

MITOCHONDRIAL DNA VARIATION IN THE PITCHER PLANT FLY *SARCOPHAGA SARRACENIAE*: EXPLORING POSSIBLE INFLUENCES OF HOST SPECIFICITY AND GEOGRAPHIC STRUCTURING

by

Joshua Parker

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Joshua Parker

Greenville, NC

May, 2019

Approved by:

Dr. Trip Lamb

Department of Biology, East Carolina University

I hereby declare I am the sole author of this thesis. It is the result of my own work and is not the outcome of work done in collaboration, nor has any of it been submitted elsewhere for another degree.

Signed: Joshua Parker

Date: 5/2/2019

Mitochondrial DNA variation in the pitcher plant fly *Sarcophaga sarraceniae*: Exploring possible influences of host specificity and geographic structuring

Joshua D. Parker

Department of Biology, East Carolina University, Greenville N.C. 27858, U.S.A.

ABSTRACT - North American pitcher plants (*Sarracenia*) are a recently evolved (~3 my) assemblage of 11 carnivorous species. *Sarracenia* pitchers also provide resources (food, shelter) for certain arthropods, including two genera of flesh flies (family *Sarcophagidae*) whose larvae develop within pitchers: *Fletcherimyia*, and a second, single species in the genus *Sarcophaga* (*S. sarraceniae*). *Sarcophaga sarraceniae* inhabits the entire geographic range of *Sarracenia* and appears to deposit larvae indiscriminately among various pitcher plant species whereas *Fletcherimyia* occupies smaller species ranges and shows pitcher host specificity. Is *S. sarraceniae* truly a pitcher generalist? To address this question, I examined mitochondrial DNA (mtDNA) variation in *S. sarraceniae* to test two hypotheses: 1) co-evolution—where observed mtDNA variation should be attributable to plant host fidelity, and 2) geography—where mtDNA variation should exhibit phylogeographic structure. I secured sequence data for the mitochondrial cytochrome oxidase 1 gene for 29 specimens representing 19 populations across the species range. Genetic variation was quite limited; it provided no support for the coevolution hypothesis but did show some phylogeographic structuring. These findings suggest that the symbiotic relationship between *S. sarraceniae* and *Sarracenia* may have been established fairly recently.

Acknowledgments

This thesis would not have been possible without the help of the following people and organizations. I would like to thank Dr. Michael Brewer and the University Department of Biology at East Carolina University. Lastly, I would like to give a special thanks to Dr. Trip Lamb, my research mentor and thesis advisor, for his knowledge, patience, and support. This project was funded in part by the East Carolina Department of Biology.

Table of Contents

Introduction	8
<i>Fly – pitcher symbiosis</i>	
<i>Questions and Hypotheses</i>	
Materials and Methods	10
<i>Collection Procedures</i>	
<i>DNA Extraction and Sequencing</i>	
<i>Sequence Alignment and Analysis</i>	
Results	14
DNA Sequencing and its Challenges	
Sequence Alignment	
Neighbor Joining and Haplotype Network Results	
Phylogeographic Map Results	
Discussion	19
Major Findings	
Vicariance Effects on Haplotypes	
Future Work	20
Continuing studies	
Literature cited	21

List of Tables

Table 1. Forward and reverse primers used for *cox1* gene amplification; primer sequences are written 5'–3'.

Table 2. Identification numbers and collection localities for the specimens used in this study.

List of Figures

Figure 1. Minimum spanning network for the *cox I* haplotypes, color-coded to depict the three recovered haplotype groups. Dashes indicate single nucleotide differences between connected groups.

Figure 2. Neighbor joining tree based on *cox I* sequenced samples, color-coded to depict the three haplotype groups.

Figure 3. Map of collecting localities, coded by haplotype color designations in the minimum spanning network.

