Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system

Jordan D Satler, Amanda J Zellmer, Bryan C Carstens

Understanding if the members of an ecological community have co-diversified is a central concern of evolutionary biology, as co-diversification suggests prolonged association and possible coevolution. By sampling associated species from an ecosystem, researchers can better understand how abiotic and biotic factors influence diversification in a region. In particular, studies of co-distributed species that interact ecologically can allow us to disentangle the effect of how historical processes have helped shape community level structure and interactions. Here we investigate the *Sarracenia alata* pitcher plant system, a community where many species from disparate taxonomic groups live in a mutualistic relationship inside the fluid-filled pitcher leaves. Direct sequencing of the eukaryotes present in the pitcher plant fluid enables us to better understand how a host plant can shape and contribute to the genetic structure of its associated inquilines, and to ask whether genetic variation in the associated taxa are structured in a similar manner to the host plant. We combine metagenomics with 454 sequencing to demonstrate that the pattern of genetic diversity in many, but not all, of the eukaryotic community is similar to that of *S. alata*, providing evidence that associated eukaryotes share an evolutionary history with the host pitcher plant. Our work provides further evidence that a host plant can influence the evolution of its associated commensals.

RUNNING HEAD: METAGENOMIC SAMPLING OF A PITCHER PLANT COMMUNITY

Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system

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

[1] *Department of Evolution, Ecology and Organismal Biology,*

The Ohio State University, Columbus, Ohio 43210, USA

[2]*Department of Biology*

Occidental College, Los Angeles, California 90041, USA

[*]**Corresponding Author:**

Department of Evolution, Ecology and Organismal Biology, 318 W. 12th Avenue,

The Ohio State University, Columbus, Ohio 43210, USA

Email: carstens.12@osu.edu

Abstract

Understanding if the members of an ecological community have co-diversified is a central concern of evolutionary biology, as co-diversification suggests prolonged association and possible coevolution. By sampling associated species from an ecosystem, researchers can better understand how abiotic and biotic factors influence diversification in a region. In particular, studies of co-distributed species that interact ecologically can allow us to disentangle the effect of how historical processes have helped shape community level structure and interactions. Here we investigate the *Sarracenia alata* pitcher plant system, an ecological community where many species from disparate taxonomic groups live inside the fluid-filled pitcher leaves. Direct sequencing of the eukaryotes present in the pitcher plant fluid enables us to better understand how a host plant can shape and contribute to the genetic structure of its associated inquilines, and to ask whether genetic variation in the taxa are structured in a similar manner to the host plant. We used 454 amplicon-based metagenomics to demonstrate that the pattern of genetic diversity in many, but not all, of the eukaryotic community is similar to that of *S. alata*, providing evidence that associated eukaryotes share an evolutionary history with the host pitcher plant. Our work provides further evidence that a host plant can influence the evolution of its associated commensals.

Introduction

Dynamic processes during the Pleistocene epoch have been implicated as drivers of biological diversification (e.g., Hewitt, 2000; Hewitt, 2004). Glacial cycles contributed to both landscape changes and climatic oscillations, providing strong abiotic factors that have led to speciation within many groups (e.g., Leaché & Fujita, 2010; McCormack et al., 2010). One region strongly influenced by these processes during the Quaternary is the southeastern United States, where decades of research has examined the structure of genetic variation in a diverse set of taxa (e.g., Avise et al., 1987; Avise, 2000; Burbrink, Lawson & Slowinski, 2000; Weisrock & Janzen, 2000; Jackson & Austin, 2010; Newman & Rissler, 2011). Although glaciers never extended to this latitude, changes in both flow rate and direction of flow of major rivers coupled with fluctuations in sea level influenced phylogeographic patterns in this region (reviewed in Soltis et al., 2006). Specifically, major rivers in the region have produced population genetic structure in many clades, with the Mississippi River recognized as a well-characterized biogeographic barrier (Brant & Orti, 2003; Pyron & Burbrink, 2009). The influence of landscape features coupled with the presence of large-scale barriers can be expected to isolate populations within a species, especially those with limited dispersal abilities. Consequently, plants and animals that lack the ability to traverse large bodies of water are expected to exhibit substantial population genetic structure in this region.

Complex interactions that occur within ecological communities can influence the formation and maintenance of biodiversity. For example, numerous studies have shown how host plant diversification can contribute to the diversification of associated species, typically insects (e.g., Farrell & Mitter, 1990; Wheat et al., 2007; McKenna et al., 2009; Espindola et al. 2014). These include systems where plants evolve secondary compounds in an “escape and radiate”

model of coevolution (Erlich & Raven, 1964), and systems that include mutualist organisms such as plants and their pollinators. Such interactions can result in congruent demographic histories (e.g., Smith et al., 2011) and patterns of co-diversification (e.g., Rønsted et al., 2005). While it seems clear that the ecological interactions among plants and associated arthropods (e.g., herbivores and pollinators) can potentially drive patterns of co-diversification, it is unclear how host plants may influence other commensal organisms, particularly small eukaryotes. Communities of commensal organisms in both facultative and obligate relationships may be expected to show varying evolutionary patterns attributed to the level of dependency on the host plant. Given the dynamic and topologically complex landscape of the southeastern region, the study of ecological communities that span the breadth of host affinity, dispersal ability and life history traits can help inform how taxonomically diverse communities have assembled through time, and whether present day ecological associations extend into the deep past.

Phytotelmata—water bodies contained within living plants—provide an ideal system for investigating co-diversification within an ecological community because they are self contained and discrete units (Kitching, 2000). Carnivorous pitcher plants are one such system, where decades of ecological work have documented a complex and distinct ecosystem associated with the pitcher fluid contained within the modified leaves. Pitcher plants in the genus *Sarracenia* (F. Sarraceniaeceae) contain a diverse microbiome, including groups such as bacteria, algae, protists, rotifers and arthropods (e.g., Folkerts, 1999; Miller & Kneitel, 2005; Peterson et al., 2008; Koopman et al., 2010). Their highly modified leaves form a trap that captures and digests prey items, while also providing a unique habitat for commensal organisms. Associated inquilines form complex relationships in the pitchers, with many supplying digestive enzymes that help break down decomposing prey items providing inorganic compounds for the plant (see

Adlassnig, Peroutka, & Lendl, 2010). A wide range of ecological work has investigated the communities associated with these plants, primarily in *Sarracenia purpurea*, showing community structure and interactions among the inquilines (e.g., Addicott, 1974; Bradshaw & Creelman, 1984; Buckley et al., 2003; Gotelli & Ellison, 2006; terHorst, Miller & Levitan, 2010; Miller & terHorst, 2012). Here, we focus on the Pale Pitcher Plant *Sarracenia alata*, a species distributed in patchy habitats along the gulf coast across eastern Texas, Louisiana, Mississippi and Alabama. This species is largely isolated from its congeners and occupies disjunct eastern and western regions across the Mississippi River (Fig. 1). Work by Koopman & Carstens (2010) identified population genetic structure in *S. alata*, and Zellmer et al. (2012) showed that major rivers in the region promoted diversification within the plant. Population divergence across either side of the Mississippi River is likely well into the Pleistocene, and estimated at greater than 120,000 years before present (Zellmer et al., 2012). Further analysis suggests that *S. alata* may contain two cryptic species, corresponding to populations on the eastern and western sides of the Mississippi River (Carstens & Satler, 2013). *Sarracenia alata* thus represents a particularly attractive system for investigating patterns of co-diversification, because the species exhibits strong genetic differentiation across the landscape, with significant divergence across an important biogeographic barrier (Soltis et al., 2006). In addition, longleaf pine savannahs in the south have seen a staggering amount of habitat loss in recent times (~1% of its original habitat remains; Noss, 1989). High levels of cryptic genetic diversity highlight *S. alata* as a species of interest; identifying ecologically associated taxa with a shared evolutionary history has clear conservation implications.

Phylogeographic investigations of co-distributed taxa are usually limited to particular taxonomic groups (e.g., Bell et al., 2012; Fouquet et al., 2012; Smith, Amei & Klicka, 2012;

116 Hope et al., 2014). While these studies can reveal evolutionary processes that produce patterns
117 within biogeographic regions, the conclusions drawn from such findings can be limited by the
118 shared life history traits that influence the formation of genetic structure (e.g., dispersal ability,
119 population size). Metagenomics provides a powerful approach for efficiently and rapidly
120 sampling taxonomic diversity within a habitat (reviewed in Tringe & Rubin, 2005), and thus may
121 provide comparative phylogeographic investigations with an efficient approach to the sampling
122 of taxa. Through the sequencing of environmental DNA, communities of small to microscopic
123 organisms can be directly sampled from the environment resulting in the assemblage of a data set
124 spanning a wide taxonomic breadth. Thus, when coupled with next generation sequencing
125 methods (Mardis, 2008), metagenomics greatly increases the "taxonomic toolbox" lending itself
126 well to investigations of comparative phylogeography. By analyzing a disparate assemblage of
127 taxa comprising an ecological community, our work has the potential to reveal a shared response
128 to historical events and thus evidence that evolutionary processes can shape community structure
129 and interactions through time (Smith et al., 2011). With the diverse array of microscopic
130 inquilines present within *Sarracenia* (e.g., Miller & Kneitel, 2005), pitcher plants provide an
131 ideal system for understanding how a host plant may influence genetic variation within an
132 associated community, and metagenomics provides a tool for sampling this taxonomic diversity.
133 Here we explore the process of evolutionary diversification in an ecological community.
134 We directly sample pitcher fluid from the modified leaves of *S. alata*, and apply a novel
135 approach utilizing metagenomics to test if *S. alata* has influenced genetic structure in its
136 eukaryotic commensal organisms. First, we characterize taxonomic diversity within the pitcher
137 plant fluid to get an understanding of the major lineages and their abundance in this unique
138 habitat. We then generate a comparative data set of OTUs which span the Mississippi River, and

assess the degree to which the inquiline community shares population genetic structure with the host plant. We hypothesize that if eukaryotes associated with *S. alata* are ecologically dependent on the plant, then the evolutionary history of the commensals should exhibit population genetic structure largely congruent with that of *S. alata*. Alternatively, if taxa do not share an evolutionary history with *S. alata*, community members should have unique population genetic structure indicating an idiosyncratic response to landscape processes driving diversification in the region.

