# GENOMIC CONSEQUENCES OF HYBRIDIZATION AND CARNIVORY IN THE PITCHER PLANT GENUS SARRACENIA

by

ETHAN BALDWIN

(Under the Direction of Jim Leebens-Mack)

## ABSTRACT

Hybridization has long fascinated evolutionary biologists. Hybridization leads to novel combinations of genes which can lead to diverse evolutionary outcomes; maladaptive genetic combinations may promote the reinforcement of species boundaries, while adaptive combinations may drive new adaptations in existing populations or even generate new species. With the continual decline in sequencing costs and advances in analytical methods, it is becoming increasingly feasible to determine how hybridization has contributed to the evolution of lineages. The North American pitcher plant genus *Sarracenia* is a charismatic and ecologically significant group of carnivorous plants known to frequently form fertile hybrids in nature. Surprisingly little is known about the extent of interspecific gene flow across the genus and what mechanisms are maintaining species boundaries in the face of hybridization. In this dissertation, I investigate interspecific gene flow in *Sarracenia* plastid and nuclear genomic compartments, assess hybrid unfitness and genetic architecture as a reproductive barrier, and develop genomic resources for *Sarracenia*. I first assembled whole plastid genomes and generated phylogeny, and employ coalescent simulations that reveals rampant introgression of the chloroplast genome has occurred. Next, I assemble the first two nuclear genomes for *Sarracenia*. Despite large, repeat-rich genomes, I found that Sarracenia has lost thousands of genes —notably those involved in photosynthesis and pathogen recognition—consistent with functional shifts associated with carnivory and prey-derived nutrient uptake. Using over 3000 single-copy loci, I reconstructed a robust species phylogeny and applied phylogenetic network estimation and quartet-based D-statistics, revealing widespread episodes of nuclear gene flow among sympatric species. Finally, I investigated the genetic basis of ecologically important pitcher traits in an F2 mapping population between *S. rosea* and *S. psittacina*, revealing that many of these traits are controlled by relatively simple genetic architecture. Additionally, a common garden experiment revealed that prey-derived nitrogen uptake was significantly reduced in hybrids compared to *S. rosea*, suggesting that hybrid pitcher morphologies reduce prey capture success and likely contribute to post-zygotic reproductive isolation. Together, these results demonstrate that interspecific hybridization has been a pervasive force in Sarracenia evolution, while divergent prey capture strategies may result in hybrid unfitness, contributing to post-zygotic reproductive isolation.

INDEX WORDS: Sarracenia, carnivorous plants, hybridization, gene flow, introgression, genome assembly, QTL, genome architecture

# GENOMIC CONSEQUENCES OF HYBRIDIZATION AND CARNIVORY IN THE PITCHER PLANT GENUS SARRACENIA

by

ETHAN BALDWIN

B.S., University of Florida, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2025

# GENOMIC CONSEQUENCES OF HYBRIDIZATION AND CARNIVORY IN THE PITCHER PLANT GENUS SARRACENIA

by

ETHAN BALDWIN

| | |
|---|---|
| Major Professor: | Jim Leebens-Mack |
| Committee: | Andrea Sweigart |
| | John Burke |
| | Lauren Eserman-Campbell |

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2025

# DEDICATION

I dedicate this dissertation to my wife, Emily. Your unwavering love and support have made getting through the lows possible and made celebrating the highs worth it. All of the struggles that I have been through seem trivial when I know that I can come home to my best friend every day. Thank you so much for coming on this journey with me.

# ACKNOWLEDGEMENTS

I want to thank my advisor, Jim Leebens-Mack, for being an outstanding mentor and providing the opportunity for me to do this work under his direction. Jim has been exceptionally generous with his time and resources, which has allowed me to work on this project that is outside the scope of his current work. This opportunity to conceptually develop a project essentially from scratch is a rare privilege that has fostered independence, creativity, and curiosity in myself as a developing scientist. I am eternally grateful to Jim for granting me this opportunity.

I also want to thank my lab mates throughout the years: Mason, Rick, Phil, Dakota, Summer, Chazz, Jeremiah, Deanna, Bruce, Monstera, and Ryan. Many lab meetings, conversations, wet lab sessions, and conferences have contributed to this work. Without the support, comradery, and thoughtful considerations they have provided, none of this would have been possible. In particular, I want to thank Mason for recruiting me into Jim's lab and perpetually being a source of knowledge on *Sarracenia* that I could lean on.

I want to acknowledge the late Russell Malmberg's research on *Sarracenia* and the foundation he has provided for my work and future researchers at UGA who are interested in this wonderful genus. I also want to give a huge thank you to Will Rogers for carrying on Russell's legacy while providing endless support for my project through propagating and maintaining the *Sarracenia* mapping population, consistently generating new ideas driven by his unending curiosity about *Sarracenia*, and always being extraordinarily generous with his time.

Lastly, I want to thank my parents, who have been an unwavering support system since (literally) day one. Even when they probably have no idea what I have been doing for the past five years, they continue to nod along enthusiastically knowing that I am happy with what I am doing.

# TABLE OF CONTENTS

Page

CHAPTER

# LIST OF TABLES

Page

# LIST OF FIGURES

## CHAPTER 1

## INTRODUCTION AND LITERATURE REVIEW

### The genus Sarracenia

The pitcher plant genus *Sarracenia* L. (Sarraceniaceae, Ericales) is a charismatic group of carnivorous plants native to eastern North America. Each of the ~10 species of *Sarracenia* produces pitcher-shaped leaves that attract, capture, and digest insect prey, providing the plant with mineral nutrients. Due to this unique adaptation they are commonly called pitcher plants, although pitcher-shaped carnivorous leaves have evolved convergently in at least two other lineages (*Nepenthes* L., *Cephalotus* Labill.) (Albert et al. 1992). *Sarracenia* species perform vital ecosystem services, providing microenvironments for numerous inquiline arthropod species (Harvey and Miller 1996) and serving as host for moths in the genus *Exyra* (Lepidoptera, Noctuidae), which spend most of their lives in *Sarracenia* pitchers (Stephens and Folkerts 2012).

The pitchers of *Sarracenia* species are diverse in form and function (Fig. 1). Except for *S. psittacina,* all species' pitchers are pitfall traps, which, as the name suggests, act as a pit which insect prey will fall into and be digested in. *S. psittacina* has traps that are comparable to lobster traps. The pitchers have a small, funneled opening which protrudes into the trap itself, making it easy for insects to crawl into but difficult for them to find their way out. The only other pitcher plant with a lobster pot trap is *Darlingtonia californica,* another member of Sarraceniaceae. Of the species with pitfall traps, most have hoods that prevent an excess of rainwater from entering the pitcher.