Introduction

Fly – pitcher symbiosis

The North American pitcher plants are a recently evolved (~3 million yrs) assemblage of 11 carnivorous species (Stephens et al., 2015). Their tubular leaves (pitchers) are pitfall traps that lure, catch, and digest insect prey. Despite this functional role, pitchers alternatively provide resources (food, shelter) for some insects, including certain species that require the plants to complete their life cycle (Folkerts, 1999). Among these obligate symbionts are nine species of flesh fly (family Sarcophagidae) that develop inside the pitcher and consume its prey. One species, *Sarcophaga sarraceniae*, uses all 11 pitcher plant species, whereas the other eight species (genus *Fletcherimyia*) target particular pitcher species as their host (Dahlem, 2006). One explanation is *Fletcherimyia* flies have been pitcher symbionts for a longer period of time, allowing them to co-evolve with specific pitcher plant hosts. Another possibility is that *S. sarraceniae*—supposedly one species—may comprise several distinct genetic lineages that are morphologically indistinguishable.

Questions and Hypotheses

To better understand the role of *S. sarraceniae* as a pitcher plant associate, I examined population genetic variation throughout the fly's geographic range and among pitcher plant hosts. Limited genetic variation would suggest the fly-pitcher plant association may have been established recently. Conversely, extensive genetic variation would indicate a longer symbiotic history, and that *S. sarraceniae* may represent several distinct lineages. If extensive genetic structuring is detected, I will test two hypotheses that could account for the genetic variation.

Co-evolution hypothesis: Genetic structuring has been established by fly fidelity to particular pitcher plant species.

Co-evolution is frequently seen in systems where two lineages are linked through a set of shared life histories, habitats, and even symbiotic hosts (Cruaud & Rasplus, 2016). In some cases, the cospeciation that occurs may lead to the development of multiple cryptic species that use similar resources and by standard field techniques are identified as a single species. In this case, these species can most easily be differentiated through genetic sequencing techniques, like those used in this study.

Geographic hypothesis: Genetic structuring has been established by geologic or historical climatic barriers that have limited fly dispersal.

Regarding geographic structuring of fly lineages, five well known geography influenced divergence patterns are known (Avice et al., 1987). One of these patterns is the long-term zoogeographic barrier discontinuous pattern, where large genetic breaks occur based on separation of related groups by geographic barriers. In response to the vicariant separation, gene flow between populations becomes restricted and the two lineages, though close geographically, will diverge due to the effects of genetic drift.

Both hypotheses could contribute to genetic divergence. To test the co-evolution hypothesis, I will construct a phylogenetic tree for fly genotypes and compare the fly tree to a recently published phylogenetic tree for the pitcher plants. If genetic structuring has been influenced primarily by pitcher plant co-evolution, then I would predict fly lineages should correspond closely to particular pitcher plant species. If geological/historical climatic events influenced genetic structuring, then I would predict the fly phylogenetic tree should show geographic patterning consistent with established biogeographic areas. With the data collected by this study, along with its respective conclusions, a greater understanding of the history behind this species' lack of pitcher plant fidelity and biogeographic structuring will be achieved.

Materials and Methods

Collection Procedures

I obtained (or secured from other biologists) 95 specimens of *S. sarraceniae*, representing 20 populations across the species' range (Newfoundland to Texas). In the field, flies were captured around pitcher plants using a hand net. Captured flies were humanely euthanized by exposure to acetone within the confines of a ceramic-lined kill jar. They were then preserved in Invitrogen RNAlater Stabilization Solution to minimize the degradation of DNA.

DNA Extraction and Sequencing

To assess genetic variation, the mitochondrial gene cytochrome oxidase I (*coxI*) was sequenced for samples from each locality. *cox I* is one of the more rapidly evolving genes of the mitochondrial genome, which should allow for identification of single nucleotide polymorphisms (SNP's), reflecting population genetic variation across the range of *S. sarraceniae*. I then developed two primers, S-1460 and S-3014, for amplifying and sequencing of the *cox I*, which represent modification of Simon et al.'s (1994) primers 1460 and 3014. Prior to doing so, I demonstrated that they work well for both sequencing and PCR by using them to amplify and sequence DNA from previously collected samples. Each primer pair works in tandem to sequence a specific portion of the *coxI* gene from opposite directions, providing a more accurate reading of the DNA within the sample.

Table 1. Forward and reverse primers used for *coxI* gene amplification; primer sequences are written 5’–3’.

