers who apply our method to empirical systems to consider the results of both PPS and K-F ANOVA in identifying outlier taxa, in addition to results of other phylogeographic anlayses.

Insights into the evolution of Páramo entomofauna

We applied the community tree approach to investigate an understudied entomofauna in the Páramo of Costa Rica. This high-elevation habitat is isolated and very distinct from the surrounding habitat types, and harbors a large degree of endemism. The nearest extant

alpine habitats are located in southern Guatemala (~900km distant) and Colombia (~650km distant). Intervening mountain ranges are low (~2000m) and separated by deep valleys (<1000m). During the Pleistocene climate oscillations, montane habitat zones fluctuated over 1000 vertical meters lowering the Páramo zone to nearly 2000m elevation (Islebe & Hooghiemstra, 1997). This may have resulted in the physical connection of many currently isolated patches while others remained isolated, but the alpine system as a whole remained isolated from northern Central and South America. However, the community tree estimated from the Páramo fly data suggests that geological affinities are more important in structuring this community than the Pleistocene climate fluctuations. The Cerro de la Muerte (CLM) and Cerro Chirripó (CHI) sampling sites are sister in the community tree, with those from the Volcán Irazú (VIR) more distantly related. CLM and CHI are part of the same mountain chain (i.e., the Cordillera de Talamanca), while VIR belongs to the Cordillera Volcánica (Fig. 4). The Talamancas are thought to be young, reaching elevations capable of sustaining alpine habitat between 1-2 mya (Driese et al., 2007; but see Montes et al., 2012), and the branch lengths of the community tree are generally consistent with this timing (assuming that the mutation rates in the Páramo flies are similar in magnitude to those of *Drosophila*). The community tree analysis provides a fundamental improvement in the phylogeographic inference, as an examination of gene trees alone does not suggest that 9 of these species have a common pattern of diversification (Suppl. Fig. 2). However, it is also not entirely clear how many Páramo dipterans share this history of codiversification, because so many of our delimited species were only sampled from one location. It could be that these species do not occur in the other localities, or it could be that they are sufficiently rare as to not have been sampled.

At present, our approach to identifying outlier species and estimating a community tree does not explicitly incorporate species-specific θ values. We suspect that it will be important to expand of our approach to include the estimation of species-specific θ values, particularly when multilocus data are used to estimate species trees within each species prior to a community tree analysis. However, failing to do does not appear to affect the analysis of the Páramo fly data, as we identify the same outlier species with and without scaling (not shown). In addition, the results of simulation testing did not change appreciably when input data were scaled prior to the estimation of the community tree (Supplemental Table 1). It is unclear whether this results from phylogenetic conservatism of effective population sizes in Páramo dipterans, the use of mitochondrial data only from each species, or from some bias introduced by choices made during the ABGD species delimitation analysis. Further simulation testing is necessary to characterize the influence of parameterizing species-specific θ values in the community tree analysis.

Future directions

Community phylogeography extends the species tree analytical framework to include multiple codistributed species, and thus can be used to identify species that have diversified in response to the same historical events. Our method is similar in spirit to that of Dolman and Joseph (2012), who identified groups of species using msBayes and then estimated divergence times of these groups using a coalescent isolation with migration model. It could also be considered a phylogeographic analog to historical biogeographic analyses such as area cladograms or vicariance biogeography (e.g., Wiley 1988), in that both treat geographical areas as OTUs in a multispecies analysis that is interested in the

inference of historical processes. While we have applied our approach to organisms that, as members of a phylogenetic clade share many aspects of their life history, a natural application of community trees may be systems where species share specialized obligatory interactions such as pollination (e.g., Smith et al. 2011) or parasitism (e.g., Whiteman et al. 2007).

Genomic data are the future of phylogeographic studies (e.g., Garrick et al. 2015), and the community tree framework should be expanded to explicitly include multilocus data from each species. We believe that such an expansion is possible, and are working to expand this method by building software that estimated species trees from gene trees within each species, concurrently with a community tree from these species trees. However, the most immediate application of the community tree is likely to be the reanalysis of existing single-locus data sets. With over 20,000 investigation published to date (Web of Science search of title=phylogeograph* conducted March 8, 2016), the reanalysis of existing data may allow us to address the most compelling question that underlies comparative phylogeography: whether geographic factors, intrinsic features of organismal biology or ecological relationships can best explain why some species exhibit phylogeographic congruence while others do not (Smith et al. 2011).