Material and Methods

Genetic Sampling

Pitcher fluid samples were collected following Koopman & Carstens (2010) during the spring and summer of 2009 from part of the plant's distribution. Specifically, samples were collected from 40 individuals across four locales (Abita Springs, Cooter's Bog, Kisatchie, Talisheek; Fig. 1) in Louisiana during June and August, as pitcher diversity peaks at this time (Koopman et al., 2010). In addition, fluid was collected from ten individuals per month for five months (April through August) from Lake Ramsey, resulting in 50 samples, for a total sampling effort of 90 individuals from five locales. Sampling was originally designed to investigate both spatial (all five locales) and temporal (Lake Ramsey) dynamics, however, we focus on just spatial patterns in this study. DNA was extracted using the Powersoil DNA Isolation Kit (MO Bio). The large subunit 28S rRNA region was amplified for each fluid sample using the following primer combination (LS1F: GTACCCGCTGAACTTAAGC ; LS4R: TTGTTCGCTATCGGTCTC; modified from Hausner, Reid & Klassen, 1993), targeting a

roughly 330 base pair (bp) region. Each pitcher fluid sample was labeled with MID tags to allow for multiplexing of individuals. PCRs were performed in triplicate and then pooled to prevent PCR bias, and subsequently sequenced on a 454 Life Sciences Genome Sequencer FLX (Roche) at Engencore Genomics Facility (University of South Carolina, Columbia) using 1/8th of a plate. Raw sequences were initially processed using Mothur (Schloss et al., 2009) to sort sequences by individual, remove low quality reads, and identify unique sequences for each individual. Chloroplast data for *S. alata* was gathered from a previous study (see Carstens & Satler, (2013) for details).

Bioinformatics

To quantify the taxonomic diversity present within the pitchers, sequences were clustered into operational taxonomic units (OTUs) through *de novo* assembly. Metagenomic studies commonly use *de novo* assembly for generating OTUs (e.g., O'Brien et al., 2005; Bik et al., 2012; Zimmerman & Vitousek, 2012), and this allowed for a rough characterization of the number of taxonomic units present within the pitcher fluid.

Sequences were first combined within each of the sampling locales (i.e., we restricted clustering to those sequences collected from within each sample site), thereby treating each of the five sites as a separate population (see Fig. 1). Reads were trimmed to 275 bp, discarding any sequences below this threshold to remove potential bias associated with clustering samples of unequal sequence size. For consistency, we only analyzed sequences from Lake Ramsey collected at the same time periods as from the other sampling sites. Trimmed sequences were assembled into clusters using the UPARSE algorithm (Edgar, 2013); this pipeline been shown to outperform commonly used clustering methods such as Mothur and QIIME, and to work well

under a solely *de novo* clustering approach. Within each locale, identical reads were collapsed and abundance values recorded (i.e., the number of times each unique read appeared in the data set). Sequences were then clustered based upon a 97% threshold, with the most represented sequences (based on abundance values) used to form initial OTU clusters, using a dynamic programming algorithm to find clusters with the maximum score. The percent similarity threshold is subjective, but since it is required for *de novo* assembly, we justify this value by noting that (i) it was recommended by the author for *de novo* assembly in UPARSE (Edgar, 2013), (ii) it falls within the range used to delimit fungi with this locus (see Sota et al., (2014) and references within), a group expected to be well represented within the pitcher fluid, and (iii) chimeric detection is increasingly difficult when this value is decreased. The clustering step in the pipeline ("cluster_otus") uses UPARSE-OTU, an algorithm that simultaneously determines the OTU clusters while removing chimeric sequences from the data set, a potential problem due to errors with pyrosequencing.

Following OTU clustering, a single sequence from each cluster was used with a Basic Local Alignment Search Tool (BLAST) search to gather taxonomic identification for each of the clusters. Although there is a concern with the incompleteness of public databases, and that searches could return spurious matches (Koski & Golding, 2001; Tringe & Rubin, 2005), at a higher taxonomic level (e.g., Class, Order), we can be reasonably confident that sequence matches reveal organismal affinity. In addition, for the purposes of this study, a qualitative assessment of higher level identification is sufficient to understand taxonomic diversity present within the pitcher fluid. A custom python script was used to search for taxonomic identities among the OTUs. For all BLAST searches, sequences representing the centroid of the original OTU searches (in UPARSE) were queried against the NCBI nucleotide database, using the

208 Megablast search algorithm, saving the top hit from each search. OTUs with BLAST hits were
209 grouped by higher-level identification, generally at the level of Class or Order, to identify the
210 variety of organisms present within the pitcher fluid. After summarizing taxonomic identity
211 within the pitcher fluid at each site, rarefaction curves were generated with the package vegan
212 (Oksanen et al., 2015) in R (R Core Team 2015), as a means to test if the taxonomic diversity
213 had been adequately captured with our sampling efforts.
214
215 **Population Structure**
216 The major goal of this study is to identify OTUs that span the Mississippi River, and test
217 if the landscape processes that have influenced diversification in *S. alata* have influenced the
218 sampled organisms in a similar manner. To generate a comparative data set, all raw sequences
219 were combined and OTUs were assembled with UPARSE following the steps outlined above
220 (i.e., all sequences were clustered in a global analysis, regardless of sampling location). This data
221 set included all sequences generated from Lake Ramsey, as we were interested in collecting taxa
222 with widespread distributions. If taxa were time dependent, they would be restricted to Lake
223 Ramsey (during the months when only this locality was sampled) and removed following our
224 filtering process (see below); however, taxa stable in these communities would comprise
225 additional sequence information for comparative analysis. Following initial OTU clustering, the
226 data set was reduced to those taxa that contained at least ten sequences per OTU with a minimum
227 of three sequences on either side of the Mississippi River. These thresholds were used to
228 maximize the number of OTUs represented in the final data set while still containing enough
229 sequence data for statistical inference, both within and across sampling sites. In addition, it is
230 expected that any potential chimeric sequences not removed in the clustering step will fall below

these thresholds, further reducing the potential for error with our final OTUs. Each OTU was aligned with MAFFT (Katoh & Standley, 2013), using either the L-INS-i (< 200 sequences) or FFT-NS-i (> 200 sequences) algorithm. To survey taxonomic diversity among the retained OTUs, a BLAST search was conducted on each of the OTUs following the same steps as outlined above.

Data were summarized within each of the OTUs in order to characterize genetic variation and quantify population genetic structure. Standard population genetic summary statistics (nucleotide diversity (π), Watterson's theta (Θ_w), and Tajima's *D*) were calculated with the package Pegas (Paradis, 2010) in R. Several approaches were used to explore the partitioning of genetic variation among the OTUs. G_{ST} (Nei, 1973) values were generated to estimate the degree of population differentiation among the locales, and were calculated with the R package gstudio (Dyer, 2012). The level of genetic partitioning was assessed with an analysis of molecular variation (AMOVA; Excoffier, Smouse & Quattro, 1992), because the G_{ST} is an analog to F_{ST} values (Nei, 1973). AMOVAs take into account the amount of variation in the sequence data, thereby extracting more information to determine the level of spatial structuring within the taxa. AMOVAs were calculated in the program SPADS (Dellicour & Mardulyn, 2013), with 10,000 permutations to generate levels of significance. Hierarchical levels tested included (i) sampling locales within each region (i.e., side of the Mississippi River), (ii) sampling locales within total distribution, and (iii) between regions. In addition, the amount of allelic sorting on either side of the Mississippi River was calculated using the genealogical sorting index (GSI; Cummings, Neel & Shaw, 2008). This method is commonly applied to tests of taxonomic distinctness; it is applied here to quantify levels of lineage sorting within each side of the river, with higher levels of sorting suggesting greater population genetic structure indicative of a longer period of population