These species are relatively tall (30-70 cm), and too much water would cause the pitchers to fall over. On the other hand, *S. purpurea* and *S. rosea* produce pitfall trap pitchers that are much shorter (9-16 cm) and allow rainwater to fill the pitcher. These differences, in addition to differences in color, size, nectar production, etc., likely contribute to differences in prey captured (Folkerts 1992).

**Figure 1.1**. Pitcher trap diversity in Sarracenia. Left: *S. leucophylla*. Top middle: *S. flava*. Bottom middle: *S. minor*. Top right: *S psittacina*. Bottom right: *S. purpurea*.

*Sarracenia* species are endemic to eastern North America and occur in wet, nutrient-poor habitats with open canopies that are often maintained by fire (Folkerts 1982). It is estimated that 97% of suitable habitat for *Sarracenia* has been lost due to human activity (Folkerts 1982). Due to this unprecedented habitat loss, almost all species are of conservation concern, including three taxa that are federally endangered (*S. alabamensis, S. jonesii* and *S. oreophila*) and three additional taxa that are being considered for federal listing (*S. rubra ssp. wherryi*, *rubra ssp. gulfensis,* and *S. purpurea var. montana*). Conservation of *Sarracenia* species is complicated by taxonomic uncertainty, especially in the "rubra complex", a group of up to 10 closely related taxa with subtle yet distinct differences in morphology that include some of the most imperiled taxa (*S. alabamensis, S. jonesii*) (Case and Case 1976). While this dissertation does not attempt to resolve any taxonomic issues, implications for the taxonomy of some members of the rubra complex are discussed in relation to the phylogeny in chapter 4.

### Hybridization and species boundaries in *Sarracenia*

Hybridization is common in *Sarracenia*—a morphological hybrid has been found in the wild for almost every sympatric pair of species (Bell 1952). Although *Sarracenia* has high potential for interspecific gene flow, this possibility remains largely unexplored. Two studies exist that examine gene flow between *Sarracenia* species at specific sites using microsatellite loci (Furches et al. 2013;Rentsch and Holland 2020). Furches et al. (2013) found that the species found in their study site (*S. leucophylla, S. alabamensis ssp. wherryi,* and *S. alata*) were exchanging genes when they occurred close to one another. Rentsch and Holland (2020) did not detect gene flow between *S. minor* and *S. flava*. Both studies are limited by a small geographic focus and the use of only eight microsatellite loci. A genus-wide and genomic scale exploration of interspecific gene flow is still necessary to elucidate the full landscape of introgression in *Sarracenia*.

Although the propensity for *Sarracenia* to hybridize is well documented, little is known about what mechanisms maintain cohesive species in the face of hybridization. Bell (1952) hypothesized that differences in flowering phenology is the strongest barrier to reproduction, although this does not account for the formation of F1 hybrids. Another possible reproductive barrier, at least between some species pairs, is hybrid unfitness due to intermediate pitcher morphology.

## Aims

Throughout this dissertation, I aim to elucidate the evolutionary consequences of hybridization in *Sarracenia* through a genomic lens. While pursuing this aim, I assemble the first complete plastid and nuclear genome sequences in *Sarracenia* and uncover how carnivory has contributed to the evolution of these genomes. The first question I seek to answer is if gene flow occurs in *Sarracenia*, and if so, to what extent and between which species. In chapter 2, I construct a phylogeny from whole plastid genome sequences and examine whether the phylogeny is consistent with patterns expected under a model with gene flow occurring. I then turn to the nuclear genome to assess the history of interspecific exchange of nuclear genes in *Sarracenia*. In order to perform these analyses and build a foundation for future genomics work in *Sarracenia* I generated reference genomes for *S. rosea* and *S. psittacina*. Chapter 3 details the sequencing and assembly of two chromosome-level reference genomes for both species, while describing the unique genomic impact of the shift from obligate autotrophy to carnivory. In chapter 4, I estimate a phylogeny and perform an in depth phylogenomic analysis of nuclear gene flow. Finally, I examine the genetic architecture of pitcher traits that contribute to the different prey capture strategies of *S. rosea* and *S. psittacina*, while discussing their potential role in hybrid unfitness and as a barrier to gene flow.

# CHAPTER 2

# RAMPANT CHLOROPLAST CAPTURE IN SARRACENIA REVEALED BY PLASTOME PHYLOGENY[1]

---

[1] Baldwin, E., McNair, M, and Leebens-Mack, J. 2023. *Frontiers in Plant Science*. 14. Reprinted here with permission of the publisher.

# ABSTRACT

Introgression can produce novel genetic variation in organisms that hybridize. Sympatric species pairs in the carnivorous plant genus *Sarracenia* L. frequently hybridize, and all known hybrids are fertile. Despite being a desirable system for studying the evolutionary consequences of hybridization, the extent to which introgression occurs in the genus is limited to a few species in only two field sites. Previous phylogenomic analysis of *Sarracenia* estimated a highly resolved species tree from 199 nuclear genes, but revealed a plastid genome that is highly discordant with the species tree. Such cytonuclear discordance could be caused by chloroplast introgression (i.e. chloroplast capture) or incomplete lineage sorting (ILS). To better understand the extent to which introgression is occurring in *Sarracenia*, the chloroplast capture and ILS hypotheses were formally evaluated. Plastomes were assembled de-novo from sequencing reads generated from 17 individuals in addition to reads obtained from the previous study. Assemblies of 14 whole plastomes were generated and annotated, and the remaining fragmented assemblies were scaffolded to these whole-plastome assemblies. Coding sequence from 80 homologous genes were aligned and concatenated for maximum-likelihood phylogeny estimation. The plastome tree is extremely discordant with the published species tree. Plastome trees were simulated under the coalescent and tree distance from the species tree was calculated to generate a null distribution of discordance that is expected under ILS alone. A t-test rejected the null hypothesis that ILS could cause the level of discordance seen in the plastome tree, suggesting that chloroplast capture must be invoked to explain the discordance. Due to the extreme level

of discordance in the plastome tree, it is likely that chloroplast capture has been common in the evolutionary history of *Sarracenia*.