Conclusions

The community tree framework provides a novel viewpoint in comparative phylogeography. It is not intended to replace *post hoc* comparisons of results from individual species, but offers a complimentary perspective that should be most valuable in comparison of species that share a distinct habitat and have similar life history

characteristics. In particular, by modeling a common history and explicitly seeking to identify species that do not share this history, the community tree framework offers an approach to the identification of *evolutionary communities* (*sensu* Smith et al. 2011) – species that share both ecological interactions and evolutionary history. Thus, community trees may provide increased integration to comparative phylogeography and community ecology. While we suspect that improvements to our approach are both possible and required, the community tree framework represents a novel direction for the field of phylogeography.

Acknowledgements

We thank members of the Carstens Lab, M. Daly, B. Titus, G. Silva-Arias, and D. Weisrock for conversations related to this topic. We thank AE Burbrink and three anonymous reviewers for comments that improved previous versions of this manuscript. Funding was provided by a grant from the American Alpine Club to NMR and start-up funds from The Ohio State University to BCC. This work was supported in part by an allocation of computer resources from the Ohio Super-computer Center.

Data Archiving

Sequence data generated as a part of this work have been submitted to Dryad under access number DOI: 10.5061/dryad.r5nf0. The computational steps to measure and visualize the cost of community constraints were summarized in the R package "CommT", which is available via CRAN: http://cran.r-project.org/web/packages/CommT/index.html

Literature Cited

Arbogast, B. and G. J. Kenagy. 2001. Comparative phylogeography as an integrative approach to historical biogeography. J. Biogeography 28:819-825.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489-522.

Avise, J. C. 2000. Phylogeography: The history and formation of species. Harvard Univ. Press, Cambridge, MA.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312: 570-572.

Bell, R. C., MacKenzie, J. B., Hickerson, M. J., Chavarria, K. L. Cunningham, M., Williams, S., and C. Moritz. 2011. Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. Proceedings of the Royal Society B 279:991-999.

Bowen, B. W., Shanker, K., Yasuda, N., Malay, M. C. D., von der Heyden, S., Paulay, G., Rocha, L. A., Selkoe, K. A., Barber, P. H., Williams, S. T., Lessios, H. A., Crandall, E. D., Bernardi, G., Meyer, C. P., Carpenter, K. E., and R. J. Toonen. 2014. Phylogeography unplugged: Comparative surveys in the genomic era. Bulletin of Marine Science 90:13-46.

Brown, B. V. 2010. Manual of Central American Diptera. NRC Research Press.

Brown, J. M. 2014. Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. Systematic Biology 63:334–348.

Brown, J. H. 1995. Macroecology. University of Chicago Press.

Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz. 2009. Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. Science 323:785-789.

Carstens, B. C. and C. L. Richards. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. Evolution 61:1439-1454.

Carstens, B.C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. A. Koch, C. D. McMahan, B. J. Nelson, C. E. Newman, J. D. Satler, G. Seeholzer, K. Prosbic, D. C. Tank, and J. Sullivan. 2013. Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. Molecular Ecology 22:4014-4028.

Carstens, B.C., S. J. Brunsfeld, J. R. Demboski, J. D. Good, and J. Sullivan. 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. Evolution 59:1639-1652.

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and de Hoon, M. J. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422-3.

Colwell, R. K., G. Brehm, C. L. Cardelus, A. C. Gillman, and J. T. Longino. 2008. Global warming, elevational range shifts and lowland biotic attrition in the wet tropics. Science 322:258-261.

DeChaine E.G., and A. P. Martin. 2005. Historical biogeography of two alpine butterflies in the Rocky Montains: broad-scale concordance and local-scale discordance. J. Biogeography 32:194301956.

Degnan J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecology Evolution 24:332-340.

Dolman, G. and Joseph, L. 2012. A species assemblage approach to comparative phylogeography of birds in southern Australia. Ecology and Evolution, 2: 354–369.

Driese, S. G., Orvis, K. H., Horn, S. P., Li, Z. H. and Jennings, D. S. 2007. Paleosol evidence for Quaternary uplift and for climate and ecosystem changes in the Cordillera de Talamanca, Costa Rica. Palaeogeography Palaeoclimatology Palaeoecology 248: 1- 23.