254 isolation. GSI values range from 0 (no sorting) to 1 (monophyletic on either size of barrier), with
255 p-values indicating the extent to which genetic structure recovered is more than would be
256 expected by chance alone. An input genealogy is required to calculate the GSI; these were
257 estimated using Maximum Likelihood (ML) with RAxML v7.2.8 (Stamatakis, 2006; Stamatakis,
258 Hoover & Rougemont, 2008). Depending on the number of sequences in the OTU, models of
259 sequence evolution included either GTRCAT (> 200 sequences) or GTRGAMMA (< 200
260 sequences). Each ML tree was then input to the GSI web server, with 10,000 permutations to
261 generate levels of significance. In addition, isolation by distance (IBD) values were calculated to
262 see if there was a correlation between genetic and geographic distance, using the IBDWS v3.23
263 web server (Jensen, Bohonak & Kelley, 2005). Genetic distance matrices were calculated using a
264 Kimura 2-parameter (K2P) substitution model for each OTU; geographic matrices were
265 constructed measuring the Euclidean distance between sampling locales in kilometers with the
266 distance measurement tool in Google Earth (www.google.com/earth/, last accessed 18 July
267 2015). Finally, we used a chi-squared goodness of fit test to see if the number of OTUs with
268 significant population genetic structure across the various analyses was more than would be
269 expected by chance alone (assuming $\alpha = 0.05$). This allowed us to test the null hypothesis that
270 there is no correlation of population structure between the members of the eukaryotic community
271 and the host plant.
272
273
274 **Results**
275 **Genetic Sampling**

276 High-throughput sequencing resulted in a total of 26,399 sequences across 90 sampled
277 pitchers. Following demultiplexing and quality control of samples, an average of 101 unique
278 sequences were retained per pitcher (range: 12 – 199) for a total of 9,045 sequences. A fasta file
279 containing all 9,045 sequences, as well as all OTU matrices from the comparative data set (see
280 below), has been deposited at Dryad (####).
281
282 **Taxonomic Diversity**
283 To remove biases associated with the clustering of length variable sequences, all
284 sequences were trimmed to 275 bp (discarding any reads below this threshold), reducing the data
285 set from 9,045 sequences to 8,991 sequences. Lake Ramsey contained a disproportionately larger
286 percentage of the total number of sequences (49%); however, to compare samples collected from
287 the same time periods, we only analyzed those samples from June and August, reducing the
288 number of sequences from Lake Ramsey from 4,398 to 2,286, resulting in a total of 6,879
289 sequences. OTU clustering at the 97% sequence identity within each locale resulted in a median
290 of 66 OTUs per sample site (324 total), ranging from 48 (Cooter's Bog) to 82 (Lake Ramsey)
291 total OTUs, with an average of 21 sequences per OTU when averaged across all sites. The
292 majority of OTUs had a close hit in the BLAST search (97%), although a small number of OTUs
293 (13) did not contain a match in the database (Fig. 2). Taxonomic diversity ranged across the tree
294 of life, with many OTUs containing hits to fungi, and to a lesser extent, various arthropod
295 groups, including insects and mites. In addition, numerous other groups were recovered in the
296 searches, including protozoans, nematodes, an annelid and even a vertebrate (*Sus scrofa*, wild
297 boar). Rarefaction curves for each sample site suggest that OTU diversity has not yet been
298 reached, indicating that the pitcher plant community was not fully sampled in any of the sites

(Fig. 3). Although fewer sequences per site likely prevented us from obtaining representatives from the full diversity of species within each pitcher, wider spatial sampling helped us achieve our goal of sampling a large number of eukaryotic species for a comparative data set (see below),

Population Structure

A global clustering effort was completed to generate a comparative data set for taxa that span the Mississippi River. As we were interested in widespread taxa, we used all sequences collected from Lake Ramsey—including those collected from additional time periods—resulting in the use of the full data set (8,991 sequences). Following *de novo* clustering, UPARSE produced 323 OTUs of which 65 contained a minimum of ten sequences and of these, 31 OTUs contained at least three representatives on either side of the river. BLAST hits of a single sequence from each of the 31 OTUs indicate that fungi and mites are the most well represented taxa (Table 1). One OTU did not contain a significant BLAST hit, and with parameters relaxed, poorly matched a portion of the sequence to multiple disparate taxonomic groups. Since we detected it in multiple pitchers, it seems unlikely that this OTU represents a chimeric sequence. Given the incompleteness of taxonomic databases, however, we retained this OTU for downstream analysis, resulting in a final dataset of 31 OTUs (see Table S1 for the sequencing distribution among locales). In this final set, the number of sequences per OTU ranged from 14 to 2,507, with a median of 54 (average of 225 sequences; Table 1).

A range of genetic variation is present in the sampled OTUs (Table 1). For example, estimates of nucleotide diversity (π) range from ~0.001 to 0.05, a fifty fold difference. Tajima *D* values are negative for most taxa (median = -2.0101), with 21 of these values significant, indicative of an excessive number of segregating sites in the data sets. Negative Tajima *D* values

322 can be interpreted as resulting from a rapid demographic expansion, or from natural selection, on
323 the marker itself or on a linked gene. This could also be the result of population structure in those
324 OTUs, as collapsing separate populations can increase the number of segregating sites in a taxon.
325 Among taxonomic groups, all fungi have a negative Tajima *D* value, with the majority (73%)
326 being significant. Of note are the Tajima *D* values for the arthropods, where all three insects have
327 significantly negative values, and all seven mites have negative values, with three out of seven
328 being significant.

329 There are varying levels of population structure across the taxonomic groups. Roughly
330 half of the fungi contain significant partitioning of genetic variance at the level of the sampling
331 locale, with two taxa also significant at the level of locales within regions (Fig. 4; Table S2).
332 Sequence-based F statistics display similar patterns, with G_{ST} values ranging from 0.003 – 0.280
333 (average G_{ST} = 0.101), suggesting population genetic structure is evident on either side of the
334 Mississippi River in many taxa. Despite this structure, there is considerable sharing of alleles
335 across the Mississippi River in the fungi, although some of the species contain greater sorting
336 than would be expected by chance (see GSI results; Fig. 4; Table S2). Furthermore, genetic
337 diversity in all but two of the fungi is not correlated with geographic distance (Table S2). Results
338 in the mites are similar, with roughly half of the taxa sampled showing a significant amount of
339 genetic variation distributed among the locales, as well as across the Mississippi River (Fig. 4,
340 Table S2). F statistics in the mites are slightly lower than those in the fungi (average G_{ST} =
341 0.081). This structure is also evident in the GSI results, with allelic sorting in most taxa higher
342 than would be expected by chance (Fig. 4; Table S2). Patterns among the fungi (roughly half of
343 the OTUs), mites and insects generally reflect those of the host plant, with the remaining taxa
344 showing essentially no evidence for this shared genetic structure. Chi-squared goodness of fit

345 tests show that more taxa share population genetic structure with the host plant than would be
346 expected by chance in many of the analyses (Table 2).
347
348
349 **Discussion**
350 Investigations into the evolutionary history of host plants and their associated insects
351 have provided evidence for co-diversification over long time-periods (e.g., Weiblen & Bush,
352 2002) in addition to demographic patterns that suggest a concerted response to abiotic factors
353 over shorter periods of time (e.g., Smith et al., 2011). Inspired by such studies, we sampled a
354 diverse set of organisms (representing similarly diverse ecological interactions) associated with
355 the Pale Pitcher Plant in order to investigate the extent to which this ecological community has
356 co-diversified. Within *Sarracenia alata*, previous work has demonstrated that populations are
357 genetically structured across the landscape (Koopman & Carstens, 2010). Major rivers are
358 important drivers of diversification both in the region (Soltis et al., 2006) and for *S. alata*
359 (Zellmer et al., 2012), and analysis of the barcode data presented here demonstrates that slightly
360 under half of the eukaryotes sampled share similar population genetic structure with *S. alata*.
361 Our results show that a core eukaryotic community exhibits congruent patterns of
362 population genetic structure, with many taxa displaying significant genetic structuring at the
363 level of the sampling locality (based on Φ_{ST}); approximately half of the microscopic fungi and
364 half of the mites are structured in a manner similar to that of the host plant (Fig. 4). Given the
365 dispersal capabilities of fungal spores (e.g., Peay, Kennedy & Bruns, 2008), this degree of
366 population genetic structure is strikingly high (but see Taylor et al., 2006). Fungi are ubiquitous
367 in terrestrial habitats, however, with many species associated with soils and plants. Fungal

368 species have also been recovered from pitcher plant leaves, demonstrating their known presence
369 within these microhabitats (reviewed in Adlassnig, Peroutka & Lendl, 2010). Multiple mite
370 species from the family Histiostomatidae have been described from *Sarracenia* pitcher plants
371 (Hunter & Hunter, 1964; Fashing & O'Connor, 1984), and act as prey consumers within the
372 pitchers. Approximately half of the mites identified here exhibited population genetic structure
373 similar to that of *S. alata*, reflecting structure seen both among sample sites and across the
374 Mississippi River (Fig. 4). Other members of this core group include two of the three sampled
375 insects (all three share a closest BLAST hit to ants), with general strong support across analyses
376 for co-diversification. Ants comprise a large component of prey items for *Sarracenia* (Newell &
377 Nastase, 1998; Ellison & Gotelli, 2009), and one could interpret these results as being indicative
378 of general biogeographic structuring in the longleaf pine savannah habitat. Such results are
379 illustrative of the challenges associated with understanding if interactions among organisms have
380 driven the shared responses to historical events reflected in patterns of co-diversification.
381 Although ants represent a major prey item of *Sarracenia* plants, our field work suggests
382 that the plant does not specialize on any particular species of ant. Therefore, ants in general may
383 be considered to have a relatively weak ecological association with the plant, and these results
384 may highlight the strong influence that landscape and abiotic factors have on diversification in
385 the region. Teasing these two interpretations (i.e., ecological association vs. landscape and
386 abiotic factors) apart is non-trivial, yet an understanding of the strength of ecological association,
387 habitat affinity, and dispersal ability can lend insight on this issue. Given the ecological
388 associations shared among many of the inquilines with the host plant in this system, shared
389 population structure does provide evidence that ecology plays a role in shaping diversification
390 patterns through time. As the Mississippi River is an important evolutionary barrier to this

system (and many groups across the region), diversification across the river may have taken place via the mechanism of oxbow lake formation, where changes in the river channel moved a portion of the habitat from the eastern to the western side of this barrier.

