MOLECULAR ECOLOGY

Do ecological communities disperse across biogeographic barriers as a unit?

Journal:	*Molecular Ecology*
Manuscript ID	MEC-17-0003
Manuscript Type:	Original Article
Date Submitted by the Author:	01-Jan-2017
Complete List of Authors:	Satler, Jordan; The Ohio State University, Evolution, Ecology, and Organismal Biology Carstens, Bryan; The Ohio State University, Evolution, Ecology and Organismal Biology
Keywords:	Phylogeography, RAD sequencing, co-diversification, allele frequency spectrum, Sarracenia

RUNNING HEAD: DIVERSIFICATION ACROSS A BIOGEOGRAPHIC BARRIER

Do ecological communities disperse across biogeographic barriers as a unit?

Jordan D. Satler[1,2,*] and Bryan C. Carstens[1]

[1] *Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA*

[2] *Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA*

[*] **Corresponding Author:**

Department of Ecology, Evolution, and Organismal Biology, 251 Bessey Hall, Iowa State University, Ames, IA 50011, USA

Email: jordansatler@gmail.com

Keywords: RAD sequencing, phylogeography, co-diversification, allele frequency spectrum, *Sarracenia*

Abstract

Biogeographic barriers have long been implicated as drivers of biological diversification, but how these barriers influence co-occurring taxa can vary depending on factors intrinsic to the organism and in their relationships with other species. Due to the interdependence among taxa, ecological communities present a compelling opportunity to explore how interactions among species may lead to a shared response to historical events. Here we collect single nucleotide polymorphism (SNP) data from five commensal arthropods associated with the *Sarracenia alata* carnivorous pitcher plant, and test for co-diversification across the Mississippi River, a major biogeographic barrier in the southeastern United States. Population genetic structure in three of the ecologically dependent arthropods mirrors that of the host pitcher plant, with divergence time estimates suggesting two of the species (the pitcher plant moth *Exyra semicrocea* and a flesh fly *Sarcophaga sarraceniae*) dispersed synchronously across this barrier along with the pitcher plant. Patterns in population size and genetic diversity suggest the plant and ecologically dependent arthropods dispersed from east to west across the Mississippi River. In contrast, species less dependent on the plant ecologically show discordant phylogeographic patterns. This study demonstrates that ecological relationships may be an important predictor of co-diversification, and supports recent suggestions that organismal trait data should be prominently featured in comparative phylogeographic investigations.

Introduction

Comparative phylogeographic investigations can elucidate the historical processes that shape and structure biological diversity. A common approach is to infer population genetic structure and estimate parameters in a geographic context that devotes particular attention to biogeographic barriers. In such a context, similarities in the demographic histories across species are indicative of a shared response to landscape changes (Avise *et al.* 1987; Sullivan *et al.* 2000), while idiosyncratic patterns suggest an independent response. Although this framework is enticing in its simplicity, the details of how to compare demographic histories across species are key. Initial approaches for this comparison utilized gene trees, with similarity in pattern being suggestive of a common response to a historical event (Avise 2000; Arbogast & Kenagy 2001). While shared spatial patterns can indicate a similar history, temporal information is necessary to demonstrate a shared response in time (Edwards & Beerli 2000). Subsequent researchers have generally taken one of two approaches, either by estimating divergence times independently from each species for comparison (e.g., Carstens *et al.* 2005; Smith *et al.* 2012) or by using probabilistic models to estimate the number of divergence episodes (e.g., Hickerson *et al.* 2006, 2007). In addition to methodological considerations such as these, the nature of the biogeographic barrier itself is an important consideration.

Hard biogeographic barriers contribute to the diversification of biota by blocking the movement of individuals, providing physical barriers to gene flow, and provide an opportunity to understand how communities responded to a shared historical event (e.g., Pyron & Burbrink 2010). One compelling example is the formation of the Isthmus of Panama, which occurred around 3 Ma (Coates *et al.* 2005; but see Bacon *et al.* 2015 for

an alternative interpretation). The formation of the isthmus drastically altered marine environments on the Pacific and Caribbean sides and represents a hard barrier for marine organisms (Leigh *et al.* 2014), with phylogeographic studies demonstrating that most germinate species diverged prior to the formation of the barrier (Cowman & Bellwood 2013). The investigation also demonstrates that divergence was asynchronous, suggesting that intrinsic differences across species played a role in their response to this geological process (Knowlton & Weigt 1998; Lessios 2008). Another example of a hard barrier is in Baja California, where phylogeographic breaks are recovered in many taxa in the vicinity of the Vizcaino Desert 28–30° N latitude (e.g., Riddle *et al.* 2000; Garrick *et al.* 2009). Upton and Murphy (1997) proposed the presence of a mid-peninsular seaway 1 Ma to explain genetic patterns recovered in the side-blotched lizards (genus *Uta*), and although support is mixed for the presence of this seaway (e.g., Hafner & Riddle 2005; Crews & Hedin 2006; Leaché *et al.* 2007; Lindell *et al.* 2006), phylogeographic data from many taxa support a vicariant event in this region. Like the rise of the Isthmus of Panama, the formation of this putative seaway likely occurred after the ancestor of these species occupied Baja California.

Other biogeographic barriers are more porous. For example, Wallace (1852) observed discontinuities in the distributions of various monkey species in the Amazon basin, and proposed the riverine barrier hypothesis, where rivers act as barriers promoting genetic and taxonomic divergence. Although rivers are physical barriers for some taxa, the degree to which this is true is dependent on both intrinsic species' characteristics and characteristics of the river such as flow rate and direction, either of which may fluctuate over time. Consequently, support for riverine barriers has been mixed, with some

investigations supporting major rivers as biogeographic barriers (e.g., Burbink *et al.* 2000; Jackson & Austin 2010) and others demonstrating their permeability (e.g., da Silva & Patton 1998; Funk *et al.* 2007).

Major rivers dominate the landscape of the southeastern United States, including the largest river system in North America, the Mississippi River (Coleman 1988). The Mississippi River has its origins in the Mesozoic era, has been present (in some form) since the Jurassic (Mann & Thomas 1968), and more recently has been influenced by sea level fluctuations (Coleman 1988) and dramatic changes in the course of the river channel (Mann & Thomas 1968). While this ancient river has been identified as a biogeographic barrier in a variety of taxa (reviewed in Soltis *et al.* 2006), it is necessarily porous, as many species belonging to clades that have arisen since the Jurassic are distributed on both sides of the river. The inherent permeability of this barrier could be influenced by intrinsic factors, such as species-specific dispersal abilities, or extrinsic factors, such as oxbow lake formation that may have the effect of transferring land from one side of the river to the other (e.g., Gascon *et al.* 2000). Taken as a whole, the age and permeability of the Mississippi River necessitates that east–west divergence in terrestrial taxa would be due to dispersal and colonization, not vicariance (Pyron & Burbrink 2010). Since intrinsic species traits directly influence dispersal and colonization, riverine barriers provide an ideal setting for investigating how such traits influence biogeographic patterns.

Here we utilize the *Sarracenia alata* pitcher plant community to investigate whether intrinsic species traits dictate species response to porous biogeographic barriers. The carnivorous pitcher plant *S. alata* is restricted to bogs and fens in longleaf pine

savannahs along the Gulf Coast from eastern Texas to western Alabama, with a distribution bisected by the Mississippi River and Atchafalaya basin. Leaves of the plant are pitcher-shaped, an adaptation for the capture and digestion of prey (e.g., Darwin 1875; Ellison & Gotelli 2001), and also provide habitat for non-prey species (inquilines) that interact ecologically with the plant (reviewed in Adlassnig *et al.* 2011; also see Folkerts 1999). Some inquilines rely upon the pitcher plant for their entire life cycle (e.g., moths, flesh flies) while others are opportunistic predators that intercept prey from the plant (e.g., spiders) but are not restricted to the unique habitat provided by *S. alata*. The varying degrees of dependence on the plant led Satler and Carstens (2016) to suggest that ecological relationships may predict the degree of phylogeographic congruence in this community. If true, then the demographic history of dependent inquilines should reflect that of the plant, while the demography of the opportunistic inquilines should reflect intrinsic species traits related to dispersal ability.