Drummond, A. J., M. A. Suchard, D. Xie, A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molec. Biol. Evol. 29:1969–1973.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M. and S. Markowitz. 2011. Geneious v5.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7.

Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.

Espindola, A., Carstens, B. C., and N. Alvarez. 2014. Comparative phylogeography of mutualists and the effect of the host on the genetic structure of its partners. Biological Journal of the Linnean Society, 113:1021-1035.

Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? Molecular Biology and Evolution 23:691-700.

García-París, M., Good, D. A., Parra-Olea, G. and D. B. Wake. 2000. Biodiversity of Costa Rican salamanders: implications of high levels of genetic differentiation and phylogeographic structure for species formation. Proceedings of the National Academy of Sciences 97: 1640-1647.

Garrick, R. C., D. M. Rowell, C. S. Simmons, D. M. Hillis and P. Sunnucks. 2008. Fine-scale phylogeographic congruence despite demographic incongruence in two low-mobility saproxylic springtails. Evolution 62:1103-1118.

Garrick, R.C., I. A. S. Bonatelli, A. E. Morales, T. A. Pelletier, M. F. Perez, E. Rice, J. D. Satler, R. E. Symula, M. T. C. Thomé and B. C. Carstens BC. 2015. The evolution of phylogeographic datasets. Molecular Ecology 24:1164-1171.

Gelman, A., Carlin, J. B., Stern, H.S., and D. B. Rubin. 2009. Bayesian data analysis, 2nd Ed. Chapman & Hall/ CRC Texts in Statistical Science. Boca Raton, FL, USA.

Gruenstaeudl, M., N. M. Reid, G. Wheeler, and B. C. Carstens. 2015. Posterior predictive checks of coalescent models: P2C2M, an R package. Molecular Ecology Resources, in press.

Heled, J. and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. Molec. Biol. Evol. 27:570–580.

Hewitt, G. M. 2004. The structure of biodiversity - insights from molecular phylogeography. Frontiers in Zoology 1:4.

Hickerson, M. J., and C. W. Cunningham. 2005. Contrasting Quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. Evolution 59:344-360.

Hickerson, M. J., E. A. Stahl, and H. A. Lessios. 2006. Test for simultaneous divergence using Approximate Bayesian Computation. Evolution 60:2435-2453.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.

Hudson, R. R. and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557-1565.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Hugall, A., C. Moritz, A. Moussalli and J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proc. Natl. Acad. Sci. USA 99:6112-6117.

Islebe, G. A. and H. Hooghiemstra. 1997. Vegetation and climate history of montane Costa Rica since the last glacial. Quaternary Science Reviews 16:589-604.

Ivanova, N. V., Dewaard, J. R. and P. D. Hebert. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes 6: 998-1002.

Janzen, D. H. 1973. Sweep samples of tropical foliage insects: effects of seasons vegetation types elevation time of day and insularity. Ecology 54: 687-708.

Janzen, D. H., Hallwachs, W., Blandin, P., Burns, J. M., Cadiou, J.-M., Chacon, I., Dapkey, T., Deans, A. R., Epstein, M. E., Espinoza, B., Franclemont, J. G., Haber, W. A., Hajibabaei, M., Hall, J. P. W., Hebert, P. D. N., Gauld, I. D., Harvey, D. J., Hausmann, A., Kitching, I. J., Lafontaine, D. O. N., Landry, J.-F., Lemaire, C., Miller, J. Y., Miller, J. S., Miller, L. E. E., Miller, S. E., Montero, J., Munroe, E., Green, S. R., Ratnasingham, S., Rawlins, J. E., Robbins, R. K., Rodriguez, J. J., Rougerie, R., Sharkey, M. J., Smith, M. A., Solis, M. A., Sullivan, J. B., Thiaucourt, P., Wahl, D. B., Weller, S. J., Whitfield, J. B., Willmott, K. R., Wood, D. M., Woodley, N. E. and J. J. Wilson. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Molecular Ecology Resources 9:1-26.

Joly, S. 2012. JML: testing hybridization from species trees. Molecular Ecology Resources 12**:**179–184.

Kappelle, M. and S. P. Horn. 2005. Páramos de Costa Rica. INBio Costa Rica.