## INTRODUCTION

Evolutionary biologists have long been interested in hybridization as a process that generates biodiversity. Hybridization leading to introgression introduces genetic information to a species, which increases genetic variation for selection to act on and provides opportunity for adaptive evolution (Pease et al. 2016;Grant and Grant 2019;Meier et al. 2019). Organisms that readily hybridize may be subject to these evolutionary forces. However, the formation of hybrids does not imply that introgression (transfer of genome segments between hybridizing species) is occurring, as hybrids must reproduce with the parental population and introgressed alleles must survive in the face of natural selection and genetic drift. Identifying the extent to which hybridizing taxa are exchanging genetic material sheds light on the processes that generate and maintain variation within them.

*Sarracenia* L. is a genus of 8-11 species of carnivorous plants native to North America. It is one of the three extant genera in the family Sarraceniaceae, with species forming tube shaped traps adapted to catch and digest insects. Due to this unique adaptation they are commonly called pitcher plants, although pitcher-shaped carnivorous leaves have evolved convergently in at least two other lineages (*Nepenthes* L., *Cephalotus* Labill.) (Albert et al. 1992). Most *Sarracenia* species occur sympatrically with at least one other species, and all species pairs can produce fertile hybrids (Bell 1952). Hybrids between sympatric species are frequently observed in nature (Bell 1952), and population genetics studies using a few microsatellite loci have shown evidence of gene-flow between species at some sites and not at others (Furches et al. 2013;Rentsch and Holland 2020). The forces maintaining species boundaries are not well known, but it is possible that outbreeding depression is contributing to species coherence in the face of hybridization. *Sarracenia* hybrids exhibit intermediate pitcher morphology which may decrease prey capture efficacy. Another possible factor contributing to the maintenance of species boundaries is asynchronous flowering phenology (Bell 1952).

*Sarracenia* diverged from the rest of Sarraceniaceae an estimated 23 MYA, with most of the diversification within *Sarracenia* occurring between 1-3 MYA (Ellison et al. 2012). Given the rapid speciation, significant gene tree discordance is expected due to incomplete lineage sorting (ILS) (Degnan and Rosenberg 2009). Despite this, Stephens et al. (Stephens et al. 2015) estimated a multi-species coalescent phylogeny using 199 nuclear genes that resolved most of the species relationships with high support. This study also presented a plastome tree that was highly discordant with the nuclear tree; no species was reciprocally monophyletic. Cytonuclear discordance such as this can be the result of ILS or introgression of the plastid genome, otherwise referred to as chloroplast capture.

Although the plastome phylogeny estimated in Stephens et al. (Stephens et al. 2015) is relatively well supported, the analysis was limited by the recovery of only 42kbp of plastome sequence limited to the long single copy and short single copy regions of the plastome. To confirm that the extreme cytonuclear discordance observed in the Stephens et al. (Stephens et al. 2015) phylogenies was not an artifact of a lack of data, we reassembled plastomes from those sequencing reads using an alternative assembly pipeline to recover more sequence. Seventeen additional accessions are added to this analysis. The cause of cytonuclear discordance is formally assessed using a coalescent based simulation approach to distinguish between ILS and chloroplast capture. Additionally, whole plastomes are assembled and gene content evolution is assessed within the context of carnivory.

## METHODS

### Sequence data

Leaf tissue was obtained from 17 individuals in total: 11 accessions were obtained from the Atlanta Botanical Garden's living conservation collection (*S. oreophila*, *S. jonesii*, *S. alata*, *S. alabamensis* and *S. rubra*) and six accessions were obtained from two field sites (*S. rubra subsp. rubra* and *S. rubra subsp. viatorum*). DNA was extracted from silica dried samples using the Qiagen DNeasy Plant Mini Kit. Library prep was performed using the Kapa Biosystems HyperPlus Kit using iTru adapters (Glenn et al. 2019). Libraries were pooled at equal concentrations and enriched for putative single-copy orthologs enrichment using the Angiosperm**s**353 bait set (Johnson et al. 2019). The enriched pool was sequenced on an Illumina NextSeq 500 at the Georgia Genomics and Bioinformatics Core using a High Output 300 cycle flow cell generating 150bp paired-end reads.

In addition, sequencing reads from Stephens et al. **(Stephens et al. 2015)** were downloaded from NCBI Short Read Archive. The Stephens et al. data set includes 71 accessions of *Sarracenia* and 4 accessions of outgroups in Sarraceniaceae (*Heliamphora minor* and *Darlingtonia californica*).

### Plastome assembly

All raw reads were trimmed using Trimmomatic (v. 0.39) (Bolger et al. 2014). Both the new data set and the data set obtained from Stephens et al. were sequenced from libraries enriched for targeted nuclear loci. However, the majority of the reads from both data sets are off-target. Stephens et al. reported an average of 1.6% of reads on target, and analysis of the new data set revealed that less than 1% of the reads were on target. The large proportion of off-target reads enable the assembly of the plastome.

Initial de-novo plastome assembly was attempted with GetOrganelle (v. 1.7.5.2) (Jin et al. 2020). GetOrganelle often produced two assembly versions differing only in the orientation of the short single copy regions (SSC). SSC orientation was determined by aligning assemblies to the reference plastome (*Clethra* L. *delavayi*, Genbank accession NC_041129) using MUMmer (v. 4.0.0) (Kurtz et al. 2004), and only the assemblies with concordant SSC orientation were retained.

GetOrganelle did not generate complete de-novo plastome assemblies from every sample. In these cases, the following reference-based pipeline was used. Reads were aligned to one of the complete *Sarracenia* plastome assemblies using BWA (v. 0.7.17) (Li et al. 2009). The aligned reads were then extracted and assembled de-novo using SPAdes (Bankevich et al. 2012). Afin (https://github.com/afinit/afin) was used to extend the resulting contigs and fuse any contigs with significant overlap. At this stage, assemblies were either mostly complete (1-3 contigs consisting of the large single copy region (LSC), short single copy regions (SSC), and one IR), or they were more fragmented. The mostly complete assemblies were manually pasted together. The IR boundaries were verified by mapping reads to the assemblies and identifying the coordinate where half of the reads spanned the IR and LSC and the other half spanned the IR and SSC.

### Plastome annotation

Complete plastome assemblies were annotated using PGA (Qu et al. 2019). Fragmented assemblies were aligned to one of the complete, PGA annotated plastomes using the Minimap2 (v. 2.17) (Li 2018) plugin in Geneious. The “transfer annotation” function was used before generating a consensus sequence.