Gene Region	Forward Primer	Reverse Primer
<i>cox I</i>	S_1460 gat tta cag tct att gcc taa att tc	S_3014 gct taa atc cat tgc act aat ctg

Fly DNA was extracted, using Qiagen’s DNeasy Blood and Tissue Kit. After the extraction process, each DNA sample was amplified with by means of the Polymerase Chain Reaction (PCR). The amplified *coxI* fragment was then analyzed through gel electrophoresis to assess the size of the amplified DNA fragments from each sample. With the newly amplified DNA samples, we used a microvolume spectrophotometer, also known as a “nanodrop” to determine the double stranded DNA concentration within each sample. Gene sequencing is typically conducted at DNA concentrations of 15–40 nanograms per microliter; thus, I used the concentration readings provided to dilute samples accordingly. Once the dilutions were completed, I submitted our samples to the ECU Genomics Core for sequencing.

Sequence Alignment and Network Analysis

Sequence data was analyzed for population genetic variation, divergence, and phylogeographic structure. To do so, the dataset was initially aligned, using the *Clustal X* software program (Larkin et al., 2007). A minimum spanning network was generated using PopART (Population Analysis with Reticulate Trees; <http://popart.otago.ac.nz/index.shtml>) to graphically display relationships among haplotypes.

Table 2. Identification numbers and collection localities for the specimens used in this study.

ID #	Locality Name	County, State	locality	date
ACL 004	Eller Seep	Clay Co., NC	Gilbert's Way off Hwy 175; left (N) side of road	7-VIII-2017
ACL 012	Eller Seep	Clay Co., NC	Gilbert's Way off Hwy 175; left (N) side of road	7-VIII-2017
ACL 179	Eglin AFB	Santa Rosa, FL	Weaver Creek	13-VII-2018
ACL 018	Chatom	Washington Co., AL	4.2 mi. S. Chatom, HY17; right side of road heading south, seepage ditch	9-VIII-2017
ACL 179	Eglin AFB	Santa Rosa, FL	Weaver Creek	13-VII-2018
TM 022-27	Shaken Creek	Pender Co., NC	Williams Rd on right off NC HHY 50 from Maple Hill	13-VII-2017
TM 076	McClure bog	Henderson Co., NC	942 Etowah School Rd	26-VII-2017
KS 1	Kisatchie National Forest	Louisiana		
KS 35	Kisatchie National Forest	Louisiana		
RD 2	Red Dirt National Wildlife Management Preserve	Louisiana		
RD 18	Red Dirt National Wildlife Management Preserve	Louisiana		
PT 11	(Pitcher) Big Thicket National Preserve	Texas		

PT 12	(Pitcher) Big Thicket National Preserve	Texas		
CB 1	Cooter's Bog	Louisiana		
BL 1	Bouton Lake	Texas		
BL 35	Bouton Lake	Texas		
KR 727649	Benjies Lake Trail	Cape Breton Highlands, Nova Scotia	46.746 N 60.817 W	29-Jun-2013
MF 856418	Bog Trail	Kouchibouguac National Park, New Brunswick	46.812 N 64.951 W	06-Jun-2013
MF 858620	Berry Hill Campground	Gros Morne National Park, Newfoundland and Labrador	49.626 N 57.922 W	11-Jul-2013
MF 859790	Berry Hill Campground	Gros Morne National Park, Newfoundland and Labrador	49.626 N 57.922 W	20-Jul-2013
HM 412879	Division No. 7	Glovertown, Newfoundland and Labrador	48.5429 N 53.9776 W	11-Jul-2009
AS 184	Abita Springs	Louisiana		
DS 1	De Soto National Forest	Mississippi		
DS 17	De Soto National Forest	Mississippi		
DS 35	De Soto National Forest	Mississippi		
TB 30	Talisheek	Louisiana		

Results

DNA Sequencing

I submitted amplified DNA to the ECU Genomics Core, but many of the samples did not sequence properly and thus did not yield useable sequence data. Following additional rounds of amplifications and dilutions, DNA samples were again submitted to the ECU Genomics core but yielded more unusable results. Only 10 of our 95 samples were incorporated in the dataset.