Investigations into *S. alata* have demonstrated that genetic diversity in the plant is structured largely due to the influence of multiple rivers that divide its range into several regions. Results from chloroplast DNA, microsatellites, and SNP data (Koopman & Carstens 2010; Zellmer *et al.* 2012) indicate that population genetic structure is largely promoted by major rivers, and estimates of the pattern of diversification demonstrate that the deepest divergence within *S. alata* occurs on either side of the Mississippi River, dates to the mid-Pleistocene (Zellmer *et al.* 2012), and may be indicative of independent evolutionary lineages (Carstens & Satler 2013). In addition, evidence suggests that the pitcher plant has a center of origin in the east, and dispersed across the Mississippi River in an east-to-west manner (Zellmer *et al.* 2012).

We sample five arthropod species: a moth, two flesh flies, and two spiders (Table 1). Inquilines include the moth, which spends its entire life cycle in the pitcher plant leaves (Jones 1921; Stephens *et al.* 2011), and the flesh flies, which are also tightly associated with the plant leaves (Dahlem & Naczi 2006). Notably, both the moth and flies are poor flyers and dispersal limited (Folkerts 1999; Stephens & Folkerts 2012; Krawchuk & Taylor 2003; Rasic & Keyghobadi 2012). In contrast, the spider species sampled here are widespread and opportunistic predators, abundant in the habitat but evidently not dependent on *S. alata*, since either can be found in a variety of other microhabitat in the region.

Community diversification hypotheses

Here we test the null hypothesis (H_0) is that there is no correlation between ecological interaction and a shared evolutionary history. The null predicts that we should observe discordant phylogeographic patterns among the species, and would suggest that ecological associations are fluid in time and do not influence the evolutionary history of a particular species. Alternatively (H_1), it may be that strong ecological interactions result in a shared evolutionary history, in which case the phylogeographic inferences in the obligate inquilines (moth and flies) should reflect that of the host plant, while the spiders would display discordant evolutionary patterns. This alternative hypothesis follows Smith *et al.* (2011), who proposed that evolutionary communities of species that are ecological interdependent may exist. A corollary of this hypothesis is that divergence time estimates from the obligate mutualists should postdate estimates from *S. alata*, reflecting the need for suitable habitat following dispersal across the Mississippi River. A third hypothesis

(H_2) is that species-specific traits influence evolutionary patterns, such that similar species would be expected to share population genetic structure and demographic patterns (e.g., Papadopoulou & Knowles 2016). Here we would predict that similar patterns would be seen *within* taxonomic groups (e.g., flies, spiders), but would be dissimilar *across* the groups. These hypotheses can be tested using estimates of population genetic structure and demographic parameters.

Material and Methods

Taxon sampling

Five arthropod species were collected from 12 possible localities throughout the distribution of *S. alata* (Fig. 1; Table 1; see Table S1 for detailed sampling information). Arthropods included a moth (*Exyra semicrocea*), two flesh flies (*Sarcophaga sarraceniae, Fletcherymia celarata*), and two spiders (*Misumenoides formosipes*, *Peucetia viridans*). All individuals were captured in the vicinity of the plant, usually resting on or just under the lid of the pitcher. For the flesh flies, males were pinned in the field with their genitalia extracted to confirm proper identification. Of those specimens, three legs were removed and placed in 95% EtOH for DNA preservation. All other specimens were preserved directly in 95% EtOH for DNA preservation.

DNA sampling preparation and processing

Genomic DNA was extracted from the specimens (either leg tissue or full body soakings) using a Qiagen DNeasy kit. Between 24 and 26 individuals were selected per species for sequencing, with samples that span the distribution of the species and roughly

equal numbers on either side of the Mississippi River (see Table S1). A double digest restriction-site associated DNA sequencing (ddRADseq) protocol (modified from Peterson *et al.* 2012) was used to generate genomic sequences. Specifically, genomes were digested with two restriction enzymes (SbfI and MspI) to reduce the number (but generate higher coverage) of the potential suite of homologous loci. Following restriction enzyme digest and ligation of internal barcodes, libraries were amplified through polymerase chain reaction. After confirming amplification of the sequence libraries via gel electrophoresis, size selection was conducted with a Blue Pippin (Sage Sciences) targeting fragments between 300 and 600 base pairs. Samples were quantified using a bioanalyzer and qPCR to confirm quality library prep, and sequenced using an Illumina HiSeq with single end 100 base pair reads.

Following demultiplexing, raw sequence reads were filtered with AlienTrimmer v0.4.0 (Criscuolo & Brisse 2013) to remove reads with adapter contamination. All retained reads were then trimmed to 80 base pairs with the FASTX-Toolkit v0.0.14 (Gordon & Hannon 2010) to account for uneven barcode lengths and remove potentially low-quality base pairs towards the end of sequences. Next, sequence reads were analyzed with Pyrad v3.0.66 (Eaton 2014) using parameter settings that were consistent for all five species. Pyrad is an automated pipeline that takes as input raw sequence reads and outputs loci, alleles, and SNPs. Base calls with a Phred score below 20 were replaced with Ns; up to four Ns were allowed for a read to be retained. A clustering threshold of 88% was used to assemble reads into loci. RAD sequencing is prone to missing data due to mutations in the restriction enzyme sites as well as allelic dropout, and this missing data can bias parameter estimates in downstream analyses (Arnold *et al.* 2013). However,

because retaining only those loci with 100% coverage can also bias parameter estimates (because such loci are likely to be evolving more slowly than genome-wide averages), we allowed for some missing data in our analysis and retained loci with a minimum of 75% coverage across individuals.

Population genetic structure

To infer population genetic structure within the species, we used STRUCTURE v2.3.4 (Pritchard *et al.* 2000), which assigns individuals into clusters by maximizing linkage equilibrium within clusters and minimizing linkage disequilibrium between clusters. Our analyses were conducted at the $K = 2$ clustering level, reflecting our understanding that populations of *S. alata* east and west of the Mississippi River comprise two distinct lineages (Zellmer *et al.* 2012; Carstens & Satler 2013). Our prediction, particularly for obligate commensal species, is that genetic structure at this level will also reflect this east–west division of habitat. For each species, we converted allelic data into haplotypes at each locus, utilizing the information contained in linked SNPs when more than one SNP was present within a locus. If any allele contained one or more Ns, we adopted the conservative approach of treating this sequence as missing due to ambiguity in allelic assignment. Analyses were conducted using an admixture model with correlated allele frequencies, sampling location information for each species, a burn-in of 1×10^5 generations and subsequent sampling for 5×10^5 generations. Each analysis was repeated 10 times, and results were processed and summarized with the pophelper package (Francis 2016) in R (R Core Team 2015).

An Analysis of Molecular Variance (AMOVA; Excoffier *et al.* 1992) was conducted on each species to assess the level of genetic partitioning across the landscape. Specifically, we tested for genetic partitioning (i) within each locality, (ii) between localities on either side of the Mississippi River, and (iii) within each side of the river. STRUCTURE haplotype files were converted to Arlequin files using PGDspider v2.0.7.1 (Lischer & Excoffier 2012). AMOVA analyses were conducted in Arlequin v3.5.1.2 (Excoffier *et al.* 2005), with distance matrices calculated using the number of different alleles per locus and 10,000 permutations to assess significance.