Knowles, L. L. and W. P. Maddison. 2002. Statistical Phylogeography. Molecular Ecology 11:2623-2635.

Kuhner, M. K. and J. Felsenstein. 1994. Simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Molec. Biol. Evol. 11:459-468.

Lewis, P.O., Xie, W,, Chen, M-H., Fan, Y. and L. Kou. 2014. Posterior predictive Bayesian model selection. Systematic Biology 63:309–321.

Luteyn, J. L. 1999. Páramos: A Checklist of Plant Diversity Geographic Distribution and Botanical Literature New York Botanical Garden Press.

Montes, C., Cardona, A., McFadden, R., Morón, S., Silva, C., Restrepo-Moreno, S., Ramírez, D., Hoyos, N., Wilson, J. and Farris, D. 2012. Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. Geological Society of America Bulletin 124: 780-799.

Morin, P. J. 2011. Community Ecology, 2nd Ed. Wiley-Blackwell Press, Somerset, NJ, USA.

Moritz, C., J. L. Patton, C. J. Conroy, J. L. Parra, G. C. White, and S. R. Beissinger. 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science 322:261-264.

Morrone, J. J. and J. V. Crisci. 1995. Historical Biogeography: Introduction to Methods. Annual Review of Ecology and Systematics 26:373-401.

Nordborg, M. 2001. Coalescent theory. Handbook of statistical genetics, edited by D. J. Balding, M. Bishop and C. Cannings. Wiley, Chichester, UK

Oaks, J. R., J. Sukumaran, J. A. Esselstyn, C. W. Linkem, C. D. Siler, M. T. Holder and R. M. Brown. 2012. Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. Evolution 67:991-1010.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E. and J. C. Morrison. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of 69 terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51: 933-938.

Paradis, E., J. Claude and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinf. 20:289-290.

Puillandre, N., M. V. Modica, Y. Zhang, L. Sirovich, M. C. Boisselier, C. Cruaud, M. Holford and S. Samadi. 2012. Large-scale species delimitation method for hyperdiverse groups. Molecular Ecology 21:2671-2691.

R Development Core Team (2013) R: a language and environment for statistical computing. (R Foundation for Statistical Computing). Vienna, Austria. Available at: http://www.r-project.org

Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Computational Applications Biosciences 13:235-238.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645-1656.

Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57:4-14.

Reid, N. M., Brown, J. M., Satler, J. D., Pelletier, T. A., McVay, J. D., Hird, S. M. and B. C. Carstens. 2014. Poor fit to the multi-species coalescent model is widely detectable in empirical data. Systematic Biology 63: 322-333.

Reid, N. M. 2012. Model-based approaches to discovering diversity: New implementations, tests of adequacy and an empirical application to Central American Diptera. Ph. D. Dissertation, Louisiana State University.

Riddle, B. R., D. J. Hafner, L. F. Alexander, and J. R. Jaeger. 2000. Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. Proc. Natl. Acad. Sci. USA 97:14438-14443.

Savage, J. M. and F. Bolaños. 2009. An enigmatic frog of the genus *Atelopus* (Family Bufonidae) from Parque Nacional Chirripó, Cordillera de Talamanca, Costa Rica. Rev. Biol. Trop. 57: 381-386.

Sklenar, P., Luteyn, J. L., Ulloa, C., Jorgensen, P. M. and M. O. Dillon. 2005. Flora Generica de los Páramos: Guia Ilustrada de las Plantas Vasculares. New York Botanical Garden Press.

Smith, C. I., S. Tank. W. Godsoe, J. Levenick, E. Strand, T. Esque and O. Pellmyr. 2011. Comparative phylogeography of a coevolved community: Concerted population expansions in Joshua Trees and Four Yucca Moths. PLoS ONE 6:e25628.

Stiles, G. F. and A. F. Skutch. 1989. A Guide to the Birds of Costa Rica. Cornell University Press, Ithaca.

Streicher, J. W., Crawford, A. J. and C. W. Edwards. 2009. Multilocus molecular phylogenetic analysis of the montane *Craugastor podiciferus* species complex (Anura: Craugastoridae) in Isthmian Central America. Molecular Phylogenetics and Evolution 53: 620-630.