While slightly under one-half of the sampled taxa share population genetic structure with *S. alata*, there are other taxa with discordant evolutionary histories. Many of the fungal taxa exhibit little to no population structure, and we suspect that these microscopic species are widespread and not restricted to the pitcher plant bog habitats. Their dispersal ability is likely to be higher than the larger members of this community, allowing them to escape the influence of biogeographic barriers. Other microscopic eukaryotes exhibit no evidence of population structure, including two protizoans and the sampled nematodes, suggesting that biogeographic barriers do not provide an obstacle for long-distance dispersal in these taxa (Finlay, 2002). In addition, one insect species demonstrates a lack of population structure. Further investigation of the BLAST result for this OTU (hit to *Solenopsis xyloni* in original search; see Table 1) indicates that this OTU is an identical match to the invasive red fire ant (*Solenopsis invicta*). Given the devastating impact and colonization power of the red imported fire ant, the lack of population structure is likely a product of their recent introduction to the southeastern United States (from South America). *Solenopsis invicta* has grown explosively and displaced native species in the region (Porter & Savignano, 1990; Stuble, Kirkman & Carroll, 2009) and the lack of structure recovered is consistent with the expectations of an invasive species. Clearly, ecological association and dispersal ability both play a role in the level of congruence detected in population genetic structure across species, although quantifying these two factors, especially dispersal for microscopic eukaryotes, remains a challenge.

Phylogeographic patterns within a species can be informative, but in aggregate, the results across many species make it possible to identify community responses to landscape changes. To date, phylogeographic researchers have not fully utilized metagenomics as a tool for increasing the taxonomic breadth of a comparative study. The *S. alata* system is ideal for such studies, as each pitcher provides a self-contained and discreet habitat, where micro- and macroscopic organisms can live and persist in an ecological entanglement. The increased sampling facilitated by metagenomic approaches allowed us to identify a core evolutionary community within *S. alata*, and the simplest explanation for this congruence is that the core community has diversified in unison because the constituent members are ecologically dependent on *S. alata*. As such, the OTUs sampled represent an example of shared evolutionary patterns across an ecological community, and suggests that co-diversification is not limited to specialized interactions such as plants and pollinators. The recent discovery of cryptic diversity within *S. alata* (Carstens & Satler, 2013), together with the work presented here, highlights the need for conserving species like pitcher plants, which play a role in the survival of many different organisms. Such systems contain species that have been ecologically interdependent over evolutionary time scales, thus the loss of substantial diversity in the pitcher plant could lead to loss of diversity in its commensal species.

Given the power of comparative analysis for phylogeographic research, metagenomics can be leveraged to increase our knowledge of the evolutionary processes that lead to biogeographic patterns around the world. In particular, environmental sampling can provide access to taxa spanning the range of ecological and life history traits, as well as greater spatial sampling, which can provide evidence of the landscape processes that have structured species and communities in a region (Bermingham & Moritz, 1998). Potential pitfalls, however, do

436 remain when applying metagenomics for such an analysis. In this study, a large number of
437 sampled OTUs are fungi (Fig. 2), which could be indicative of their ubiquity in nature, but could
438 also be due to our use of primers originally developed from fungal genomic resources. The need
439 to isolate specific gene fragments with primers could have biased the taxonomic sampling, which
440 may have also contributed to the non-asymptotic nature of the rarefaction curves, although this is
441 more likely due to a relatively small number of sequences from next generation sequencing with
442 the sampling strategy used in this study. In addition, challenges exist when using *de novo*
443 assembly for generating a taxonomic data set, particularly with the requirement of a percent
444 threshold to determine the placement of sequences within OTUs. Although some values are
445 commonly used for certain groups, it is unlikely that a single cutoff is appropriate across the tree
446 of life. Further exploration of the correlation between sequence similarity and taxonomic identity
447 across diverse groups is necessary to better place sequences with their proper OTU. However, as
448 demonstrated here, metagenomic data can be beneficial for phylogeographic studies, with careful
449 and transparent analysis of the data providing valuable insight into the diversification of a region,
450 or in our case, an ecological community composed of a diverse set of lineages.

451 Remarkably, the co-diversification described here may extend beyond the eukaryotic
452 members of this ecosystem. Koopman & Carstens (2011) provide evidence that the phylogenetic
453 community structure in the bacterial microbiome reflects the population genetic structure of the
454 plant. Since the bacterial microbiome is dominated by Enterobacteriaceae (Koopman et al.,
455 2010), a family commonly found in animal guts, it could be that the insect members of the core
456 community facilitate colonization of bacteria in the pitchers (which are sterile before opening;
457 see Peterson et al., 2008). If the core arthropods seed the pitchers with Enterobacteria, these
458 microbes may produce enzymes that contribute to the digestive function of the pitcher. Since

459 these complex ecological interactions have likely persisted for hundreds of thousands of years
460 (based on estimates from *S. alata*), our work underscores the importance of investigating the
461 evolutionary relationships of ecological communities.
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481

Acknowledgements

We thank members of the Carstens lab, particularly Margaret Hanes (formerly Koopman) and Sarah Hird, for comments and discussion regarding the manuscript. We thank the AE and reviewers for comments that helped improve this manuscript.

505 **References**

506 Addicott JF. 1974 Predation and prey community structure: an experimental study of the effect
507 of larvae on the protozoan communities of pitcher plants. *Ecology* **55**, 475–492.
508 Adlassnig W, Peroutka M, Lendl T. 2011 Traps of carnivorous pitcher plants as a habitat:
509 composition of the fluid, biodiversity and mutualistic activities. *Annals of Botany* **107**, 181–
510 194. (doi:10.1093/aob/mcq238)
511 Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987
512 Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics
513 and systematics. *Ann. Rev. Ecol. Syst.* **18**, 489–522.
514 Avise JC. 2000 Phylogeography: the history and formation of species. Cambridge, MA: Harvard
515 University Press.
516 Bell RC, MacKenzie JB, Hickerson MJ, Chavarria KL, Cunningham M, Williams S, Moritz C.
517 2012 Comparative multi-locus phylogeography confirms multiple vicariance events in co-
518 distributed rainforest frogs. *Proc. R. Soc. B* **279**, 991–999. (doi:10.1038/rspb.2011.1229)
519 Bermingham E, Moritz C. 1998 Comparative phylogeography: concepts and applications. *Mol.*
520 *Ecol.* **7**, 367–369. (doi:10.1046/j.1365-294x.1998.00424.x)
521 Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. 2012 Sequencing our
522 way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* **27**, 233–243.
523 (doi:10.1016/j.tree.2011.11.010)
524 Bradshaw WE, Creelman RA. 1984 Mutualism between the carnivorous purple pitcher plant and
525 its inhabitants. *American Midland Naturalist* **112**, 294–304.

526 Brant SV, Orti G. 2003 Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda*
527 (Insectivora: Soricidae): past fragmentation and postglacial recolonization. *Mol. Ecol.* **12**,
528 1435–1449. (doi:10.1046/j.1365-294X.2003.01789.x)
529 Buckley HL, Miller TE, Ellison AM, Gotelli NJ. 2003 Reverse latitudinal trends in species
530 richness of pitcher-plant food webs. *Ecology Letters* **6**, 825–829. (doi:10.1046/j.1461-
531 0248.2003.00504.x)
532 Burbrink FT, Lawson R, Slowinski JS. 2000 Mitochondrial DNA phylogeography of the
533 polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept.
534 *Evolution* **54**, 2107–2118. (doi:10.1111/j.0014-3820.2000.tb01253.x)
535 Carstens BC, Satler JD. 2013 The carnivorous plant described as *Sarracenia alata* contains two
536 cryptic species. *Biol. J. Linn. Soc.* **109**, 737–746. (doi:10.1111/bij.12093)
537 Cummings MP, Neel MC, Shaw KL. 2008 A genealogical approach to quantifying lineage
538 divergence. *Evolution* **62**, 2411–2422. (doi:10.1111/j.1558-5646.2008.00442.x)
539 Dellicour S, Mardulyn P. 2014 SPADS 1.0: a toolbox to perform spatial analyses on DNA
540 sequence data sets. *Molecular Ecology Resources* **14**, 647–651. (doi:10.1111/1755-
541 0998.12200)
542 Dyer RJ. 2012 The gstudio package. Virginia: Virginia Commonwealth University.
543 Edgar RC. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads.
544 *Nature Methods* **10**, 996–998. (doi:10.1038/nmeth.2604)
545 Ehrlich PR, Raven PH. 1964 Butterflies and plants: a study in coevolution. *Evolution* **18**, 586–
546 608.
547 Ellison AM, Gotelli NJ. 2009 Energetics and the evolution of carnivorous plants—Darwin's
548 'most wonderful plants in the world'. *J. Exp. Bot.* **60**, 19–42. (doi:10.1093/jxb/ern179)