In addition, summary statistics were calculated from the data with the python library DendroPy v4.1.0 (Sukumaran & Holder 2010). These included number of segregating sites, nucleotide diversity (π), Watterson's theta (Θ_w), and Tajima's *D*. Samples were partitioned based on side of river (east or west), reflecting our understanding of the diversification of the host pitcher plant.

Estimating population divergence, population size, and gene flow

Phylogeographic concordance factors suggest that multiple arthropods are concordant with *S. alata* (Satler & Carstens 2016). In order to explore this suggestion, parameters including population divergence (τ), population size (N_e), and gene flow (2Nm) were estimated from the SNP data in each species using allele frequency spectrum (AFS) methods (Gutenkunst *et al.* 2009; Excoffier *et al.* 2013). One recently developed method, *fastsimcoal2* (FSC2; Excoffier *et al.* 2013), uses coalescent simulations to calculate the likelihood of the observed AFS given a demographic model using the likelihood calculation developed by Nielsen (2000). Simulations suggest that FSC2 is

computationally efficient and produces accurate parameter estimates (Excoffier *et al.* 2013). As models are user-specified, the flexibility of FSC2 makes it appealing to apply to the analysis of data from non-model species where the correct model is unknown (Thomé & Carstens 2016).

Model selection has become an integral part of phylogeography in large part because the utility of parameter estimation to the inferences process relies on the appropriateness of the analytical models (e.g., Fagundes *et al.* 2007; Carstens *et al.* 2013). Because populations of *S. alata* have been isolated on either side of the Mississippi River for a considerable amount of time (Zellmer *et al.* 2012; Carstens & Satler 2013), we assumed a two-population model, grouping samples on either side of the biogeographic barrier into populations, but consider several models containing different combinations of parameters (e.g., τ, N_e, 2Nm) in each species. FSC2 calculates a composite likelihood with the assumption SNPs are in linkage equilibrium, and thus any genetic linkage may bias this calculation and invalidate model comparisons. To satisfy this assumption, we randomly selected one SNP per locus to generate an unlinked AFS. We then conducted model-selection on seven variants of the isolation-with-migration (IM) model (Fig. 2) using Akaike information criterion (AIC; Akaike 1974) and model probabilities calculated following Burnham and Anderson (2002). Parameter estimates were subsequently generated via model-averaging (i.e., weighted by the probabilities for each of the models), allowing for estimates of a particular parameter to contribute to the overall parameter estimate in proportion to its model probability. Since we were concerned that reducing our data set to only unlinked SNPs might leave us with too few SNPs to accurately estimate parameters of interest, we also estimated parameters using

the traditional IM model (Fig. 2C) for each species incorporating all of the SNPs. Linkage among SNPs affects the calculation of the likelihood, not parameter estimation, so linked SNPs are not expected to bias parameter estimation when a single model is used.

Analyses were conducted with *fastsimcoal* v25221 (Excoffier *et al.* 2013). We constructed a folded allele frequency spectrum from minor allele counts as we did not have sequence data from outgroups. Fixed numbers of alleles for all populations are required for generating the observed AFS, however, only including SNPs with 100% coverage would drastically reduce (and likely bias) our sampling. To account for missing data while maximizing the number of SNPs, we required that 75% of alleles were present within each population (east and west) for the SNP to contribute to the AFS. Given these criteria, building of the observed AFS took place in three ways: (i) if either population had fewer alleles than the set threshold, that SNP was discarded, (ii) if either population had the same number of alleles as the threshold, the allele frequencies were calculated (for the total SNP) and the minor allele count was used in the AFS, (iii) if either population had a greater number of alleles than the threshold, the alleles were subsampled *with replacement* until the necessary number of alleles (matching the threshold) were sampled, and then the minor allele was counted. For the SNP that met either criterion *ii* or *iii*, the proper cell was populated in the AFS with the minor allele counts from each population. Although this down sampling procedure allowed us to include more SNPs in our analysis, it had the undesirable effect of subsampling some alleles such that they appeared monomorphic in a particular subsampling replicate. To account for variation in generating the observed AFS, we replicated the AFS building procedure 10 times.

Replication serves two purposes: (i) it accounts for variation in the subsampling process, and (ii) allows us to generate confidence intervals on parameter estimates for across species comparisons.

To convert parameter estimates to real values, we assumed a mutation rate of 8.4 X 10^{-9} estimated from *Drosophila* flies (Haag-Liautard *et al.* 2007). Species-specific generation length estimates were gathered from the literature to scale parameters to real values. Specifically, we used two generations per year for the moth (Moon *et al.* 2008) and flies (Rango 1999; Rasic & Keyghobadi 2012), and one generation per year in the spiders (Foelix 1982); we discuss later implications of uncertainty in these estimates. We also counted the number of invariant sites in the sequence data to populate the monomorphic cell. All FSC2 analyses were run on the Oakley cluster at the Ohio Supercomputer Center (https://osc.edu). Each analysis (for each AFS replicate per model) was repeated 50 times, to take into account stochasticity in the simulated AFSs (as recommended by Excoffier *et al.* 2013). The run with the highest composite likelihood was then selected as the best run (among the 50), and parameter estimates from these runs were recorded. Custom python and bash scripts (available on Dryad; ####) were written to generate the observed AFS, prepare each analysis, and collate and summarize the results.

Results

DNA sequencing

We sequenced either 24 (*S. sarraceniae, F. celarata*, *M. formosipes*) or 26 (*E. semicrocea*, *P. viridans*) individuals of each species using two HiSeq lanes and a partial MiSeq lane resulting in ~310 million sequence reads. Following demultiplexing and quality control, we retained ~215 million reads for *de novo* assembly. Four individuals were poorly sequenced (one each from *F. celerata* and *P. viridans*, and two from *M. formosipes*), and they were removed from downstream analyses. Using an 88% within-species clustering threshold and requiring at least six reads before calling a cluster, an average of between 1617 and 4615 clusters were generated per species (Table 2). Our final data sets—requiring at least 75% of individuals—contained between 383 and 1037 loci, and between 617 and 2055 variable sites for analysis. SNP files and data sets are available on Dryad (####).

Population genetic structure

STRUCTURE results vary by species, but consistently reflect the ecology of the species in question (Fig. 3). The moth (*E. semicrocea*) is partitioned into two groups on either side of the Mississippi River, with a similar pattern recovered in one of the flies (*S. sarraceniae).* Population structure in the other fly species (*F. celarata*) is minimal, as essentially no structure is seen at the $K = 2$ level. This result, however, appears to be an artifact of the uneven sampling on either side of the Mississippi River, as only five flies were sampled from west of the Mississippi River (Table S1; see Puechmaille 2016 for discussion of how such uneven sampling can bias STRUCTURE results). When we randomly subsampled individuals in the eastern locales to be similar in number to the sample sizes in the west, genetic partitions were geographically clustered, recovering the

east–west split (subsampling replicated 10 times, with STRUCTURE analyses run as outlined above; Fig. S1). In contrast to the insects, neither spider species exhibited appreciable genetic structure, with STRUCTURE plots discordant with geography.

Results from the AMOVA are consistent with those from STRUCTURE. In both *E. semicrocea* and *S. sarraceniae,* there is significant genetic structure at all three levels of the analysis, demonstrating strong population genetic structure in each species (Table 3). Population structure in the other fly (*F. celarata*) suggests significant association among localities, although lack of permutations precludes us from testing for significance at the other hierarchical levels (see Fitzpatrick 2009). In the spiders, genetic data are not significantly structured at any of the hierarchical levels, consistent with results from STRUCTURE and our inference of a loose association between the spider species and the pitcher plant. Various summary statistics are consistent with these results. For the moth and flies, genetic diversity is consistently higher in the east than in the west (Table 4). This is in contrast to the spiders, where genetic diversity is higher in the west than the east (*M. formosipes*) or is similar on either side of the river (*P. viridans*). All species show negative Tajima's *D* values, although the standard deviations encompass small positive values for all calculations.