Sullivan, J., E. Arellano and D. S. Rogers. 2000. Comparative phylogeography of Mesoamerican highland rodents: Concerted versus independent response to past climate fluctuations. American Naturalist 155:755-768.

Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294-299.

Wakeley. J., and J. Hey. 2000. Estimating ancestral population parameters. Genetics 145:847-855.

Wares, J. P. 2010. Natural distributions of mitochondrial sequence diversity support new null hypotheses. Evolution 64:1136-1142.

Wiley, E. O. 1988. Parsimony analysis and vicariance biogeography. Systematic Zoology 37: 271-290.

Whiteman, N. K., Kimball, R. T., and P. G. Parker. 2007. Co-phylogeography and comparative population genetics of the threatened Galapagos hawk and three ectoparasite species: ecology shapes population histories within parasite communities. Molecular Ecology 16:4759-4773.

Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis F.R.S. Phil. Trans. Roy. Soc. Lond. B 213:21-87.

Table 1. Results of simulation testing. For each treatment with community diversification, three types of error are shown. Accurate detection measures the rate of detection of species with an evolutionary history distinct from that of the community; failed detection measures the rate that species with an evolutionary history distinct from that of the community are not detected; false identification measures the identification of species as outliers from a co-diversifying community when they in fact share a history of codiversification.

Treatment	PPS			K-F ANOVA		
	accurate detection	*failed detection*	*false identification*	*accurate detection*	*failed detection*	*false identification*
A	n/a	n/a	0.03	n/a	n/a	0.65
B	0.50	0.50	0.01	0.90	0.00	0.10
C	0.53	0.48	0.03	0.75	0.25	0.00
D	0.41	0.59	0.00	0.05	0.95	0.00

Table 2. Sampling information for the empirical data**.** Shown for each species are the number of samples in the Cerro Chirripó (CHI), Cerro de la Muerte (CLM), and Volcán Irazú (VIR). Also shown are the number of segregating sites (S), nucleotide diversity (p), the number of haplotypes (h) and the number of base pairs (bp).

Species	**CHI**	**CLM**	**VIR**	**S**	**π**	**θ_W**	**h**	**bp**
149	4	24	7	17	0.00949	0.00715	12	616
165	9	22	37	4	0.0002	0.00141	4	595
193	1	2	15	30	0.01654	0.01565	9	605
202	17	23	2	4	0.00165	0.00156	4	598
238	7	38	10	13	0.00129	0.00476	12	600
247	1	6	10	10	0.0029	0.00493	10	603
249	67	26	33	30	0.00518	0.01049	30	589
254	16	58	11	15	0.00168	0.00515	13	585
256	2	12	11	13	0.00646	0.00576	7	602
271	2	3	9	2	0.00068	0.00105	3	602

Table 3. Summary of community tree posterior predictive simulations. From left, the species labels, the median values and error terms of the difference distribution when all species are included in the community, and the same values when species #193 is removed. Outliers at the 0.01 tolerance level are marked with an '*'.

Species	**all species (0.01)**	**sp #193 removed (0.01)**
Sp#149	2.43 ±7.62	-0.13 ±6.8
Sp#165	-17.7 ±10.55	-16.66±8.41
Sp#193	5.27 ±2.36 *	*n/a*
Sp#202	-17.93 ±8.95	-18.8±7.58
Sp#238	5.02 ±12.37	0.65±11.48
Sp#247	0.8 ±3.49	0.78±3.22
Sp#249	33.93 ±30.05	26.18±22.99
Sp#254	-1.87 ±20.78	-5.3±16.2
Sp#256	-4.58 ±5.21	-4.79±4.6
Sp#271	-4.84 ±3.23	-3.98±2.41

Figure 1. (A) Distribution of land above 2000m in Costa Rica (shaded area) and distribution of Páramo (shapes nested within shaded area). Study sites are indicated by yellow stars and three letter abbreviations - VIR, CLM and CHI. (B) Distribution of counts of individuals, sequences and species delimited by ABGD across sites: individuals collected (Ind), individuals sequenced (Seq), number of species (Spp: lumped/split), distance between sites (Dist) and the maximum height of mountain passes between sites (gray boxes).