549 Espíndola A, Carstens BC, Alvarez N. 2014 Comparative phylogeography of mutualists and the
550 effect of the host on the genetic structure of its partners. *Biol. J. Linn. Soc.* **113**, 1021–1035.
551 (doi:10.1111/blj.12393)
552 Excoffier L, Smouse PE, Quattro JM. 1992 Analysis of molecular variance inferred from metric
553 distances among DNA haplotypes: application to human mitochondrial DNA restriction data.
554 *Genetics* **131**, 479–491.
555 Farrell B, Mitter C. 1990 Phylogenesis of insect/plant interactions: have *Phyllobrotica* leaf
556 beetles (Chrysomelidae) and the Lamiales diversified in parallel? *Evolution* **44**, 1389–1403.
557 Fashing NJ, OConnor BM. 1984 Sarraceniopus - a new genus for Histiostomatid mites
558 inhabiting the pitchers of the Sarraceniaceae (astigmata: Histiostomatidae). *International*
559 *Journal of Acarology* **10**, 217–227.
560 Finlay BJ. 2002 Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–
561 1063. (doi:10.1126/science.1070710)
562 Folkerts D. 1999 Pitcher plant wetlands of the southeastern United States. Pp 247–275 in Batzer
563 DP, Rader RB, Wissinger SA, eds. Invertebrates in Freshwater Wetlands of North America:
564 Ecology and Management. John Wiley and Sons, Inc., New York, NY.
565 Fouquet A, Noonan BP, Rodrigues MT, Pech N, Gilles A, Gemmell NJ. 2012 Multiple
566 Quaternary refugia in the eastern Guiana shield revealed by comparative phylogeography of
567 12 frog species. *Syst. Biol.* **61**, 461–489. (doi:10.1093/sysbio/syr130)
568 Gotelli NJ, Ellison AM. 2006 Food-web models predict species abundances in response to
569 habitat change. *PLoS Biol.* **4**, e324. (doi:10.1371/journal.pbio.0040324)
570 Hausner G, Reid J, Klassen GR. 1993 On the subdivision of *Ceratocystis* s.l., based on partial
571 ribosomal DNA sequences. *Can. J. Bot.* **71**, 52–63.

572 Hewitt G. 2000 The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.
573 (doi:10.1038/35016000)
574 Hewitt GM. 2004 Genetic consequences of climatic oscillations in the Quaternary. *Phil. Trans.*
575 *R. Soc. Lond. B* **359**, 183–195. (doi:10.1098/rstb.2003.1388)
576 Hope AG, Ho SYW, Malaney JL, Cook JA, Talbot SL. 2014 Accounting for rate variation
577 among lineages in comparative demographic analyses. *Evolution* **68**, 2689–2700.
578 (doi:10.1111/evo.12469)
579 Hunter PE, Hunter CA. 1964 A new *Anoetus* mite from pitcher plants. *Proc. Entomol. Soc.*
580 *Wash.* **66**, 39–46.
581 Jackson ND, Austin CC. 2010 The combined effects of rivers and refugia generate extreme
582 cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution* **64**,
583 409–428. (doi:10.1111/j.1558-5646.2009.00840.x)
584 Jensen JL, Bohonak AJ, Kelley ST. 2005 Isolation by distance, web service. *BMC Genetics* **6**,
585 13. v.3.23 http://ibdws.sdsu.edu/ (doi:10.1186/1471-2156-6-13)
586 Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7:
587 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
588 (doi:10.1093/molbev/mst010)
589 Kitching RL. 2000. Food webs and container habitats: the natural history and ecology of
590 phytotelmata. Cambridge University Press.
591 Koopman MM, Carstens BC. 2010 Conservation genetic inferences in the carnivorous pitcher
592 plant *Sarracenia alata* (Sarraceniaceae). *Conserv. Genet.* **11**, 2027–2038.
593 (doi:10.1007/s10592-010-0095-7)

Koopman MM, Fuselier DM, Hird S, Carstens BC. 2010 The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. *Appl. Environ. Microbiol.* **76**, 1851–1860. (doi:10.1128/AEM.02440-09)

Koopman MM, Carstens BC. 2011 The microbial phyllogeography of the carnivorous plant *Sarracenia alata*. *Microb. Ecol.* **61**, 750–758. (doi:10.1007/s00248-011-9832-9)

Koski LB, Golding GB. 2001 The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52**, 540–542. (doi:10.1007/s002390010184)

Leaché AD, Fujita MK. 2010 Bayesian species delimitation in west African forest geckos (*Hemidactylus fasciatus*). *Proc. R. Soc. B* **277**, 3071–3077. (doi:10.1098/rspb.2010.0662)

Mardis ER. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387–402. (doi:10.1146/annurev.genom.9.081307.164359)

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL. 2011 Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution* 65, 184–202. (doi:10.1111/j.1558-5646.2010.01097.x)

McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD. 2009 Temporal lags and overlap in the diversification of weevils and flowering plants. *Proc. Nat. Acad. Sci.* **106**, 7083–7088. (doi:10.1073/pnas.0810618106)

Miller TE, Kneitel JM. 2005 Inquiline communities in pitcher plants as a prototypical metacommunity. *In*: Metacommunities: spatial dynamics and ecological communities (eds Holyoak M, Leibold MA, Holt R) University of Chicago Press, Chicago, IL, pp.122.145.

Miller TE, terHorst CP. 2012 Testing successional hypotheses of stability, heterogeneity, and diversity in pitcher-plant inquiline communities. *Oecologia* **170**, 243–251. (doi:10.1007/s00442-012-2292-1)

617 Nei M. 1973 Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci.* **70**,
618 3321–3323.
619 Newell SJ, Nastase AJ. 1998 Efficiency of insect capture by *Sarracenia purpurea*
620 (Sarraceniaceae), the northern pitcher plant. *American Journal of Botany* **85**, 88–91.
621 Newman CE, Rissler LJ. 2011 Phylogeographic analyses of the southern leopard frog: the impact
622 of geography and climate on the distribution of genetic lineages vs. subspecies. *Mol. Ecol.*
623 **20**, 5295–5312. (doi:10.1111/j.1365-294X.2011.05353.x)
624 Noss RF. 1989 Longleaf pine and wiregrass: keystone components of an endangered ecosystem.
625 *Natural Areas Journal* **9**, 211–213.
626 O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. 2005 Fungal community
627 analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* **71**,
628 5544–5550. (doi:10.1128/AEM.71.9.5544-5550.2005)
629 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos
630 P, Stevens MHH, Wagner H. 2015 vegan: community ecology package. R package version
631 2.2-1. http://CRAN.R-project.org/package=vegan
632 Paradis E. 2010 pegas: an R package for population genetics with an integrated-modular
633 approach. *Bioinformatics* **26**, 419–420. (doi:10.1093/bioinformatics/btp696)
634 Peay KB, Kennedy PG, Bruns TD. 2008 Fungal community ecology: a hybrid beast with a
635 molecular master. *Bioscience* **58**, 799–810. (doi:10.1641/B580907)
636 Peterson CN, Day S, Wolfe BE, Ellison AM, Kolter R, Pringle A. 2008 A keystone predator
637 controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem.
638 *Environ. Microbiol.* **10**, 2257–2266. (doi:10.1111/j.1462-2920.2008.01648.x)

639 Porter SD, Savignano DA. 1990 Invasion of polygyne fire ants decimates native ants and
640 disrupts arthropod community. *Ecology* **71**, 2095–2106. (doi:10.2307/1938623)
641 Pyron RA, Burbrink FT. 2009 Lineage diversification in a widespread species: roles for niche
642 divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Mol. Ecol.* **18**,
643 3443–3457. (doi:10.1111/j.1365-294X.2009.04292.x)
644 R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for
645 Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
646 Rønsted N, Weiblen GD, Cook JM, Salamin N, Machado CA, Savolainen V. 2005 60 million
647 years of co-divergence in the fig-wasp symbiosis. *Proc. R. Soc. B* **272**, 2593–2599.
648 (doi:10.1098/rspb.2005.3249)
649 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
650 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber
651 CF. 2009 Introducing mothur: open-source, platform-independent, community-supported
652 software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**,
653 7537–7541. (doi:10.1128/AEM.01541-09)
654 Smith CI, Tank S, Godsoe W, Levenick J, Strand E, Esque T, Pellmyr O. 2011 Comparative
655 phylogeography of a coevolved community: concerted population expansions in Joshua trees
656 and four yucca moths. *PLoS ONE* **6**, e25628. (doi:10.1371/journal.pone.0025628)
657 Smith BT, Amei A, Klicka J. 2012 Evaluating the role of contracting and expanding rainforest in
658 initiating cycles of speciation across the Isthmus of Panama. *Proc. R. Soc. B* **279**, 3520–3526.
659 (doi:10.1098/rspb.2012.0706)