Estimating population divergence, population size, and gene flow

Model selection

We specified seven models for analysis using the unlinked AFS, all variations of the traditional isolation-with-migration models (Fig. 2). Results were similar across

374 species in that isolation-only models had low model probabilities, and for each species,
375 multiple models received appreciable support (Table 5).
376
377 Divergence times
378 As most species had strong support for one of the IM models (Table 5), parameter
379 estimates were relatively consistent across data sets regardless of whether they were
380 generated via model-averaging (from unlinked AFS) or from the full IM model (using the
381 linked AFS). In general, unlinked AFS with model-averaged parameters contained
382 slightly younger divergence times than linked AFS with the IM model, not surprising
383 given the contribution of models that did not include gene flow. For the remainder of this
384 paper, we consider parameter estimates generated from the model-averaging approach
385 with unlinked data sets, but note that results from the other analyses are similar (i.e., Figs.
386 S2–S3; Table S2).
387 Divergence times were restricted to the Pleistocene in all species (Fig. 4), with the
388 precision varying across taxa. Assuming two generations per year, the moth
389 (*E. semicrocea*) is estimated to have diverged 230,745 years before present, with a 95%
390 CI of 213,555–247,935 (Table 6). Divergence time estimates for the flies were shallower
391 than the moth. Assuming two generations per year, divergence time in *S. sarraceniae*
392 averaged 195,045 years before present (95% CI 145,995–244,096), while those in
393 *F. celarata* averaged 90,598 years before present (95% CI 84,668–96,528; Table 6). For
394 the spiders, assuming one generation per year, divergence time estimates were older than
395 the rest of the community: *M. formosipes* ~475k years before present (95% CI 435,551–
396 511,349); *P. viridans* ~245k years before present (95% CI 235,180–253,086).

Collectively, divergence time estimates span from ~90k years before present to ~475k years before present (Fig. 4).

Population sizes

As with the divergence time estimates, values are generally consistent within species regardless of whether estimates where model-averaged (with unlinked AFS; Fig. 5) or from an IM model (with linked AFS; Fig. S3, Table S2). For the moth, population sizes in the east are roughly five times as large as those in the west (520,215 vs. 102,907; Table 6). This same pattern is evident in both flesh flies, where population sizes in the east are roughly two to five times as large as those in the west (*S. sarraceniae*: 693,356–124,174; *F. celarata*: 314,109–168,863; Table 6). In contrast, population sizes in the spiders are similar on either side of the river (*M. formosipes*: 764,333 (E) – 556,845 (W); *P. viridans*: 394,613 (E) – 333,666 (W); Table 6).

Gene flow

Migration rates (2Nm) are lowest among the ecologically dependent species (Table 6). In the moth, migration is below 0.75 in either direction, suggesting little to no migration within this species. Low levels of migration are seen with the flies, although values of 1.57 for *S. sarraceniae* from west to east and 3.72 for *F. celarata* from east to west suggest higher levels of migration (Table 6). Migration rates, however, are highest within the spiders. In *M. formosipes*, $2Nm_{\text{west to east}} = 5.61$; in *P. viridans*, $2Nm_{\text{east to west}} = 4.08$, and $2Nm_{\text{west to east}} = 5.54$ (Table 6).

Discussion

Diversification patterns of the Sarracenia alata *ecological community*

Zellmer *et al.* (2012) demonstrated that divergence across the Mississippi River in *S. alata* occurred in the Pleistocene, roughly 120,000 years before present. Given findings in other studies of host plants and associated arthropods that share a phylogeographic history (e.g., Smith *et al.* 2011), we tested the prediction that obligate commensals of *S. alata* should exhibit concordant population genetic structure, as well as divergence time estimates similar to or more recent than the plant, reflecting the requirement of this specialized habitat to facilitate colonization following dispersal to the west side of the river for the arthropods.

We sampled arthropods from five divergent lineages, ranging in their association with the host pitcher plant from obligate inquiline commensals (moth and two flies) to opportunistic capture interrupters (two spiders). Both estimates of population genetic structure and parameters demonstrate that the three commensal arthropods (pitcher plant moth *E. semicrocea,* pitcher plant flies *S. sarraceniae* and *F. celarata*) exhibit an evolutionary history largely congruent with the host pitcher plant. Perhaps the most compelling result is that the pattern of genetic diversity and effective population sizes of these three species and *S. alata* are largely concordant on either side of the river, with high diversity in the east and low diversity in the west. These result support H_1, and suggests the long-term association of the ecologically dependent arthropod species with the host pitcher plant. In addition, estimates of population divergence across the Mississippi River indicate that at least two of these three arthropods (pitcher plant moth *E. semicrocea* and pitcher plant fly *S. sarraceniae*) dispersed across the Mississippi River

largely in concert with *S. alata*, as their divergence time estimates are within 35k years of one another (230 kya and 195 kya), with confidence intervals that overlap with each other and those estimated from the pitcher plant. These results suggest the association between these arthropods and *S. alata* has been stable for nearly 200,000 years.

The pattern identified in the other flesh fly (*F. celarata*) is intriguing. Population divergence estimates from *F. celarata* are more recent (~90k years before present) than those from *S. alata* and the other dipteran (*S. sarraceniae*), suggesting this species dispersed across the Mississippi River after western populations of the plant and *S. sarraceniae* were already well established. In our extensive fieldwork, we were only able to collect five flies of *F. celarata* from the western locales (all from Cooter's Bog; see Fig. 1); in contrast, we collected 73 *S. sarraceniae* individuals from the west. A series of *F. celarata* specimens is known from Warren, Texas (see Dahlem & Naczi 2006), but we were unable to locate any individuals of this species in any other western locale. These five samples are monophyletic in their mitochondrial DNA (Satler & Carstens 2016), and population genetic parameters support their east-to-west dispersal and structure (following subsampling and replication in STRUCTURE; Fig. S1). Three factors could explain these results. For one, abiotic factors in the west may play a limiting role in *F. celarata's* ability to disperse throughout the western landscape. Environmental niche models (see Zellmer *et al.* 2012) suggest an inland/coastal division (for *S. alata*), but given where *F. celarata* has been sampled, environmental differences between eastern and western locales may be contributing to the lack of presence of these flies in additional western bogs. Alternatively, the younger divergence time recovered in *F. celarata* could be an artifact caused by limited sampling. Although we were able to

sample up to 10 alleles per locus for the western individuals, limited geographic sampling combined with lower numbers of allele counts may have precluded us from generating accurate estimates of divergence times. Finally, the limited sampling and population genetic parameter estimates could reflect differing outcomes of interspecies competition. Pitcher plant flesh flies are ovolarviparous, with females depositing one larvae per pitcher. Larvae are aggressive and territorial, actively attacking other flesh fly larvae when present (Forsyth & Robertson 1975; Rango 1999; Dahlem & Naczi 2006). As the two flesh fly species fill the same ecological niche, it may be the case that *F. celarata* is less competitive in certain environments, resulting in higher numbers of *S. sarraceniae* in the west. Given our estimated divergence times, *S. sarraceniae* would have had substantially more time to become established (than *F. celarata*) at plant populations west of the Mississippi River, leading to their higher abundance in our sampling efforts. We note, however, that both species co-occur in the east, so the presence of one species does not preclude the presence of the other. Although we do not have sufficient data to conclusively determine the cause of the discordant pattern in sampling and divergence dating, all other population genetic parameters support an east-to-west dispersal in *F. celarata*, with population structure mirroring the pitcher plant highlighting the tight relationship between the flesh fly and host plant.