Figure 2. Simulation testing results. Results from a P2C2M analysis of the community tree simulations are shown. Genealogies from individual species are shown on the Y-axis of each figure, with replicates along the X-axis. Five simulation treatments are shown, from top they depict: (A) a community of 10 species that have codiversified, (B) a community of 9 species that have codiversified, with 1 outlier species, (C) a community of 8 species that have codiversified, with 2 outlier species each with their own idiosyncratic pattern, (D) a community of 6 species that has codiversified, with 4 outlier species each with their own idiosyncratic patterns of diversification, and (E) results from an analysis of 10 species with idiosyncratic histories. Dashed lines represent species that were modeled as outliers from the community diversification, while the dotted lines in Treatment E represent the absence of any shared history of diversification. For each of the species represented by this simulation, a black dot indicates an accurate detection of a species as an outlier from the community pattern of diversification, measured at a 0.01 tolerance using the LCWT

probability, and a red dot represents a false identification of a species as being an outlier from the community diversification.

Figure 3. Visualization of gene tree discord. Shown are K-F distance distributions from a representative replicate of simulation treatments A-E. Ten distributions (i.e., one for each constituent species) are shown for each treatment, with those detected as outliers shown in red. Note that as the number of species that share a history of codiversification decreases, the amount of observed gene tree discord (as measured by Kuhner-Felsenstein tree distances) increases. This is a function of the increase in the number of outlier species, and the cost of the constraint imposed by the community tree on these species. For plots of all simulation testing replicates, see Supplemental Figure 1.

Figure 4. Community diversification of Páramo flies. Figure 4A shows the cost of constraint when all data are analyzed in the community tree framework, while 4B presents these results after removing sp193. 4C shows the community tree estimated from 9 species, with tips abbreviated as follows: Volcán Irazú (VIR), Cerro de la Muerte (CLM), and Cerro Chirripó (CHI).

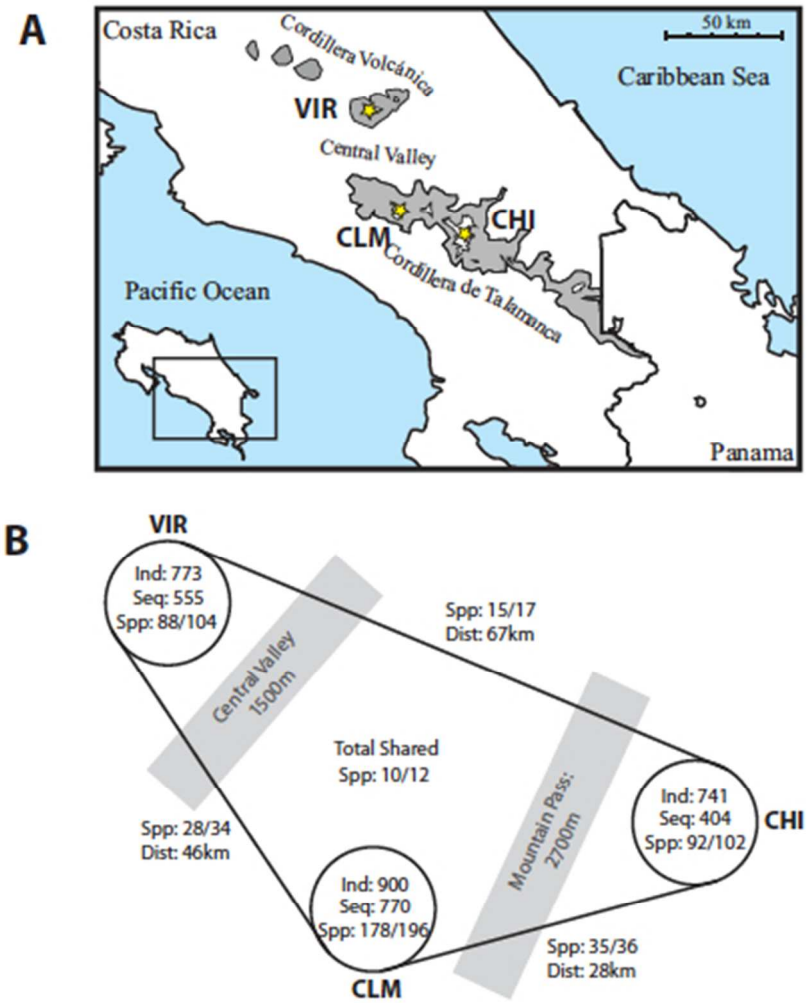


Figure 1. (A) Distribution of land above 2000m in Costa Rica (shaded area) and distribution of Páramo (shapes nested within shaded area). Study sites are indicated by yellow stars and three letter abbreviations - VIR, CLM and CHI. (B) Distribution of counts of individuals, sequences and species delimited by ABGD across sites: individuals collected (Ind), individuals sequenced (Seq), number of species (Spp: lumped/split), distance between sites (Dist) and the maximum height of mountain passes between sites (gray boxes).
151x182mm (72 x 72 DPI)