### Alignment and phylogeny estimation

Coding sequences (CDS) from 80 plastid genes were extracted from the annotated assemblies and aligned with MAFFT (v. 7.470) (Katoh and Standley 2013). All resulting gene alignments were concatenated. Regions of the concatenated alignment that were poorly aligned or had gaps in 50% or more of the samples were filtered out of the gene alignments using Gblocks (v. 0.91b) (Castresana 2000). A maximum-likelihood phylogeny was estimated from the concatenated gene alignments using IQ-Tree (v. 2.0.6) (Nguyen et al. 2015). 1000 bootstrap replicates were performed using UFBoot (Minh et al. 2013). The GTR + F + R4 substitution model was used.

### Plastome tree simulations

To differentiate between incomplete lineage sorting (ILS) and chloroplast capture, a tree simulation approach similar to Folk et al. 2017 (Folk et al. 2016) was used. Plastome trees under ILS were simulated using the dendropy python package (v. 4.5.2) (Sukumaran and Holder 2010) with the species tree from Stephens et al. (Stephens et al. 2015) as a guide tree. Since plastomes are effectively haploid and inherited uniparentally, plastomes have one quarter of the effective population size of diploid nuclear loci. Since the guide tree used for these simulations was estimated exclusively using nuclear loci, its branch lengths were scaled by four to account for the effective population size differential between plastomes and nuclear loci. A distribution of tree discordance under the null hypothesis of ILS was generated by calculating a tree distance metric (information-based generalized Robinson-Foulds distance (Smith 2020)) between 1000 simulated trees and the species tree. Then the distance between the empirical plastome tree from this study and the species tree was calculated and compared to the null distribution. Since the empirical plastome tree has samples that are not in the Stephens et al. (Stephens et al. 2015) species tree, those tips were dropped from the plastome tree to enable calculating distance.

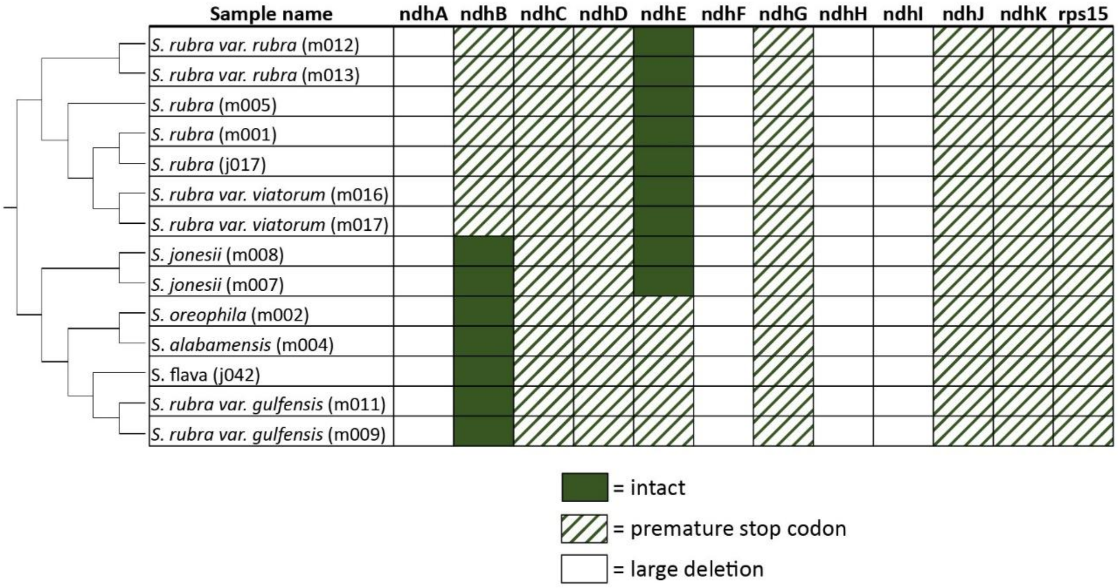
## RESULTS

### Plastome assemblies

Fourteen complete, circularized plastomes have been assembled and annotated including the following *Sarracenia* species: *S. jonesii, S. alabamensis, S. oreophila, S. rubra subsp. gulfensis, S. rubra subsp. rubra,* and *S. rubra subsp. viatorum*. Average assembly statistics for the all assemblies are shown in Table 1. The assembly pipeline for fragmented assemblies recovered an average of 114kbp of plastome sequence, almost tripling the 42kbp recovered in Stephens et al. (Stephens et al. 2015). The use of different references is one potential factor explaining this difference; this study used a complete *Sarracenia* plastome (Ericales) as a reference whereas Stephens et al. (Stephens et al. 2015) used a plastome from *Vitis vinifera* (Vitales). Eighty protein-coding genes were extracted from assemblies, and sequences were aligned for all samples, and alignments were concatenated for the phylogeny estimation.

## Pseudogenization of plastome encoded genes

All complete *Sarracenia* plastomes include some pseudogenized plastome-encoded genes. With the exception of *ndhB* and *ndhE*, all *ndh* genes either have been pseudogenized due to premature stop codons or large deletions (Figure 1). Similarly, all samples contain a premature stop codon within the *rps12* gene.



| Sample name | ndhA | ndhB | ndhC | ndhD | ndhE | ndhF | ndhG | ndhH | ndhI | ndhJ | ndhK | rps15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. rubra var. rubra* (m012) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra var. rubra* (m013) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra* (m005) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra* (m001) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra* (j017) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra var. viatorum* (m016) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra var. viatorum* (m017) | □ | ▨ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. jonesii* (m008) | □ | ■ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. jonesii* (m007) | □ | ■ | ▨ | ▨ | ■ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. oreophila* (m002) | □ | ■ | ▨ | ▨ | ▨ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. alabamensis* (m004) | □ | ■ | ▨ | ▨ | ▨ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| S. flava (j042) | □ | ■ | ▨ | ▨ | ▨ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra var. gulfensis* (m011) | □ | ■ | ▨ | ▨ | ▨ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |
| *S. rubra var. gulfensis* (m009) | □ | ■ | ▨ | ▨ | ▨ | □ | ▨ | □ | □ | ▨ | ▨ | ▨ |

■ = intact
▨ = premature stop codon
□ = large deletion

**Figure 2.1.** Status of ndh genes in all complete plastome assemblies. Filled cells represent intact genes, dashed cells represent premature stop codons, and blank cells represent genes with large deletions. Plastome tree trimmed from these samples is shown on the left.