Our findings are consistent with a growing biogeographic understanding of this system. Stephens *et al.* (2015) proposed a center of origin for *Sarracenia* in southeastern North America where the other *Sarracenia* species are distributed. In addition to being the only member of the genus found west of the Mississippi River, population genetic patterns in *S. alata* support this hypothesis, with colonization of the west from eastern

populations. Results from the insects are consistent with this scenario. This raises the question: how did the *S. alata* community disperse across the Mississippi River? *Sarracenia* seeds are tiny and lack modifications for long-range dispersal (Ellison 2001). Ellison and Parker (2002) recovered most seeds of *Sarracenia purpurea* within five cm of the parent plant, suggesting limited seed dispersal in these plants. We follow Zellmer *et al.* (2012) in suggesting that a likely scenario is the course of the river changed to effectively move some habitat from the east side to the west via the process of oxbow lake formation (e.g., Gascon *et al.* 2000). The lower Mississippi River is a dynamic system, with tremendous change in movement and flow during the Pleistocene (Mann & Thomas 1968; Coleman 1988). Such a process would provide the opportunity for mature plants and their commensal arthropods to move as a single unit across the river.

Intrinsic species traits and porous biogeographic barriers

Two spider species included here are markedly incongruent with the demographic patterns evident in *S. alata* and its commensal arthropods. In each species the rates of gene flow across the Mississippi River are high (Table 6), which leads to population sizes on either side of the river that differ from patterns in the *S. alata* community. We suspect that intrinsic species traits related to dispersal explain these differences, as both spiders can travel long distances via ballooning. Divergence time estimates in the crab spider (*M. formosipes*) are much older (~475k years before present) than the other species, while those in the green lynx spider (*P. viridans*) are older, but similar with those of other species analyzed here. As both spiders exploit the insect-attracting abilities of *Sarracenia* (Folkerts 1999) and are commonly found in association to the pitcher plant but not

limited to this specialized habitat, it seems clear that ongoing gene flow within the region have produced the discernable lack of population genetic structure in these species.

Challenges with comparing divergence times across a biogeographic barrier

Investigating the timing of diversification across biogeographic barriers is of central importance to the discipline, as a clustering of divergence times suggests a shared response to a historical event (Bermingham & Moritz 1998). Accurately estimating divergence times is particularly challenging when the focal species are sampled from disparate taxonomic groups. Methods incorporating the coalescent model allow the timing of population divergence to be directly estimated, potentially leading to more precise inferences of community divergence (Hickerson *et al.* 2006), but rely on external information that may be unknown in non-model species. While phylogeography has assumed since Edwards and Beerli (2000) that more data would lead to more precise estimates of population divergence and thus facilitate comparative studies that span biogeographic barriers, comparative investigations require two types of information (mutation rate and generation length) to convert estimates to values that can be compared across species. Within the same taxonomic groups, these values are typically assumed to be the same across taxa (e.g., Smith *et al.* 2014; Papadopoulou & Knowles 2015), but in studies such as ours, a comparison of species that are only distantly related to one another is complicated by a lack of information about these values. Here, we utilized a direct estimate of the mutation rate from *Drosophila* flies (Haag-liutard *et al.* 2007) for lack of a better option, but note that its relevance to the distantly related dipterans, lepidopterans, and arachnids analyzed here is suspect because the three groups likely diverged before

the Cambrian (e.g., Rehm *et al.* 2011). Perhaps a larger concern is generation length. In this study, we investigated small arthropods where there is little existing information about life history traits. Previous works suggests that araneomorph spiders have one generation per year (Foelix 1982), but we are less certain about the remaining arthropods. The moth and the flies are reported to have multiple generations per year, but exact values are unknown (Folkerts 1999). Moon *et al.* (2008) suggested *E. semicrocea* has two generations per year, and this value is consistent with estimates from other moths in the Noctuidae family (e.g., Spitzer *et al.* 1984). For the flies, we based our estimate of two generations per year on research conducted in another pitcher plant flesh fly (*Fletcherimyia fletcheri*) that is associated with *Sarracenia purpurae.* *Fletcherimyia fletcheri* is estimated to have one generation per year at the higher latitudes in northeastern United States and Canada (Rango 1999; Rasic & Keyghobadi 2012), where pitcher leaves are active for ~4–8 weeks (Fish & Hall 1978). But in *S. alata*, where leaves appear to be active for at least four months, we believe it is reasonable to use a value of two generations per year. This is consistent with generation time estimates in other flesh flies (in the genus *Sarcophaga*) suggesting 2–3 generations per year in the temperate regions, with generation cycles taking up to 60 days (Denlinger 1978) depending on day length and temperature (Chen *et al.* 1987; Lee *et al.* 1987). Furthermore, seasonal and yearly fluctuations in climate and environment will influence the number of generations in groups such as dipterans, which may mean that there were fewer generations per year in the cooler Pleistocene Epoch. Regardless, these assumptions clearly influence estimates of divergence time (Fig. 6).

Conclusions

Our results suggest that *S. alata* and at least two of its commensal arthropods dispersed across the Mississippi River in a concerted manner, likely facilitated via oxbow lake formation (Koopman & Carstens 2010), and suggest that these species represent an evolutionary community *sensu* Smith et al. (2011). Given the similarities in population genetic structure and effective population sizes across the members of this community, it seems clear that the evolution of each species is influenced by the other members of the community, as predicted by Darwin's tangled bank (Darwin 1859). While there are clearly methodological difficulties pertaining to the analysis of genetic data, our work demonstrates the importance of considering both the ecological relationships and the intrinsic species traits when conducting comparative phylogeographic investigations.

570

571 **Funding**

572 Funding was provided by the Society for the Study of Evolution (Rosemary Grant
573 Award for Graduate Student Research) and the Society of Systematic Biologists
574 (Graduate Student Research Award) to J.D.S., and the Louisiana EPSCoR Pfund, the
575 LSU Office of Research and Economic Development and Louisiana Board of Regents to
576 B.C.C.

577

578

579 **Acknowledgements**

580 We thank Rob Naczi and Greg Dahlem for assistance with identifying fly
581 specimens, Steve Bgdanowicz, Kelly Zamudio, and the Zamudio lab for assistance with
582 RAD sequencing, and Mike Sovic for discussion regarding *fastsimcoal2* analyses. We
583 also thank Mitchell Cruzan, Jon Eldon, and two anonymous reviewers for comments that
584 helped to improve this manuscript.

585

References

Adlassnig W, Peroutka M, Lendl T (2011) Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Annals of Botany*, **107**, 181–194.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC*, **19**, 716–723.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard Univ. Press, Cambridge, MA.

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences, USA*, **112**, 6110–6115.

Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.

Burbrink FT, Lawson R, Slowinski JB (2000) Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution*, **54**, 2107–2118.

Burnham KP, Anderson DR (2002) Model Selection and Multimodal Inference: A Practical Information Theoretic Approach, 2nd edn. Springer-Verlag, New York.

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J (2005) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.

Carstens BC, Satler JD (2013) The carnivorous plant described as *Sarracenia alata* contains two cryptic species. *Biological Journal of the Linnean Society*, **109**, 737–746.

Carstens BC, Brennan RS, Chua V, Duffie CV, Harvey MG, Koch RA, McMahan CD, Nelson BJ, Newman CE, Satler JD, Seeholzer G, Posbic K, Tank DC, Sullivan J (2013) Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology*, **22**, 4014–4028.

Chen C, Denlinger DL, Lee RE (1987) Responses of nondiapausing flesh flies (Diptera: Sarcophagidae) to low rearing temperatures: developmental rate, cold tolerance, and glycerol concentrations. *Annals of the Entomological Society of America*, **80**, 790–796.

Coates AG, McNeill DF, Aubry MP, Berggren WA, Collins LS (2005) An introduction to the geology of the bocas del toro archipelago, Panama. *Caribbean Journal of Science*, **41**, 374–391.