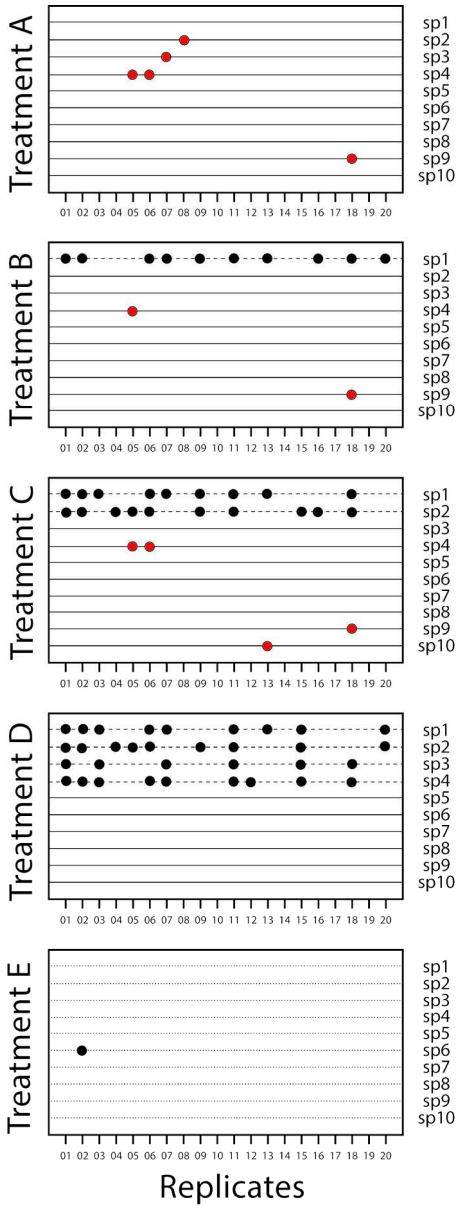
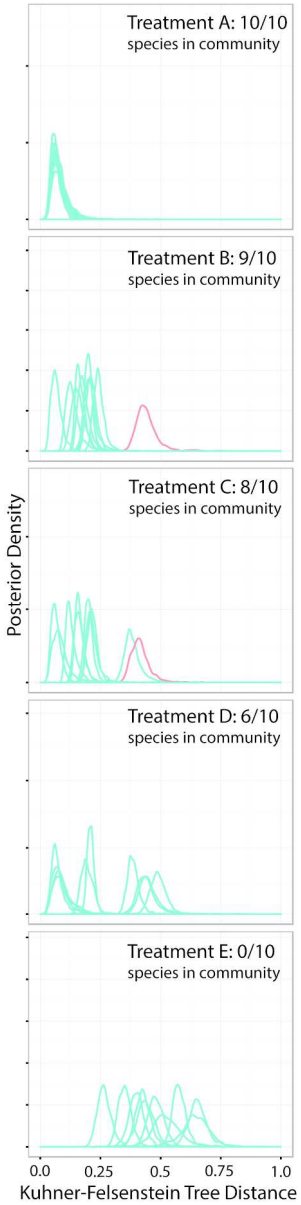
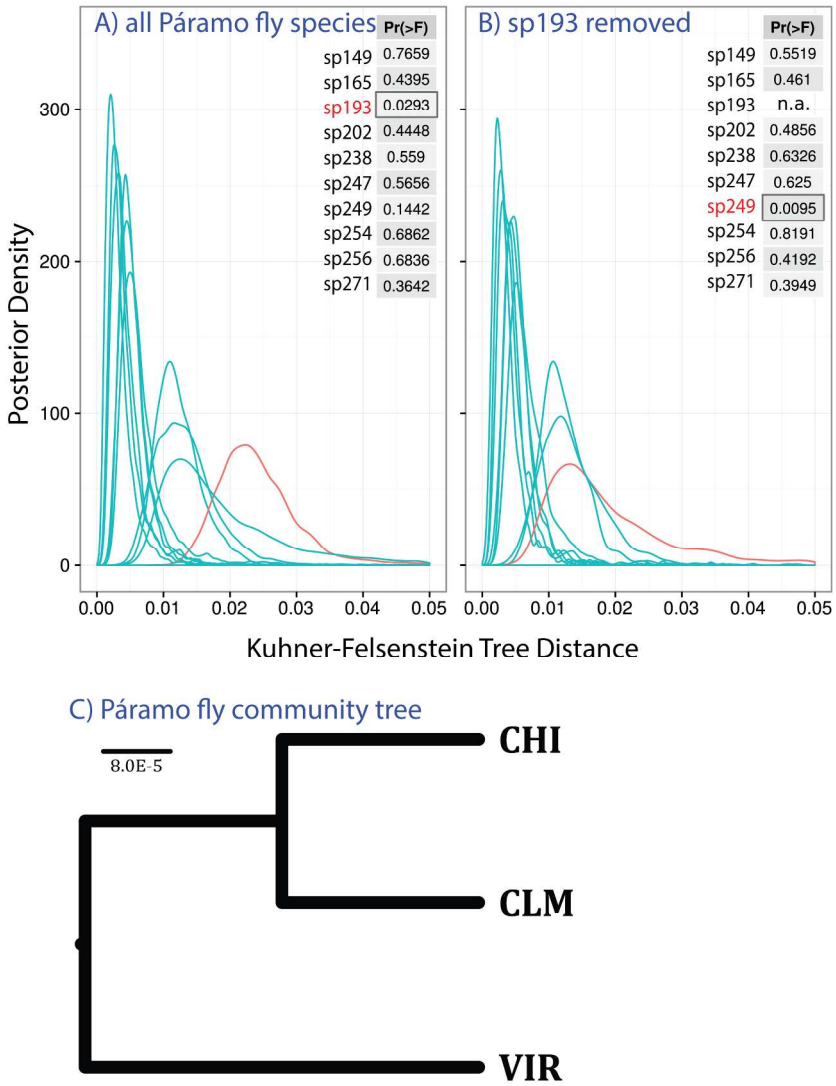