660 Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. 2006 Comparative phylogeography
661 of unglaciated eastern North America. *Mol. Ecol.* **15**, 4261–4293. (doi:10.1111/j.1365-
662 294X.2006.03061.x)
663 Sota T, Kagata H, Ando Y, Utsumi S, Osono T. 2014 Metagenomic approach yields insights into
664 fungal diversity and functioning. Species Diversity and Community Structure (pp. 1–23).
665 Springer Japan.
666 Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
667 thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.
668 (doi:10.1093/bioinformatics/btl446)
669 Stamatakis A, Hoover P, Rougemont J. 2008 A rapid bootstrap algorithm for the RAxML web
670 servers. *Syst. Biol.* **57**, 758–771. (doi:10.1080/10635150802429642)
671 Stuble KL, Kirkman LK, Carroll CR. 2009 Patterns of abundance of fire ants and native ants in a
672 native ecosystem. *Ecological Entomology* **34**, 520–526. (doi:10.1111/j.1365-
673 2311.2009.01098.x)
674 Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA
675 polymorphism. *Genetics* **123**, 585–595.
676 Taylor JT, Turner E, Townsend JP, Dettman JR, Jacobson D. 2006 Eukaryotic microbes, species
677 recognition and the geographic limits of species: examples from the kingdom fungi. *Phil.*
678 *Trans. R. Soc. B.* **361**, 1947–1963. (doi:10.1098/rstb.2006.1923)
679 terHorst CP, Miller TE, Levitan DR. 2010 Evolution of prey in ecological time reduces the effect
680 size of predators in experimental microcosms. *Ecology* **91**, 629–636. (doi:10.1890/09-
681 1481.1)

682 Tringe SG, Rubin EM. 2005 Metagenomics: DNA sequencing of environmental samples. *Nature*
683 *Reviews Genetics* **6**, 805–814. (doi:10.1038/nrg1709)
684 Weiblen GD, Bush GL. 2002 Speciation in fig pollinators and parasites. *Mol. Ecol.* **11**, 1573–
685 1578. (doi:10.1046/j.1365-294X.2002.01529.x)
686 Weisrock DW, Janzen FJ. 2000 Comparative molecular phylogeography of North American
687 softshell turtles (*Apalone*): implications for regional and wide-scale historical evolutionary
688 forces. *Mol. Phyl. Evol.* **14**, 152–164. (doi:10.1006/mpev.1999.0689)
689 Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007 The genetic
690 basis of a plant-insect coevolutionary key innovation. *Proc. Nat. Acad. Sci.* **104**, 20427–
691 20431. (doi:10.1073/pnas.0706229104)
692 Zellmer AJ, Hanes MM, Hird SM, Carstens BC. 2012 Deep phylogeographic structure and
693 environmental differentiation in the carnivorous plant *Sarracenia alata*. *Syst. Biol.* **61**, 763–
694 777. (doi:10.1093/sysbio/sys048)
695 Zimmerman NB, Vitousek PM. 2012 Fungal endophyte communities reflect environmental
696 structuring across a Hawaiian landscape. *Proc. Nat. Acad. Sci.* **109**, 13022–13027.
697 (doi:10.1073/pnas.1209872109)
698
699
700
701
702
703
704

Figure Legends

Figure 1. Sampling distribution of *Sarracenia alata* in Louisiana. Sample sites are partitioned based on side of the Mississippi River. Red circles represent Kisatchie (K) and Cooter's Bog (C) in the west; blue squares represent Lake Ramsey (L), Abita Springs (A) and Talisheek (T) in the east.

Figure 2. Taxonomic composition of the OTUs for each sample site. Each site contains the number of OTUs (*N*) and the major lineages in which they belong. See Supplemental Material for full taxonomic information.

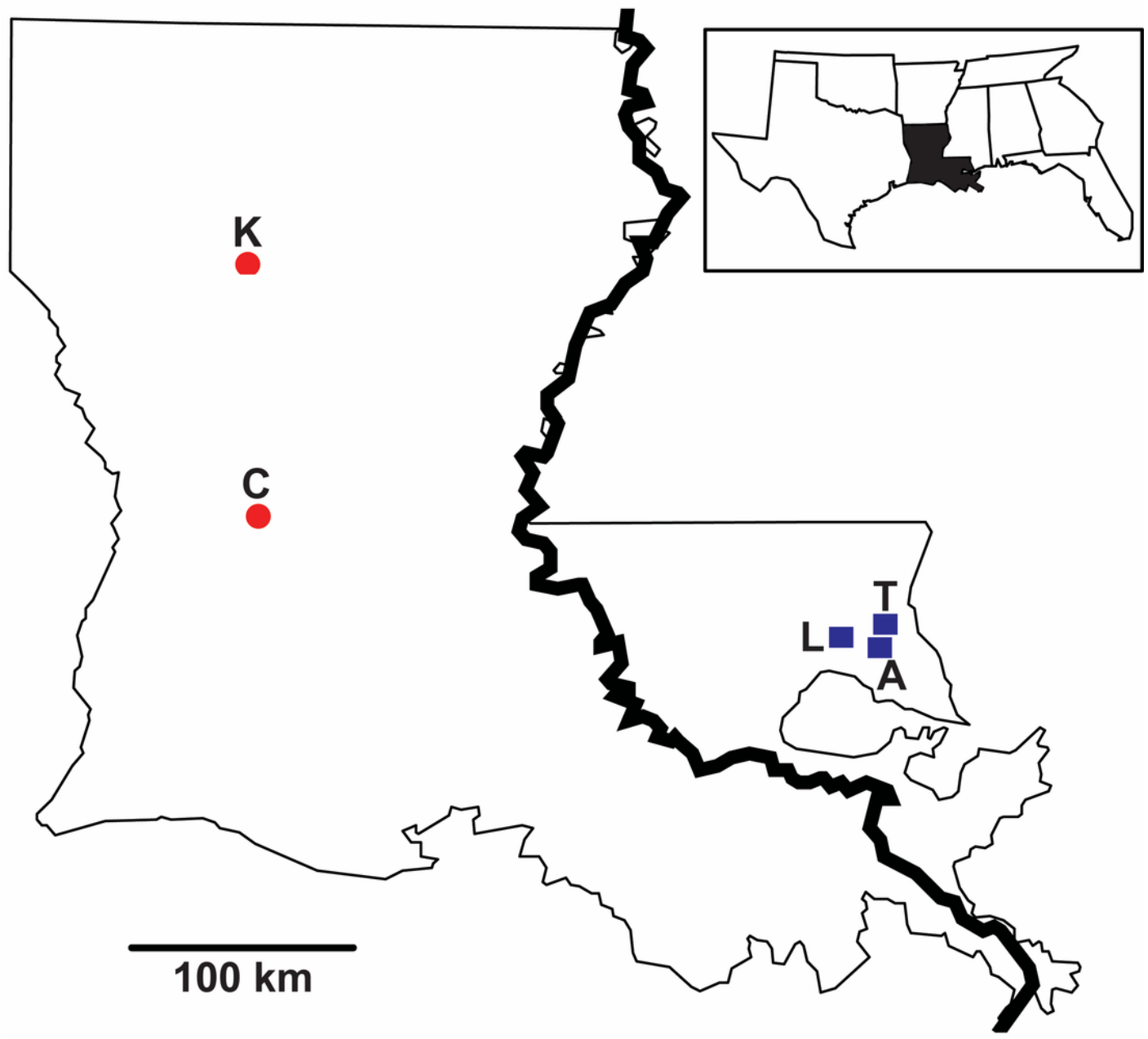
Figure 3. Rarefaction curves of OTU richness at each sampling site.

Figure 4. Population genetic structure for the inquiline community spanning the Mississippi River. Results are shown from the AMOVA and GSI analyses. AMOVA analyses show the hierarchical partitioning scheme of locales within regions ($\mathbf{\Phi_{SC}}$), locales within total distribution ($\mathbf{\Phi_{ST}}$), and between regions ($\mathbf{\Phi_{CT}}$). GSI analyses represent the amount of allelic sorting on the eastern and western sides of the Mississippi River. Dark cells indicate taxa with significant genetic structure at the corresponding level; Table S2 contains specific values from each analysis. See Carstens and Satler (2013) for sampling information for *S. alata*, as these samples were collected from throughout the plant's distribution.

1

Sampling distribution of *Sarracenia alata* in Louisiana.