Sequence alignment

In choosing additional sequence samples to include in our *coxI* dataset, I selected sequences for *S. sarraceniae* from the Canadian National Parks Data Release on NCBI's GenBank service (Hebert et al., 2016). Along with the Canadian samples, I included 67 GenBank sequence samples for flies from Texas, Louisiana, and Mississippi, reported in Satler and Carstens (2016). Lastly, the readable samples we collected from North Carolina, Alabama, and the panhandle of Florida filled in the remainder of our localities. For sequence alignment, a total of 29 samples with clean reads were selected, representing 20 distinct localities. Once aligned, these samples displayed a 324-base segment with complete alignment between all samples. Within this area, 11 regions of nucleotide incongruity were found, thus separating 11 different haplotypes among the 29 samples.

Neighbor Joining and Haplotype Network Results

From the sequence alignment results, very little genetic difference between populations was found. Though 11 distinct haplotypes were present, none differed from any other haplotype by more than three bases. Comprised of the 11 haplotype groups, three large groups were found that shared high similarity with one another, but not with those of the other two groups. The groups were labeled Western, Northeastern, and Southeastern, to account for their general

geographic range within the total species range. As seen in Figure 1, the haplotype network generated from these findings displays minimal nucleotide differences between each haplotype, which is identified as a single dash mark in the connective lines between circular groups.

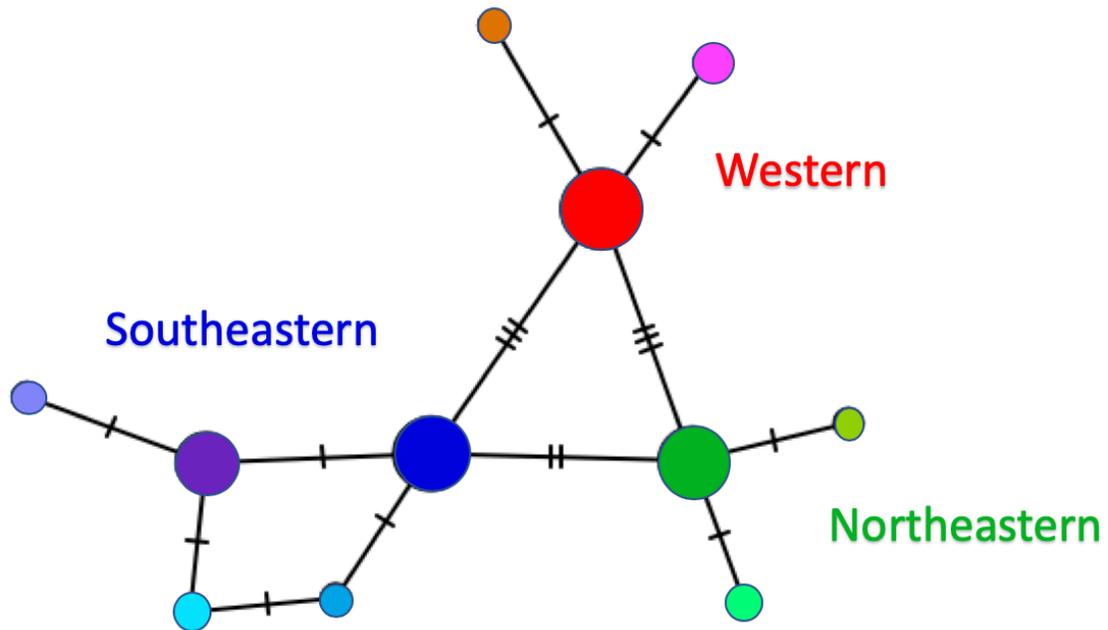


Figure 1. Minimum spanning network for the *coxI* haplotypes, color-coded to depict the three recovered haplotype groups. Dashes indicate single nucleotide differences between connected groups.

The neighbor joining tree in Figure 2 represents the same data as the haplotype network in Figure 1 but is displayed as a tree format, thus depicting relatedness through branch arrangement. Within the neighbor joining tree, the same color-coding was used to represent the major haplotypes making up the three groups.