## Plastid phylogeny

Consistent with Stephens et al. (Stephens et al. 2015), no species were found to exhibit monophyly of their plastomes, and the plastid tree is highly incongruent with the published species tree (Figure 2). Support values across the backbone of the tree are all greater than 70, and most internal nodes are highly supported as well (Figure 2). Branch

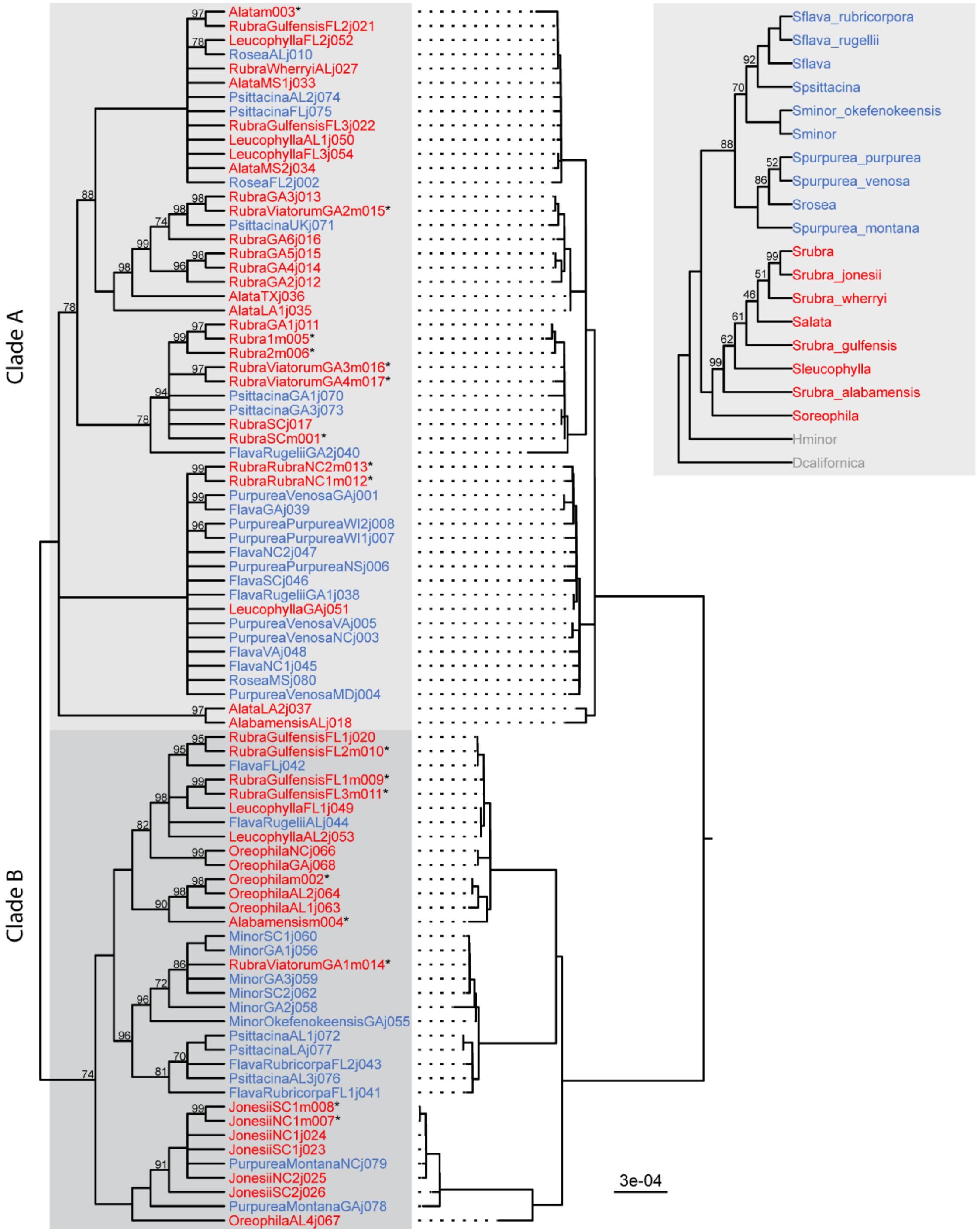


**Figure 2.2.** Maximum likelihood plastome cladogram (left) and phylogram (center) and species tree (inset, top right) from Stephens et al. (2015). Nodes on the cladogram with bootstrap values less than 70 are collapsed. Uncollapsed cladogram nodes with bootstrap values less than 100 are labelled. Tip names are either red or blue based on which of the two major clades the species belongs to in the species tree. Asterisks next to tip labels indicate samples that were newly sequenced for this study.

lengths within *Sarracenia* are generally very short in comparison to the outgroups. An exception is the split at the base of the *Sarracenia* clade. This branch splits *Sarracenia* into two distinct plastid lineages. These main lineages are arbitrarily termed clade A and clade B (Figure 2). Clade B contains all sampled individuals of *minor, oreophila, jonesii,* and *purpurea var. montana*, and clade A contains all sampled individuals of *alata* and *purpurea* (excluding *var. montana*).  All other species are split across these two main lineages (*flava, psittacina, rubra,* and *leucophylla*).

## Southern Appalachian species

*S. purpurea var. montana* and *S. jonesii* form a clade. Both taxa have distributions restricted to a small area in the southern Appalachian Mountains (Figure 3) and hybridize at sympatric sites. The only other species found in the southern Appalachians is *S. oreophila*, although it is not sympatric with *S. jonesii* or *S. purpurea var. montana*, but may have been historically (McPherson and Schnell 2011). Two *S. oreophila* accessions from Alabama are sister to the Appalachian clade, and the other *S. oreophila* accessions are placed in a clade sister to this.

## ***Sarracenia flava, S. minor, and S. psittacina***

*S. flava, S. minor,* and *S. psittacina* form a clade sister to *S. purpurea* on the species tree, however the placement of these species on the plastid tree is not congruent. All *S. minor* accessions are placed within clade B sister to the clade containing *S. oreophila, S. jonesii,* and *S. purpurea var. montana*. Some *S. flava* and *S. psittacina* accessions from the Gulf coastal plain are also placed in the *S. minor* clade, despite all *S. minor* accessions in this study originating from the Atlantic coastal plain. This could indicate either ancient introgression or retention of plastome diversity from the ancestor of these three species. *S. flava* and *S. psittacina* are scattered across the chloroplast phylogeny; both species have accessions found in clades A and B. In *S. flava*, all Gulf coastal plain accessions are found in clade B and all Atlantic coastal plain accessions are found in clade A.