Coleman JM (1988) Dynamic changes and processes in the Mississippi River delta. *Geological Society of America Bulletin*, **100**, 999–1015.

Cowman PF, Bellwood DR (2013) Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131541.

Crews SC, Hedin M (2006) Studies of morphological and molecular phylogenetic divergence in spiders (Araneae: Homalonychus) from the American southwest, including divergence along the Baja California Peninsula. *Molecular Phylogenetics and Evolution*, **38**, 470–487.

Criscuolo A, Brisse S (2013) AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics*, **102**, 500–506.

Dahlem GA, Naczi RC (2006) Flesh flies (Diptera: Sarcophagidae) associated with North American pitcher plants (Sarraceniaceae), with descriptions of three new species. *Annals of the Entomological Society of America*, **99**, 218–240.

Darwin C (1859) The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London, UK.

Darwin C (1875) Insectivorous plants. John Murray, London, UK.

Denlinger DL (1978) The developmental response of flesh flies (Diptera: Sarcophagidae) to tropical seasons. *Oecologia*, **35**, 105–107.

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Edwards SV, Beerli P (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Ellison AM (2001) Interspecific and intraspecific variation in seed size and germination requirements of *Sarracenia* (Sarraceniaceae). *American Journal of Botany*, **88**, 429–437.

Ellison AM, Gotelli NJ (2001) Evolutionary ecology of carnivorous plants. *Trends in Ecology & Evolution*, **16**, 623–629.

Ellison AM, Parker JN (2002) Seed dispersal and seedling establishment of *Sarracenia purpurea* (Sarraceniaceae). *American Journal of Botany*, **89**, 1024–1026.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences, USA*, **104**, 17614–17619.

Fish D, Hall DW (1978) Succession and stratification of aquatic insects inhabiting the leaves of the insectivorous pitcher plant, Sarracenia purpurea. American Midland Naturalist. 1978 Jan 1:172-83.

Fitzpatrick BM (2009) Power and sample size for nested analysis of molecular variance. *Molecular Ecology*, **18**, 3961–3966.

Foelix RF (1982) Biology of spiders. Harvard University Press.

Folkerts D (1999) Pitcher plant wetlands of the southeastern United States. Pp. 247–275 *in* Batzer DP, Rader RB, Wissinger SA, eds. Invertebrates in Freshwater Wetlands of North America: Ecology and Management. John Wiley and Sons, Inc., New York, NY.

Forsyth AB, Robertson RJ (1975) K reproductive strategy and larval behavior of the pitcher plant sarcophagid fly, *Blaesoxipha fletcheri*. *Canadian Journal of Zoology*, **53**, 174–179.

Francis RM (2016) pophelper: an r package and web app to analyse and visualize population structure. *Molecular Ecology Resources*. *in press*.

Funk WC, Caldwell JP, Peden CE, Padial JM, De la Riva I, Cannatella DC (2007) Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. *Molecular Phylogenetics and Evolution*, **44**, 825–837.

Gascon C, Malcolm JR, Patton JL, da Silva MN, Bogart JP, Lougheed SC, Peres CA, Neckel S, Boag PT (2000) Riverine barriers and the geographic distribution of Amazonian species. *Proceedings of the National Academy of Sciences, USA*, **97**, 13672–13677.

Gordon A, Hannon GJ (2010) "Fastx-toolkit." FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit/

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Charlesworth B, Keightley PD (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature*, **445**, 82–85.

Hafner DJ, Riddle BR (2005) Mammalian phylogeography and evolutionary history of northern Mexico's deserts. *Biodiversity, ecosystems, and conservation in northern Mexico (J.-LE Cartron, G. Ceballos, and RS Felger, eds.). Oxford University Press, New York*, 225–245.

Hickerson MJ, Stahl EA, Lessios HA (2006) Test for simultaneous divergence using approximate bayesian computation. *Evolution*, **60**, 2435–2453.

Hickerson MJ, Stahl E, Takebayashi N (2007) msBayes: pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, **8**, 268.

Jackson ND, Austin CC (2010) The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution*, **64**, 409–428.

Jones FM (1921) Pitcher plants and their moths. *Natural History*, **21**, 297–316.

715 Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the
716 Isthmus of Panama. *Proceedings of the Royal Society B: Biological Sciences*, **265**,
717 2257–2263.
718 Koopman MM, Carstens BC (2010) Conservation genetic inferences in the carnivorous
719 pitcher plant *Sarracenia alata* (Sarraceniaceae). *Conservation Genetics*, **11**,
720 2027–2038.
721 Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across
722 multiple spatial scales for three species of insects. *Oikos*, **103**, 153–161.
723 Leaché AD, Crews SC, Hickerson MJ (2007) Two waves of diversification in mammals
724 and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biology*
725 *Letters*, **3**, 646–650.
726 Lee RE, Chen C, Denlinger DL (1987) A rapid cold-hardening process in
727 insects. *Science*, **238**, 1415–1417.
728 Leigh EG, O'Dea A, Vermeij GJ (2014). Historical biogeography of the Isthmus of
729 Panama. *Biological Reviews*, **89**, 148–172.
730 Lessios HA (2008) The great American schism: divergence of marine organisms after the
731 rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and*
732 *Systematics*, 63–91.
733 Lindell J, Ngo A, Murphy RW (2006) Deep genealogies and the mid-peninsular seaway
734 of Baja California. *Journal of Biogeography*, **33**, 1327–1331.
735 Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for
736 connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–
737 299.
738 Mann CJ, Thomas WA (1968) The ancient Mississippi River. *Gulf Coast Association of*
739 *Geological Societies Transactions*, **18**, 187–204.
740 Moon DC, Rossi A, Stokes K, Moon J (2008) Effects of the pitcher plant mining moth
741 Exyra semicrocea on the hooded pitcher plant Sarracenia minor. *American Middle*
742 *Naturalist*, **159**, 321–326.
743 Nielsen R (2000) Estimation of population parameters and recombination rates from
744 single nucleotide polymorphisms. *Genetics*, **154**, 931–942.
745 Papadopoulou A, Knowles LL (2015) Species-specific responses to island connectivity
746 cycles: refined models for testing phylogeographic concordance across a
747 Mediterranean Pleistocene aggregate island complex. *Molecular Ecology*, **24**,
748 4252–4268.
749 Papadopoulou A, Knowles LL (2016) Toward a paradigm shift in comparative
750 phylogeography driven by trait-based hypotheses. *Proceedings of the National*
751 *Academy of Sciences, USA*, **113**, 8018–8024.
752 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double digest
753 RADseq: an inexpensive method for de novo SNP discovery and genotyping in
754 model and non-model species. *PloS ONE*, **7**, e37135.
755 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
756 multilocus genotype data. *Genetics*, **155**, 945–959.
757 Puechmaille SJ (2016) The program structure does not reliably recover the correct
758 population structure when sampling is uneven: subsampling and new estimators
759 alleviate the problem. *Molecular Ecology Resources*, **16**, 608–627.

Pyron RA, Burbrink FT (2010) Hard and soft allopatry: physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography*, **37**, 2005–2015.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rango JJ (1999) Resource dependent larviposition behavior of a pitcher plant flesh fly, *Fletcherimyia fletcheri* (Aldrich)(Diptera: Sarcophagidae). *Journal of the New York Entomological Society*, **107**, 82–86.

Rasic G, Keyghobadi N (2012). The pitcher plant flesh fly exhibits a mixture of patchy and metapopulation attributes. *Journal of Heredity*, **103**, 703–710.

Rehm P, Borner J, Meusemann K, von Reumont BM, Simon S, Hadrys H, Misof B, Burmester T (2011) Dating the arthropod tree based on large-scale transcriptome data. *Molecular Phylogenetics and Evolution*, **61**, 880–887.

Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proceedings of the National Academy of Sciences, USA*, **97**, 14438–14443.