Figure 2. Simulation testing results. Results from a P2C2M analysis of the community tree simulations are shown. Genealogies from individual species are shown on the Y-axis of each figure, with replicates along the X-axis. Five simulation treatments are shown, from top they depict: (A) a community of 10 species that have codiversified, (B) a community of 9 species that have codiversified, with 1 outlier species, (C) a community of 8 species that have codiversified, with 2 outlier species each with their own idiosyncratic pattern, (D) a community of 6 species that has codiversified, with 4 outlier species each with their own idiosyncratic patterns of diversification, and (E) results from an analysis of 10 species with idiosyncratic histories. Dashed lines represent species that were modeled as outliers from the community diversification, while the dotted lines in Treatment E represent the absence of any shared history of diversification. For each of the species represented by this simulation, a black dot indicates an accurate detection of a species as an outlier from the community pattern of diversification, measured at a 0.01 tolerance using the LCWT probability, and a red dot represents a false identification of a species as being an outlier from the community diversification.

85x226mm (300 x 300 DPI)



Visualization of gene tree discord. Shown are K-F distance distributions from a representative replicate of simulation treatments A-E. Ten distributions (i.e., one for each constituent species) are shown for each treatment, with those detected as outliers shown in red. Note that as the number of species that share a history of codiversification decreases, the amount of observed gene tree discord (as measured by Kuhner-Felsenstein tree distances) increases. This is a function of the increase in the number of outlier species, and the cost of the constraint imposed by the community tree on these species. For plots of all simulation testing replicates, see Supplemental Figure 1.
61x251mm (300 x 300 DPI)



Community diversification of Páramo flies. Figure 4A shows the cost of constraint when all data are analyzed in the community tree framework, while 4B presents these results after removing sp193. 4C shows the community tree estimated from 9 species, with tips abbreviated as follows: Volcán Irazú (VIR), Cerro de la Muerte (CLM), and Cerro Chirripó (CHI).
202x266mm (300 x 300 DPI)