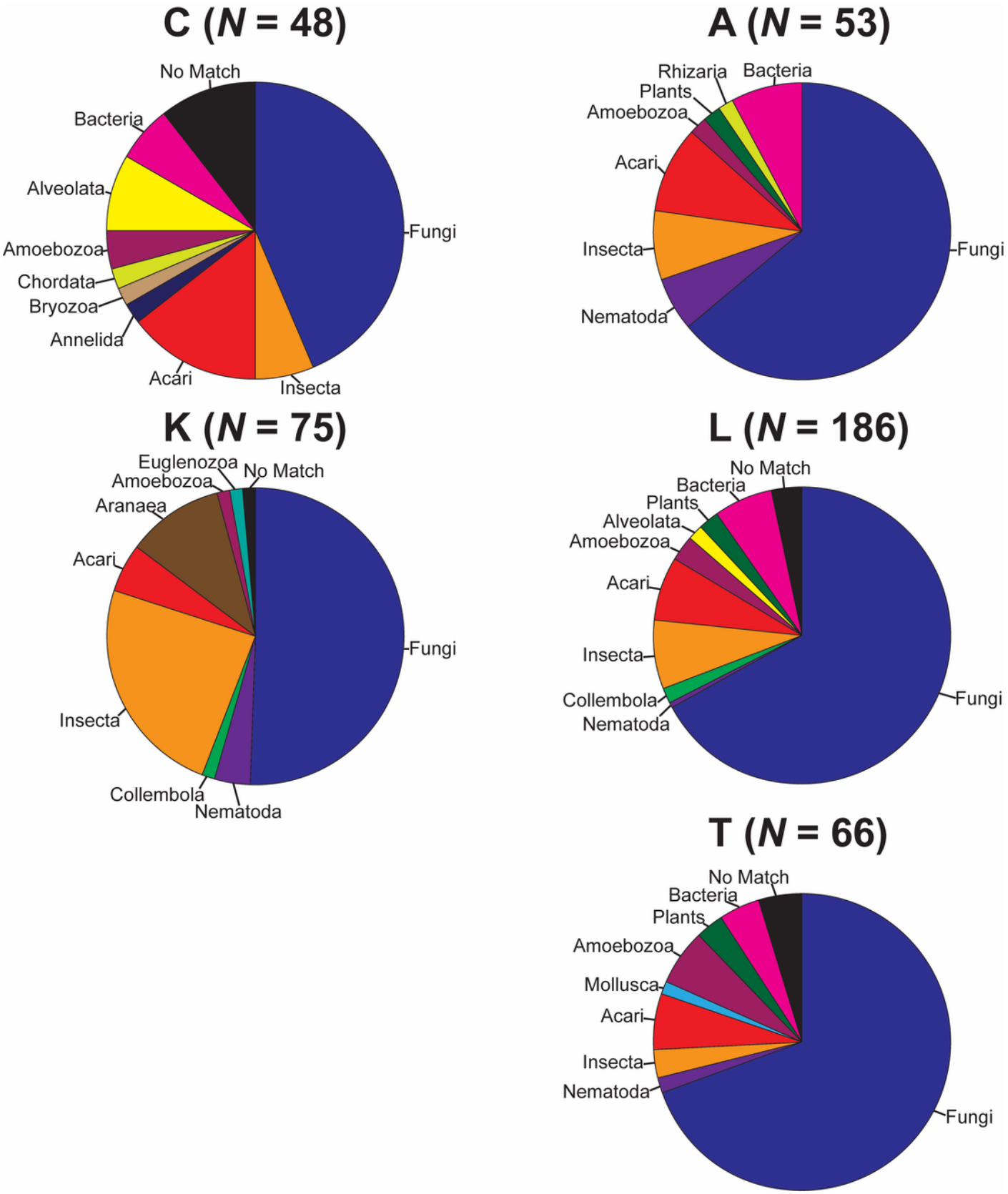
Sample sites are partitioned based on side of the Mississippi River. Red circles represent Kisatchie (K) and Cooter's Bog (C) in the west; blue squares represent Lake Ramsey (L), Abita Springs (A) and Talisheek (T) in the east.



2

Taxonomic composition of the OTUs for each sample site.

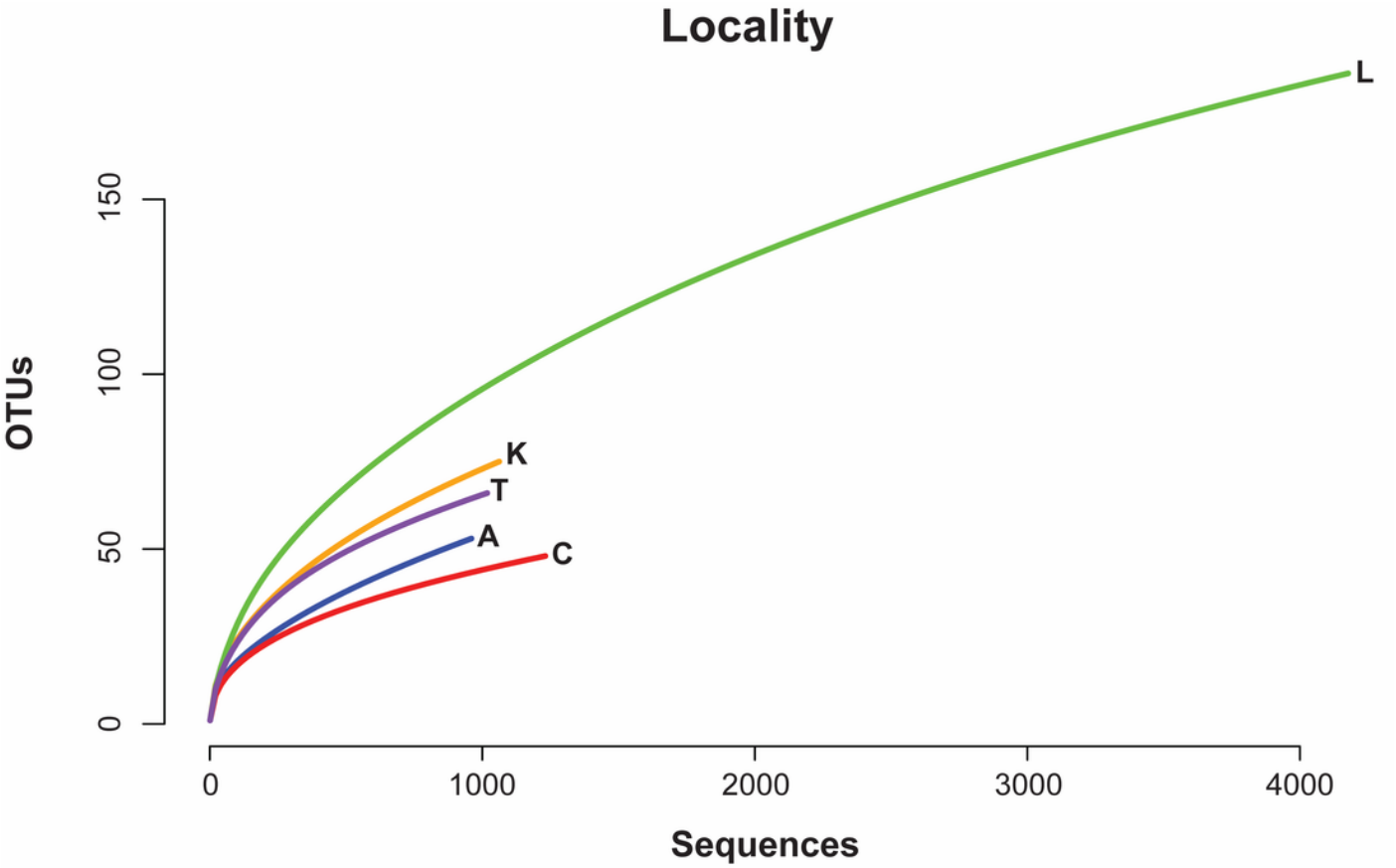
Each site contains the number of OTUs (*N*) and the major lineages in which they belong. See Supplemental Materials for full sampling information.


C (N = 48)
No Match
Bacteria
Alveolata
Amoebozoa
Chordata
Bryozoa
Annelida
Acari
Insecta
Fungi
A (N = 53)
Rhizaria
Bacteria
Plants
Amoebozoa
Acari
Insecta
Nematoda
Fungi
K (N = 75)
Euglenozoa
Amoebozoa
No Match
Aranaea
Acari
Insecta
Collembola
Nematoda
Fungi
L (N = 186)
No Match
Bacteria
Plants
Alveolata
Amoebozoa
Acari
Insecta
Collembola
Nematoda
Fungi
T (N = 66)
No Match
Bacteria
Plants
Amoebozoa
Mollusca
Acari
Insecta
Nematoda
Fungi

3

Rarefaction curves of OTU richness at each sampling site.

Rarefaction curves of OTU richness at each sampling site are shown with the number of OTUs plotted against the number of sequences.



4

Population genetic structure for the inquiline community spanning the Mississippi River.

Results are shown from the AMOVA and GSI analyses. AMOVA analyses show the hierarchical partitioning scheme of locales within regions (Φ_{SC}), locales within total distribution (Φ_{ST}), and between regions (Φ_{CT}). GSI analyses represent the amount of allelic sorting on the eastern and western sides of the Mississippi River. Dark cells indicate taxa with significant genetic structure at the corresponding level. See Table S2 for specific values from each analysis.