Satler JD, Zellmer AJ, Carstens BC (2016) Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system. *PeerJ*, **4**, e1576.

Satler JD, Carstens BC (2016) Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. *Evolution*, **70**, 1105–1119.

da Silva MNF, Patton JL (1998) Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular Ecology*, **7**, 475–486.

Smith CI, Tank S, Godsoe W, Levenick J, Strand E, Esque T, Pellmyr O (2011) Comparative phylogeography of a coevolved community: concerted population expansions in Joshua Trees and four yucca moths. *PLoS ONE*, **6**, e25628.

Smith BT, Amei A, Klicka J (2012) Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the Isthmus of Panama. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3520–3526.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83–95.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Spitzer K, Rejmánek M, Soldán T (1984) The fecundity and long-term variability in abundance of noctuid moths (Lepidoptera, Noctuidae). *Oecologia*, **62**, 91–93.

Stephens JD, Santos SR, Folkerts DR (2011) Genetic differentiation, structure, and a transition zone among populations of the pitcher plant moth *Exyra semicrocea*: implications for conservation. *PLoS ONE*, **6**, e22658.

Stephens JD, Folkerts DR (2012) Life history aspects of *Exyra semicrocea* (pitcher plant moth)(Lepidoptera: Noctuidae). *Southeastern Naturalist*, **11**, 111–126.

Stephens JD, Rogers WL, Heyduk K, Cruse-Sanders JM, Determann RO, Glenn TC, Malmberg RL (2015) Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Molecular Phylogenetics and Evolution*, **85**, 76–87.

Sukumaran J, Holder MT (2010) DendroPy: a python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.

Sullivan J, Arellano E, Rogers DS (2000) Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent response to past climatic fluctuations. *The American Naturalist*, **155**, 755–768.

Thomé MCT, Carstens BC (2016) Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences, USA*, *in press*.

Upton DE, Murphy RW (1997) Phylogeny of the side-blotched lizards (Phrynosomatidae: Uta) based on mtDNA sequences: support for a midpeninsular seaway in Baja California. *Molecular Phylogenetics and Evolution*, **8**, 104–113.

Wallace AR (1852) On the monkeys of the Amazon. *Proceedings of the Zoological Society of London*, **20**, 107–110.

Zellmer AJ, Hanes MM, Hird SM, Carstens BC (2012) Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Systematic Biology*, **61**, 763–777.

Data Accessibility

SNP files, data sets, and scripts are available on Dryad (####).

Figure legends

Figure 1. Distribution of *Sarracena alata* in the southeastern United States, with the Mississippi River represented on the map. Locales are as follows: Sundew (S), Pitcher (P), Bouton Lake (B), Red Dirt (R), Cooter's Bog (C), Kisatchie (K), Lake Ramsey (L), Abita Springs (A), Talisheek (T), De Soto (D), Franklin Creek (F), and Tibbie (Tb). Arthropods vary in the number of locales in which they are represented (see Table 1); details on the sampling distribution of each species can be found in Table S1.

Figure 2. Models used in FSC2 analyses, all variations of the isolation-with-migration model (panel C). Models varied in their included parameters, from divergence, to migration, to population size change. These models encompass several evolutionary scenarios for the species, and were selected to allow for model-selection tests prior to parameter estimation. This allows for model uncertainty to be taken into account, necessary for accurate parameter estimation in model-based inference. Models are as follows: A) Isolation only (ISO), B) Isolation with population size change in daughter populations (ISOc), C) IM model with symmetric migration (IM), D) IM model with symmetric migration and population size change in daughter populations (IMc), E) IM model with migration from west to east (IM_{WE}), F) IM model with migration from east to west (IM_{EW}), G) Island model (Island).

Figure 3. STRUCTURE results showing clustering of individuals at the $K = 2$ level for each species. Each column represents an individual. Two-letter codes below plots correspond to sampling locality.

Figure 4. Divergence time estimates from FSC2. Results show estimates of divergence times in years across the Mississippi River for each of the ten replicated data sets. Mean and 95% confidence intervals are presented from model-averaging with the unlinked AFS for each species. The host pitcher plant is estimated to have diverged at least 120,000 years before present (Zellmer *et al.* 2012).

Figure 5. Effective population size estimates from populations on either side of the Mississippi River from FSC2. Results are from the ten replicated data sets. Mean and 95% confidence intervals are presented from model-averaging with the unlinked AFS for each species.

Figure 6. Influence of generation length on divergence time estimates. Presented are estimated divergence time values (mean and 95% confidence intervals) for *S. sarraceniae* from model-averaging and unlinked AFS, scaled by number of generations per year. Between one and three generations per year would result in a divergence time similar to estimates in *S. alata*, suggesting co-diversification. This demonstrates that our inferences

are dependent on the values assumed, and highlights the difficulties inherent to conducting comparative phylogeographic investigations using parameter estimates, especially when species are from taxonomically disparate groups.

Supplemental Figure legends

Figure S1. STRUCTURE results for *F. celarata* at the $K = 2$ level, after subsampling the eastern locales to match in sample size with the west. We subsampled two individuals from each of the three eastern locales, and repeated this process to get 10 subsampled data sets. Results show a strong east–west genetic clustering across nearly all of the data sets.

Figure S2. Divergence time estimates from FSC2. Results show estimates of divergence times in years across the Mississippi River for each of the ten replicated data sets. Mean and 95% confidence intervals are presented from an isolation-with-migration model with the linked AFS for each species. The host pitcher plant is estimated to have diverged at least 120,000 years before present (Zellmer *et al.* 2012).

Figure S3. Effective population size estimates from populations on either side of the Mississippi River from FSC2. Results are from the ten replicated data sets. Mean and 95% confidence intervals are presented from an isolation-with-migration model with the linked AFS for each species.

Table 1. Sampled arthropod information. Five species were sampled from the community. Association with the plant, dispersal ability, and where they were sampled are presented here.

Species	Type	Host Association	Dispersal Ability	Sampled Locales
E. semicrocea	Moth	Specialist	Weak	B,S,P,C,R – A,L,T,D,F,Tb
S. sarraceniae	Fly	Specialist	Weak	B,P,C,K – L,T,D,Tb
F. celarata	Fly	Specialist	Weak	C – T,D,Tb
M. formosipes	Spider	Generalist	Strong	B,P,C,R,K – A,D
P. viridans	Spider	Generalist	Strong	B,P,C,R,K – A,L,T,D,F,Tb

Table 2. Genomic sequencing data. Samples were processed through Pyrad. Loci were present in at least 75% of samples for all species.

Species	Samples (N)	Reads[1]	Clusters[2] at 88%	Loci	Variable Sites
E. semicrocea	26	74161645	4480 (1602 – 10275)	715	1724
S. sarraceniae	24	41434321	2291.5 (1336 – 6704)	383	962
F. celarata	23	28878943	1617 (736 – 3884)	440	617
M. formosipes	22	21859081	4615 (1049 – 13291)	579	1953
P. viridans	25	30424227	4195 (1243 – 6835)	1037	2055

[1]Reads that passed quality filters. [2]Clusters with at least six reads; median and range are reported.

Table 3. AMOVA results. Samples were partitioned by locality and within either side of the Mississippi River. Significance was assessed with 10,000 bootstrap replicates. * represents a *p*-value between 0.05 and 0.01; ** represents a *p*-value below 0.01. For *F. celarata*, significance was not assessed at levels other than φ_{ST} due to small number of permutations (see Fitzpatrick 2009).