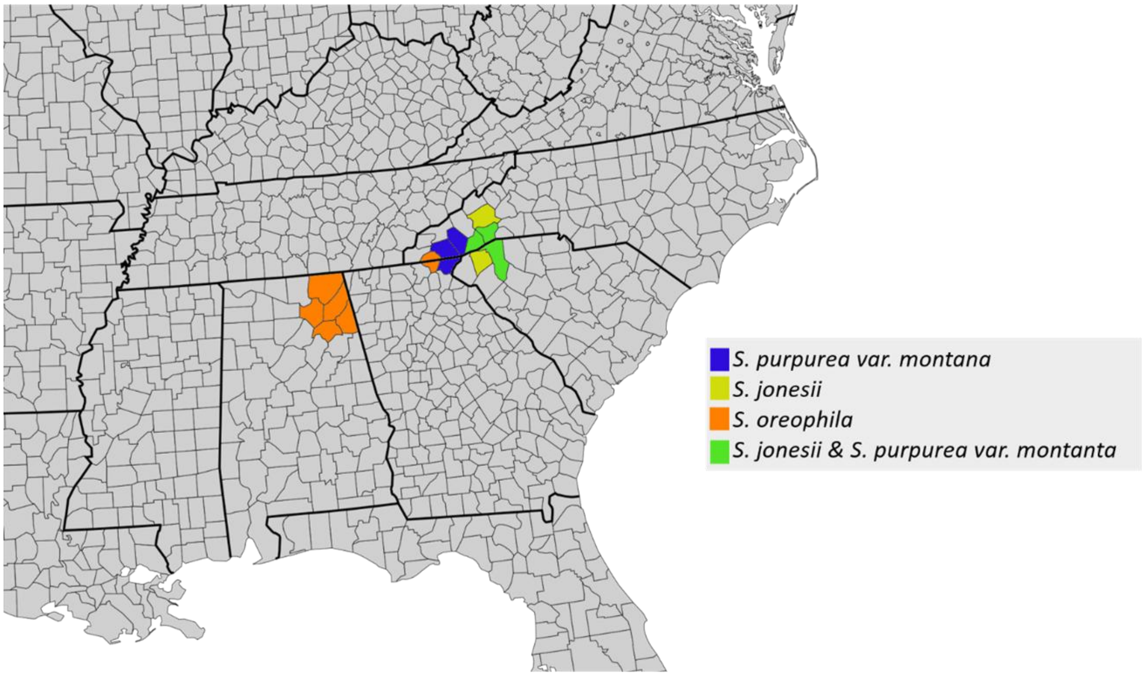


**Figure 2.3**. A county level distribution map for S. purpurea var montana, S. jonesii, and S. oreophila.

### *Sarracenia purpurea* complex

With the exception of *S. purpurea var. montana*, all *S. purpurea* accessions (including *S. rosea*) are placed in clade A. There is no discernible pattern to their placement within this lineage. This is surprising given the vast geographic range represented by these taxa; the individuals sampled for this study originate from throughout their distribution from Mississippi to Nova Scotia. Only *S. purpurea subsp.*

*purpurea* is found north of Maryland, so the relatedness of plastomes between this taxon and other species are unlikely to be the result of recent introgression.

**Plastome phylogeny simulations**

The tree distance metric that was used ranges from 0 (an identical tree) to 1 (the most distal tree). The plastome trees simulated under the pure coalescent model have distances from the species tree ranging from 0.29 to 0.56, while the distance from the empirical plastome tree is 0.73 (Figure 4). A T-test using the distribution of simulated plastome tree distances as the null distribution gives a p-value of >2.2e-16, rejecting the null hypothesis of ILS causing the discordance alone.

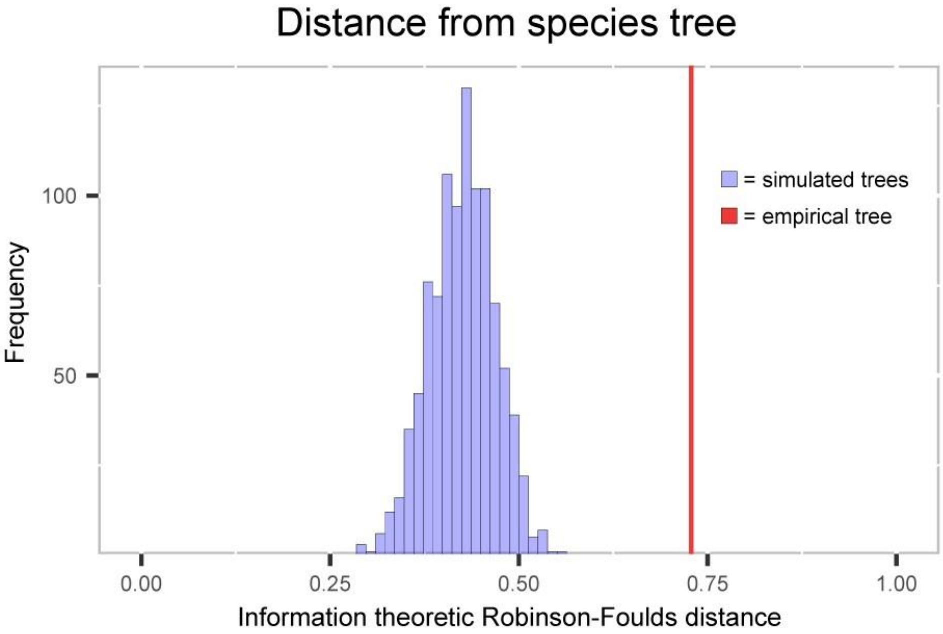


**Figure 2.4**. Histogram of information-based generalized Robinson-Foulds distance between the simulated plastome trees and the species tree. Red line shows the distance between empirically estimated plastome tree and the species tree.