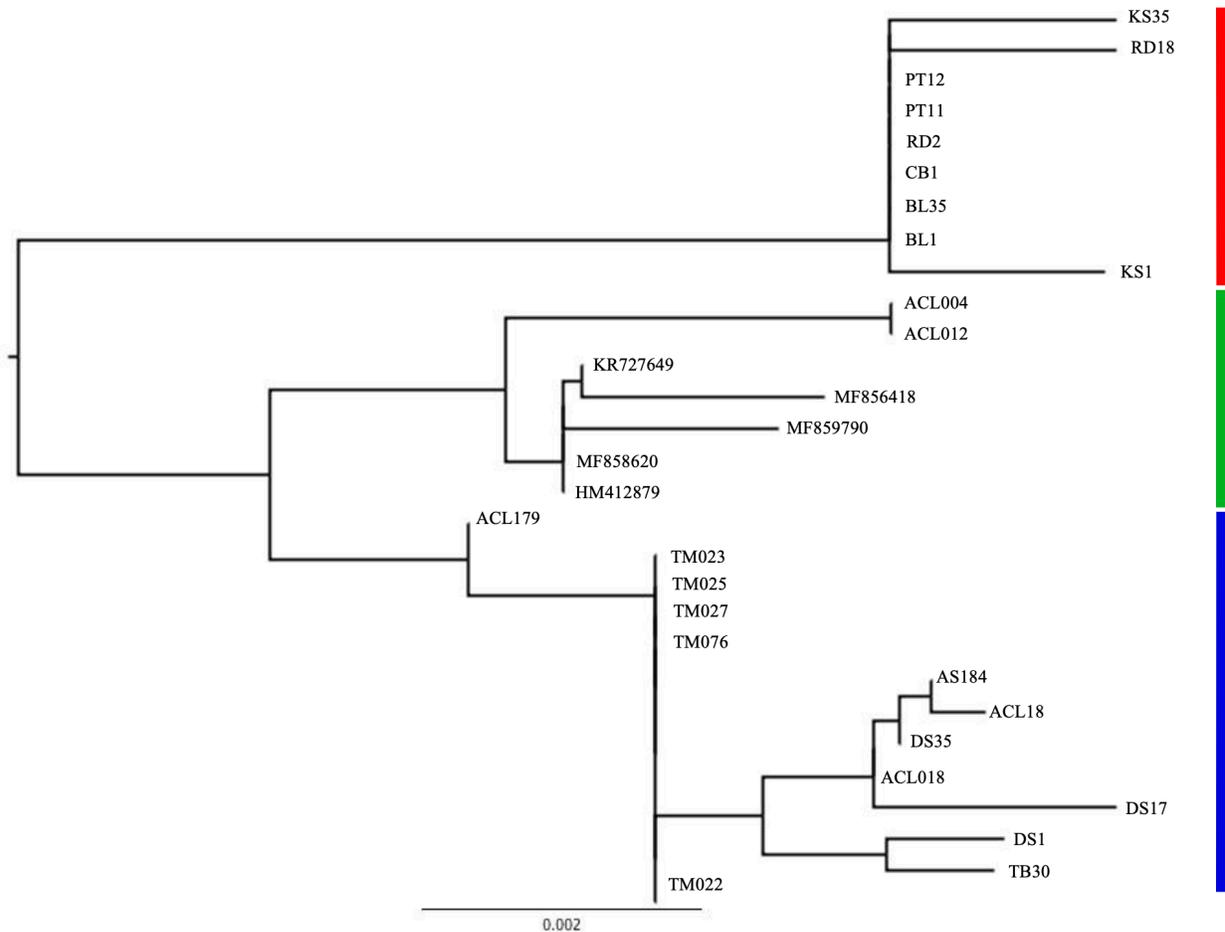


Figure 2. Neighbor joining tree based on *coxI* sequenced samples, color-coded to depict the three haplotype groups.