Species	Among locales	φ_{ST}	Among locales within regions	φ_{SC}	Within regions	φ_{CT}
E. semicrocea	34.38	0.46**	11.99	0.18**	53.63	0.34**
S. sarraceniae	16.53	0.27**	10.78	0.13**	72.70	0.17*
F. celarata	14.64	0.19**	3.92	0.05	81.44	0.15
M. formosipes	-2.78	0.15	17.94	0.17	84.84	-0.03
P. viridans	0.73	0.14	13.70	0.14	85.58	0.01

Table 4. Summary statistics, with samples partitioned west (W) and east (E) of the Mississippi River. Mean values and standard deviation for segregating sites (SS), nucleotide diversity (π), Watterson's theta (Θ_w), and Tajima's *D*.

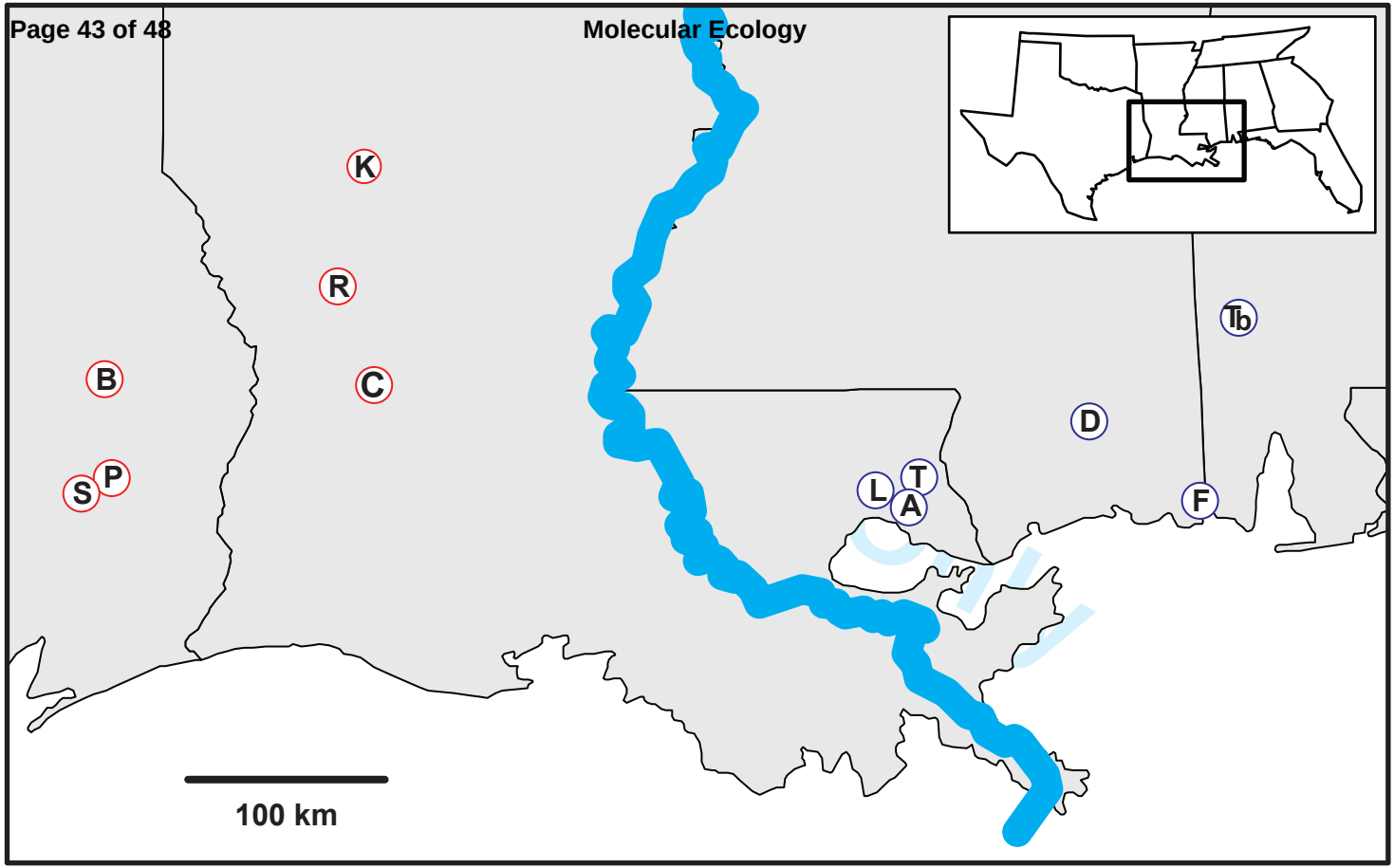
Species	**SS**		**π**		**Θ_w**		**Tajima's *D***	
	W	**E**	**W**	**E**	**W**	**E**	**W**	**E**
E. semicrocea	1.17 (1.77)	2.36 (2.47)	0.0033 (0.0058)	0.0075 (0.0094)	0.3210 (0.4952)	0.6431 (0.6876)	-0.26 (0.72)	-0.26 (0.81)
S. sarraceniae	1.62 (2.22)	2.23 (2.36)	0.0049 (0.0076)	0.0064 (0.0079)	0.4529 (0.6253)	0.6184 (0.6601)	-0.28 (0.77)	-0.44 (0.77)
F. celarata	0.51 (1.01)	1.80 (1.94)	0.0025 (0.0052)	0.0045 (0.0054)	0.1856 (0.3734)	0.4490 (0.4896)	-0.03 (0.51)	-0.38 (0.62)
M. formosipes	3.55 (3.02)	2.12 (2.34)	0.0107 (0.0099)	0.0083 (0.0099)	0.9597 (0.8163)	0.7065 (0.7828)	-0.40 (0.71)	-0.31 (0.77)
P. viridans	1.76 (1.95)	1.96 (2.07)	0.0052 (0.0065)	0.0056 (0.0068)	0.4945 (0.5487)	0.5469 (0.5755)	-0.36 (0.71)	-0.42 (0.78)

Table 5. Results from model selection tests. Values represent model probabilities generated using AIC and information theory. Only models that include migration generate any substantial support. See Fig. 2 for model details.

	Species				
Model	***E. semicrocea***	***S. sarraceniae***	***F. celarata***	***M. formosipes***	***P. viridans***
1 – ISO	0.00	0.00	0.00	0.00	0.00
2 – ISOc	0.00	0.00	0.00	0.00	0.00
3 – IM	0.83	0.41	0.30	0.39	0.70
4 – IMc	0.00	0.00	0.00	0.00	0.00
5 – IM_{WE}	0.15	0.57	0.03	0.58	0.20
6 – IM_{EW}	0.01	0.00	0.60	0.03	0.10
7 – Island	0.01	0.02	0.07	0.00	0.00

Table 6. Population genetic parameters estimates from FSC2 from model-averaging and unlinked AFS data sets. Divergence times (τ) are in years, scaled by number of generations per year, and migration rates are in 2Nm. Values were averaged across the ten replicated data sets within each species.

Species	τ		N_e WEST		N_e EAST		M_{WE}		M_{EW}	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
E. semicrocea	230,745	213,555 – 247,935	102,907	86,586 – 119,228	520,215	482,101 – 558,329	0.18	0.12 – 0.24	0.66	0.35 – 0.97
S. sarraceniae	195,045	145,995 – 244,096	124,174	89,668 – 158,680	693,356	567,537 – 819,174	1.57	1.08 – 2.06	0.75	0.00 – 1.49
F. celarata	90,598	84,668 – 96,528	168,863	114,681 – 223,046	314,109	259,511 – 368,708	0.19	0.00 – 0.38	3.72	2.94 – 4.51
M. formosipes	473,450	435,551 – 511,349	556,845	441,555 – 672,135	764,333	539,372 – 989,295	5.61	4.36 – 6.86	0.86	0.41 – 1.31
P. viridans	244,133	235,180 – 253,086	333,666	235,139 – 432,194	394,613	299,784 – 489,442	5.54	3.52 – 7.56	4.08	2.65 – 5.50


K
R
B
C
S
P
L
T
A
D
Tb
F
100 km