# DISCUSSION

## Pseudogenization of ndh genes

Independent pseudogenization or complete loss of *ndh* genes has been shown in many plant lineages, including holoparasitic, hemiparastitic, and carnivorous plant lineages (Barrett et al. 2014;Lin et al. 2017;Cao et al. 2019;Gruzdev et al. 2019;Nevill et al. 2019). Functional *ndh* genes are rarely found in non-photosynthetic parasitic plants, and the loss of *ndh* genes is strongly correlated with the transition to heterotrophy in parasitic plant lineages (Wicke et al. 2016). Since plastid encoded *ndh* genes are thought to optimize photosynthetic chemistry in fluctuating or stressful environments (reviewed in (Sabater 2021)), the loss of *ndh* genes in parasitic lineages that are no longer fully dependent on photosynthesis as a source of carbon is unsurprising. In carnivorous plants, however, evidence for significant heterotrophic uptake of carbon is limited (Rischer et al. 2002), and a transition to full heterotrophy seems unlikely, so this line of reasoning does not explain the independent pseudogenization of functional *ndh* genes across carnivorous plant lineages. It is possible that the acquisition of organic nitrogen has an interaction with photosynthetic chemistry that relaxes the need for *ndh*. As Nevill et al. (Nevill et al. 2019) noted, organic nitrogen acquisition bypasses the need to assimilate nitrate using photosynthetically-derived reductant. Alternatively, the pseudogenization of *ndh* genes in parasitic plants and carnivorous plants could be due to unrelated mechanisms. The pseudogenization of almost all of the *ndh* genes across the genus *Sarracenia* shown here provides further evidence that carnivorous plants do not require these genes. Sequencing of full plastomes from other carnivorous species would reveal if the pseudogenization of *ndh* occurs early in carnivorous plant evolution.

## Cytonuclear discordance

The plastome phylogeny in this study shows a similarly extreme level of discordance with the species tree as that of the Stephens et al. (Stephens et al. 2015) plastome phylogeny. That study ascribed the discordance to a combination of chloroplast capture and a lack of informative polymorphisms in the chloroplast sequence. A third source of discordance, ILS, is considered here. A lack of informative polymorphisms is not an issue here, as almost all the plastome coding sequences are used and the resulting phylogeny has high bootstrap values across the spine, suggesting that there is sufficient evidence that major clades within the tree are correct.

To distinguish between the two remaining sources of discordance, plastome phylogenies under ILS were simulated. The simulated phylogenies showed much lower levels of discordance with the species tree than the empirically estimated plastome. To simulate the plastome phylogenies, the branch lengths of the guide tree were multiplied by four due to the assumption that the chloroplast is inherited matrilineally in *Sarracenia* like most seed plants (Mogensen 1996). Since this assumption hasn't been empirically proven and biparental inheritance of the chloroplast is possible, simulations with branch lengths multiplied by two were performed and show similar results (supplemental data).

There is ample signal of introgression in the plastome, but Stephens et al. (Stephens et al. 2015) reported no evidence of gene flow in the nuclear data. A search through the nuclear gene trees revealed that none of the trees had a similar topology to the plastome tree. In *Sarracenia*, hybridization and backcrossing are rampant, but little signal of this process is found in the nuclear genome. Cytonuclear discordance is commonly observed and is attributed to introgression in plant and animal systems

(Rieseberg and Soltis 1991;Berthier et al. 2006;Gernandt et al. 2018), including several instances where there is limited signal for introgression in nuclear data (Winkler et al. 2013;Good et al. 2015;Folk et al. 2016;Rose et al. 2020). However, the mechanism for organellar introgressions without accompanying nuclear loci is poorly understood (Rieseberg and Soltis 1991;Folk et al. 2018). *Sarracenia* is a genus where hybridization is common and thus some level of nuclear introgression might be expected. The extreme level of chloroplast capture and lack of signal for nuclear gene flow in *Sarracenia* illustrates the comparative ease of introgression of organelles over nuclear loci.

### Geographic patterns of plastome introgression

Although the lack of monophyletic species in the plastome tree makes it difficult to interpret specific instances of plastome introgression, a handful of such instances can be elucidated using geographic context. For example, all accessions of *S. purpurea var. montana* and *S. jonesii*, two taxa restricted to a small region in the southern Appalachians, form a well-supported clade within clade B. Given that all other *S. purpurea* accessions are placed in clade A, it is likely that a plastome derived from *S. jonesii* was introgressed into *S. purpurea var. montana*. Similarly, an accession of *S. rubra* that was sampled from the Georgia fall line near *S. minor* populations is placed within the *S. minor* clade. Again, we hypothesize this to be an instance of *S. minor* plastome being introgressed into *S. rubra*. More generally, the weak species clustering in the plastome tree implies a long history of interspecific exchange of cytoplasmic genomes in *Sarracenia*.