Phylogeographic Map Results

The three major phylogeographic groups are labeled as the Northeastern, Western, and Southeastern groups. Populations that fall in the Southeastern group were found from eastern and central North Carolina to the Florida panhandle and western Mississippi. Populations that fall in the Western group occur from east Texas to eastern Louisiana. Populations forming the Northeastern group occur from extreme western North Carolina to Newfoundland, Canada. When developing the phylogeographic map based on the locality data and haplotype network, several trends appeared. Many of the localities within a close geographic area typically were

found to have the same or a very similar haplotype. This was not the case though in the localities of central and western North Carolina, which displayed localities that possessed two nucleotide differences, though they were not greatly separated by distance. Conversely, a locality from the Florida panhandle and another from eastern Louisiana were found to have the same haplotype, even though they were separated by more than twice the physical distance as that of the two North Carolina localities. More interestingly, the western North Carolina locality, so different from the central North Carolina haplotype, clusters with the Canadian haplotypes, which suggests that the Northeastern group may be distributed along the length of the Appalachian Mountains.

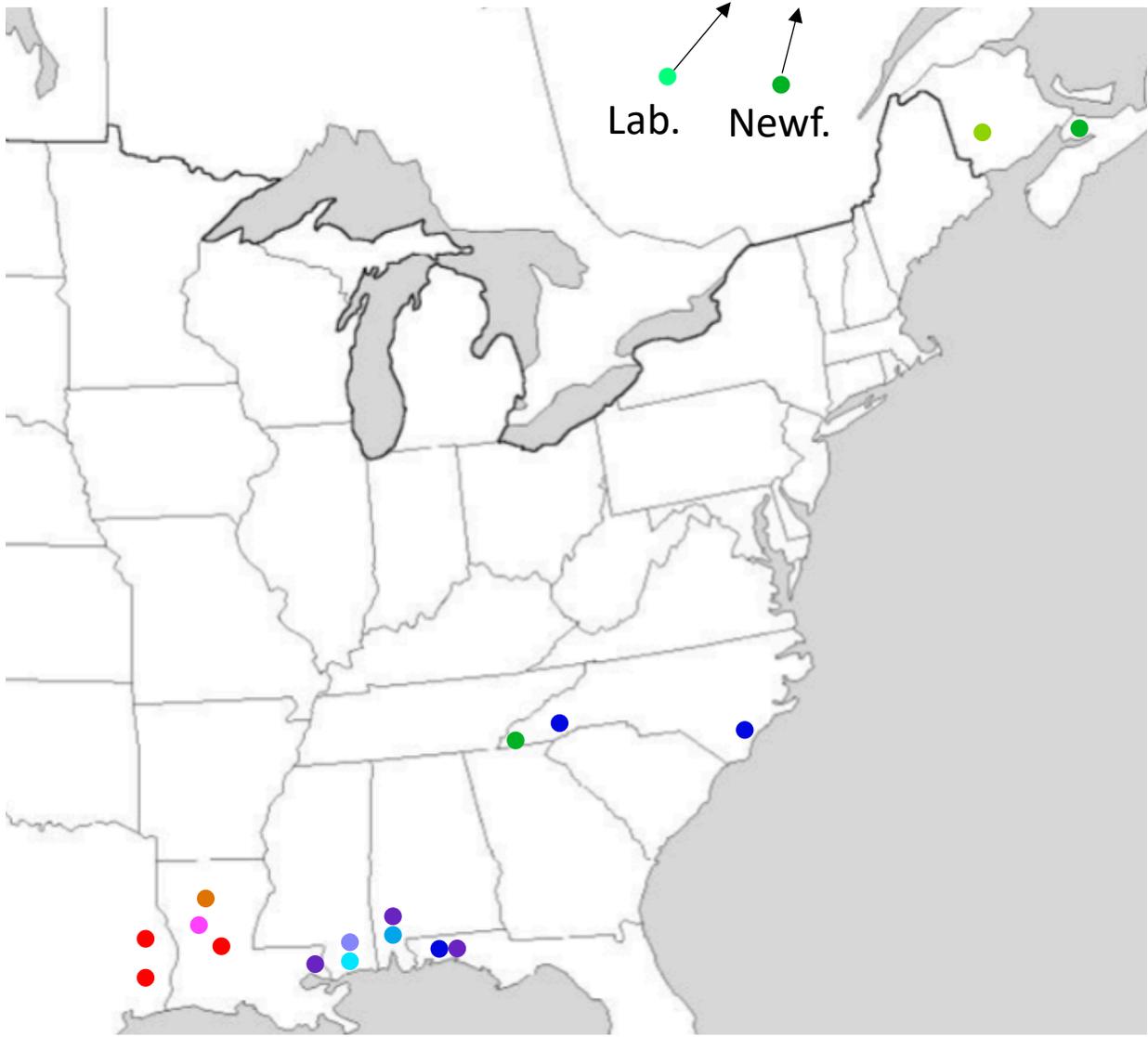


Figure 3. Map of collecting localities, coded by haplotype color designations in the minimum spanning network.