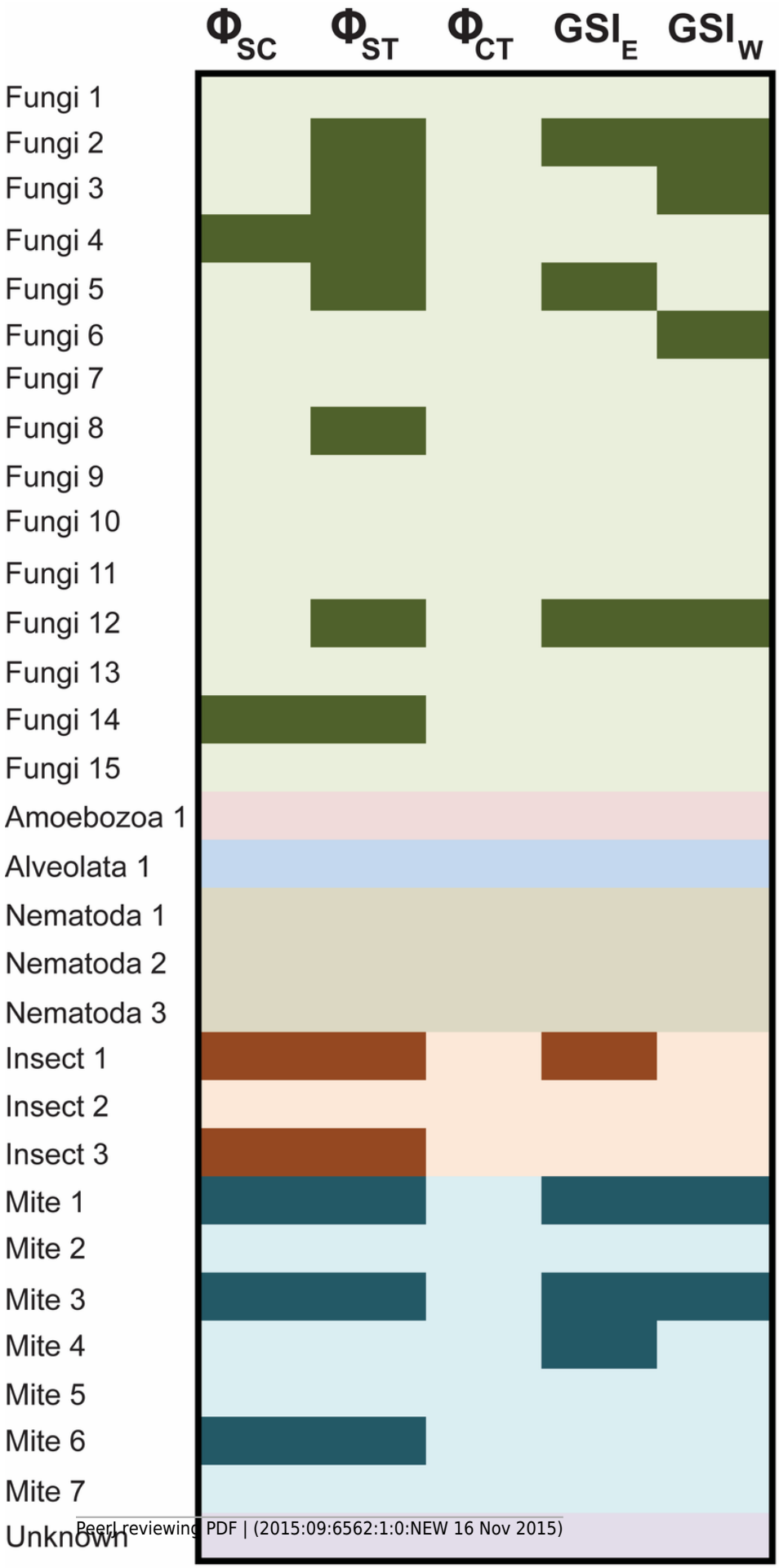

Φ_{SC}
Φ_{ST}
Φ_{CT}
GSI_E
GSI_W
Fungi 1
Fungi 2
Fungi 3
Fungi 4
Fungi 5
Fungi 6
Fungi 7
Fungi 8
Fungi 9
Fungi 10
Fungi 11
Fungi 12
Fungi 13
Fungi 14
Fungi 15
Amoebozoa 1
Alveolata 1
Nematoda 1
Nematoda 2
Nematoda 3
Insect 1
Insect 2
Insect 3
Mite 1
Mite 2
Mite 3
Mite 4
Mite 5
Mite 6
Mite 7
Unknown

Table 1(on next page)

Table 1

Taxa included in the final comparative data set. Information for OTUs include number of sequences (N), their nearest BLAST hit (except for *S. alata*), nucleotide diversity (π), Watterson's theta (Θ_w) per site, Tajima's *D*, and $G_{ST.}$ Significance of G_{ST} and Tajima's *D* (*D* following a beta distribution; Tajima, 1989) at $\alpha = 0.05$ is indicated with an asterisk (*).

Table 1. Taxa included in the final comparative data set. Information for OTUs include number of sequences (N), their nearest BLAST hit (except for *S. alata*), nucleotide diversity (π), Watterson's theta (Θ_w) per site, Tajima's *D*, and $G_{ST.}$ Significance of G_{ST} and Tajima's *D* (*D* following a beta distribution; Tajima, 1989) at $\alpha = 0.05$ is indicated with an asterisk (*).

Taxa	N	BLAST	π	Θ_w	Tajima's *D*	G_{ST}
Fungi1	52	*Cladosporium sp.* (Fungi)	0.0088	0.0244	-2.1585*	0.1099*
Fungi2	51	*Fusarium annulatum* (Fungi)	0.0130	0.0299	-2.0211*	0.2335
Fungi3	22	*Curvularia sp.* (Fungi)	0.0157	0.0272	-1.7148	0.1419
Fungi4	2507	*Candida saitoana* (Fungi)	0.0059	0.0918	-2.6902*	0.0025
Fungi5	97	*Candida saitoana* (Fungi)	0.0072	0.0203	-2.0520*	0.0514
Fungi6	168	*Candida saitoana* (Fungi)	0.0090	0.0187	-1.6674	0.0167
Fungi7	84	*Candida saitoana* (Fungi)	0.0088	0.0217	-2.0016*	0.0711
Fungi8	57	*Candida quercitrusa* (Fungi)	0.0134	0.0341	-2.0901*	0.1150
Fungi9	30	*Candida saitoana* (Fungi)	0.0010	0.0053	-2.3512*	0.0991
Fungi10	54	*Candida saitoana* (Fungi)	0.0050	0.0183	-2.3958*	0.0788
Fungi11	189	*Mucor circinelloides* (Fungi)	0.0078	0.0307	-2.3694*	0.0090
Fungi12	14	Uncultured soil fungus (Fungi)	0.0087	0.0100	-0.6488	0.2805
Fungi13	40	Uncultured fungus (Fungi)	0.0103	0.0125	-0.6810	0.2494
Fungi14	766	Fungal endophyte (Fungi)	0.0117	0.0586	-2.4374*	0.0052
Fungi15	15	*Nigrospora sphaerica* (Fungi)	0.0106	0.0188	-1.8028*	0.0451
Amoebozoa1	227	*Fuligo septica* (Amoebozoa)	0.0042	0.0347	-2.6856*	0.0411
Alveolata1	18	*Leptopharynx costatus* (Alveolata)	0.0182	0.0324	-1.8094*	0.0418
Nematoda1	79	*Nematoda sp.* (Nematoda)	0.0150	0.0317	-1.8654*	0.4101
Nematoda2	154	*Nematoda sp.* (Nematoda)	0.0130	0.0174	-0.9456	0.2359
Nematoda3	21	*Nematoda sp.* (Nematoda)	0.0188	0.0170	0.2858	0.0624
Insect1	61	*Brachymyrmex depilis* (Insecta)	0.0122	0.0285	-2.0101*	0.0346
Insect2	37	*Solenopsis xyloni* (Insecta)	0.0208	0.0409	-1.8296*	0.0723

Insect3	41	*Paratrechina hystrix* (Insecta)	0.0081	0.0208	-2.1225*	0.2694
Mite1	828	*Ovanoetus sp.* (Acari)	0.0086	0.0620	-2.5727*	0.0150*
Mite2	30	*Ovanoetus sp.* (Acari)	0.0152	0.0348	-2.1831*	0.2911
Mite3	1071	*Anoetus sp.* (Acari)	0.0071	0.0678	-2.6276*	0.0101*
Mite4	56	*Anoetus sp.* (Acari)	0.0114	0.0242	-1.7765	0.0437
Mite5	34	*Anoetus sp*. (Acari)	0.0176	0.0219	-0.7551	0.1516
Mite6	50	*Anoetus sp.* (Acari)	0.0111	0.0197	-1.4951	0.0427
Mite7	45	*Anoetus sp.* (Acari)	0.0049	0.0098	-1.5594	0.0147
Unknown	66	No BLAST Match	0.0059	0.0198	-2.2971*	0.3539
Host plant	79	*Sarracenia alata*	0.0028	0.0034	-0.4521	0.8483*

5

6

Table 2(on next page)

Table 2

A chi-squared goodness of fit test was used to measure if the number of taxa with significant population genetic structure was more than would be expected by chance alone. Under a null model we would expect a significant result 5% of the time (assuming $\alpha = 0.05$). Results show that for many analyses, there are more OTUs with significant values than expected by chance, suggesting an association between many members of the community and the host pitcher plant.

Table 2. A chi-squared goodness of fit test was used to measure if the number of taxa with significant population genetic structure was more than would be expected by chance alone. Under a null model we would expect a significant result 5% of the time (assuming $\alpha = 0.05$). Results show that for many analyses, there are more OTUs with significant values than expected by chance, suggesting an association between many members of the community and the host pitcher plant.

Test	χ^2	df	p-value	Number Significant	Total Taxa
Φ_{SC}	22.3612	1	2.26×10^{-6}	7	29
Φ_{ST}	80.8004	1	2.20×10^{-16}	12	29
Φ_{CT}	1.5263	1	0.22	0	29
G_{ST}	1.4278	1	0.23	3	31
GSI_E	20.1715	1	7.08×10^{-6}	7	31
GSI_W	13.4482	1	2.45×10^{-4}	6	31