Discussion

Broad Findings

The findings of this study suggest that *S. sarraceniae* is a single species, which in the sampled section of the gene *coxI*, has very few polymorphic differences. One possible explanation for minimal genetic divergence among populations across its range may be high vagility, leading to increased levels of gene flow between populations and minimizing the effects of genetic drift. Another plausible cause includes the recent expansion of this species and its symbiotic counterparts' recent expansion, following the glacial recession of the last ice age (Merz et al., 2013).

Vicariance Effects on Haplotypes

The haplotype network revealed three phylogeographic groups. When overlaid on the map of localities and color coded based on the network, the haplotypes that most resembled each other occurred within a close geographic proximity, with the exclusion of localities within western and central North Carolina. From the comparison of our haplotype network and phylogeographic map, we found evidence for the cause of the separation of the species range into three discrete regions. The separations of these regions in some cases appeared quite abrupt, which led us to look for vicariant factors that might minimize gene flow between populations. I found that between the western and southeastern regions, the Mississippi River split the sampled populations and likely presented itself as a difficult barrier for flies to cross. For the northeastern and southeastern regions, overall, there are not enough sampled points to lead to a firm conclusion as to what separates these populations. The exception to this is the oddity of the western and central North Carolina localities, which are genetically different, but are not geographically distant.

Future Work

Continuing Studies

Although 95 samples were collected, less than a quarter of them were used for sequence alignment. To better support our results, the next steps will include extracting DNA from the remaining samples for additional sequencing. Along with this, new localities with pitcher plants north and west of North Carolina should be sampled to increase the robustness of the sampled gene pool and fill geographic gaps in my study. This will also elucidate the possible presence of other discrete regional groups and/or display more geographic structure around physical barriers.

To further this study, the inclusion of sequencing for the mitochondrial gene *cox II*, as well as possible microsatellites throughout the genome would provide stronger results to support or refute the conclusions based off of the *cox I* gene alone.

Literature Cited

- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., . . . Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18(1), 489-522.
- Cruaud, A., & Rasplus, J. Y. (2016). Testing cospeciation through large-scale cophylogenetic studies. *Current Opinion in Insect Science*, 18, 53-59.
- Dahlem, G.A., Naczi, R.C. (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.
- Folkerts, D. (1999) Pitcher plant wetlands of the southeastern United States. In Batzer, D.P., Rader, R. B., Wissinger, S. A. (Eds.), *Invertebrates in Freshwater Wetlands of North America: Ecology and Management*. pp 247–275.
- Hebert, P. D. N., Ratnasingham, S., Zakharov, E. V., Telfer, A. C., Levesque-Beaudin, V., Milton, M. A., . . . Dewaard, J. R. (2016). Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702), 20150333-20150333.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). ClustalW and ClustalX version 2.0. *Bioinformatics*, 23, 2947–2948.
- Merz, C., Catchen, J. M., Hanson-Smith, V., Emerson, K. J., Bradshaw, W. E., & Holzapfel, C. M. (2013). Replicate phylogenies and post-glacial range expansion of the pitcher-plant mosquito, *Wyeomyia smithii*, in North America. *Plos One*, 8(9), e72262.

- Satler, J. D., & Carstens, B. C. (2016). Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. *Evolution*, 70(5), 1105-1119.
- Simons, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Floors, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87(6), 651-701. doi:10.1093/aesa/87.6.651
- Stephens, J. D., Rogers, W. L., Heyduk, K., Cruse-Sanders, J. M., Determann, R. O., Glenn, T. C., & Malmberg, R. L. (2015). Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Molecular Phylogenetics and Evolution*, 85, 76-87.