**Table 2.1**. Accession information and assembly statistics for all samples used in this study.

| Taxon | Sample ID | NCBI Biosample | Collector | Herbarium ID | Total contigs | Total length |
|---|---|---|---|---|---|---|
| ***Darlingtonia californica*** | | | | | | |
| | DarlingtoniaOR_j028 | SAMN03354578 | J. D. Stephens | UGA66 | 36 | 121605 |
| | DarlingtoniaUN1_j029 | SAMN03354579 | J. D. Stephens | N/A | 65 | 116346 |
| | DarlingtoniaUN2_j030 | SAMN03354580 | J. D. Stephens | UGA54 | 60 | 112573 |
| ***Heliamphora minor*** | | | | | | |
| | HeliamphoraVE_j031 | SAMN03354581 | J. D. Stephens | UGA55 | 45 | 130627 |
| ***S. alabamensis*** | | | | | | |
| | AlabamensisAL_j018 | SAMN03354582 | J. D. Stephens | UGA19 | 76 | 110275 |
| | Alabamensis_m004 | SAMN31020169 | E. Baldwin | 1004 | 1 | 154984 |
| ***S. alata*** | | | | | | |
| | AlataMS1_j033 | SAMN03354583 | J. D. Stephens | UGA21 | 51 | 118063 |
| | AlataMS2_j034 | SAMN03354584 | J. D. Stephens | N/A | 62 | 117941 |
| | AlataLA1_j035 | SAMN03354585 | J. D. Stephens | UGA67 | 25 | 125730 |
| | AlataTX_j036 | SAMN03354586 | J. D. Stephens | TAES253951 | 39 | 123698 |
| | AlataLA2_j037 | SAMN03354587 | J. D. Stephens | UGA60 | 38 | 126597 |
| | Alata_m003 | SAMN36416359 | E. Baldwin | 1003 | 29 | 149266 |
| ***S. flava*** | | | | | | |
| | FlavaGA_j039 | SAMN03354588 | J. D. Stephens | UGA15 | 57 | 115895 |
| | FlavaFL_j042 | SAMN03354589 | J. D. Stephens | UGA65 | 1 | 153868 |
| | FlavaNC1_j045 | SAMN03354590 | J. D. Stephens | UGA48 | 34 | 125019 |
| | FlavaSC_j046 | SAMN03354591 | J. D. Stephens | UGA45 | 18 | 129423 |
| | FlavaNC2_j047 | SAMN03354592 | J. D. Stephens | UGA50 | 22 | 129654 |
| | FlavaVA_j048 | SAMN03354593 | J. D. Stephens | UGA64 | 23 | 128338 |
| ***S. flava var. rubricorpora*** | | | | | | |
| | FlavaRubricorpaFL1_j041 | SAMN03354594 | J. D. Stephens | UGA18 | 14 | 132800 |
| | FlavaRubricorpaFL2_j043 | SAMN03354595 | J. D. Stephens | UGA18 | 15 | 128594 |
| ***S. flava var. rugelii*** | | | | | | |
| | FlavaRugeliiGA1_j038 | SAMN03354597 | J. D. Stephens | UGA26 | 52 | 117830 |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | FlavaRugeliiGA2_j040 | SAMN03354598 | J. D. Stephens | UGA44 | 26 | 126308 |
|  | FlavaRugeliiAL_j044 | SAMN03354596 | J. D. Stephens | UGA51 | 8 | 130660 |
| ***S. jonesii*** |  |  |  |  |  |  |
|  | JonesiiSC1_j023 | SAMN03354599 | J. D. Stephens | UGA32 | 24 | 125346 |
|  | JonesiiNC1_j024 | SAMN03354600 | J. D. Stephens | UGA31 | 9 | 127465 |
|  | JonesiiNC2_j025 | SAMN03354601 | J. D. Stephens | UGA33 | 66 | 116650 |
|  | JonesiiSC2_j026 | SAMN03354602 | J. D. Stephens | UGA30 | 53 | 118729 |
|  | JonesiiNC1_m007 | SAMN31020170 | E. Baldwin | 1007 | 1 | 151409 |
|  | JonesiiSC1_m008 | SAMN31020171 | E. Baldwin | 1008 | 1 | 151385 |
| ***S. minor*** |  |  |  |  |  |  |
|  | MinorGA1_j056 | SAMN03354609 | J. D. Stephens | N/A | 23 | 126435 |
|  | MinorGA2_j058 | SAMN03354610 | J. D. Stephens | UGA8 | 82 | 111289 |
|  | MinorGA3_j059 | SAMN03354611 | J. D. Stephens | UGA39 | 57 | 117630 |
|  | MinorSC1_j060 | SAMN03354612 | J. D. Stephens | UGA46 | 20 | 127646 |
|  | MinorSC2_j062 | SAMN03354613 | J. D. Stephens | UGA13 | 32 | 124716 |
| ***S. minor var. okefenokeensis*** |  |  |  |  |  |  |
|  | MinorOkefenokeensisGA_j055 | SAMN03354614 | J. D. Stephens | UGA23 | 22 | 125047 |
| ***S. oreophila*** |  |  |  |  |  |  |
|  | OreophilaAL1_j063 | SAMN03354615 | J. D. Stephens | UGA2 | 9 | 128216 |
|  | OreophilaAL2_j064 | SAMN03354616 | J. D. Stephens | UGA28 | 32 | 126491 |
|  | OreophilaAL3_j065 | SAMN03354617 | J. D. Stephens | UGA27 | 66 | 95604 |
|  | OreophilaNC_j066 | SAMN03354618 | J. D. Stephens | UGA20 | 29 | 124312 |
|  | OreophilaAL4_j067 | SAMN03354619 | J. D. Stephens | UGA24 | 40 | 120414 |
|  | OreophilaGA_j068 | SAMN03354620 | J. D. Stephens | UGA22 | 43 | 118692 |
|  | Oreophila_m002 | SAMN31020172 | E. Baldwin | 1002 | 1 | 156118 |
| ***S. psittacina*** |  |  |  |  |  |  |
|  | PsittacinaGA1_j070 | SAMN03354621 | J. D. Stephens | UGA43 | 40 | 121580 |
|  | PsittacinaAL1_j072 | SAMN03354623 | J. D. Stephens | UGA11 | 38 | 128169 |
|  | PsittacinaGA3_j073 | SAMN03354624 | J. D. Stephens | UGA10 | 23 | 129214 |
|  | PsittacinaAL2_j074 | SAMN03354625 | J. D. Stephens | UGA1 | 43 | 122414 |
|  | PsittacinaFL_j075 | SAMN03354626 | J. D. Stephens | UGA35 | 36 | 119271 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | PsittacinaAL3_j076 | SAMN03354627 | J. D. Stephens | UGA53 | 18 | 128493 |
| | PsittacinaLA_j077 | SAMN03354628 | J. D. Stephens | UGA59 | 29 | 126403 |
| ***S. purpurea ssp. purpurea*** | | | | | | |
| | PurpureaPurpureaNS_j006 | SAMN03354629 | J. D. Stephens | UGA61 | 37 | 124790 |
| | PurpureaPurpureaWI1_j007 | SAMN03354630 | J. D. Stephens | UGA47 | 39 | 119984 |
| | PurpureaPurpureaWI2_j008 | SAMN03354631 | J. D. Stephens | UGA47 | 56 | 125476 |
| ***S. purpurea ssp. venosa*** | | | | | | |
| | PurpureaVenosaGA_j001 | SAMN03354463 | J. D. Stephens | UGA12 | 28 | 126463 |
| | PurpureaVenosaNC_j003 | SAMN03354632 | J. D. Stephens | UGA49 | 33 | 124039 |
| | PurpureaVenosaMD_j004 | SAMN03354633 | J. D. Stephens | UGA62 | 61 | 121892 |
| | PurpureaVenosaVA_j005 | SAMN03354634 | J. D. Stephens | UGA63 | 48 | 123570 |
| ***S. purpurea ssp. venosa var. montana*** | | | | | | |
| | PurpureaMontanaGA_j078 | SAMN03354636 | J. D. Stephens | UGA41 | 31 | 126220 |
| | PurpureaMontanaNC_j079 | SAMN03354635 | J. D. Stephens | UGA34 | 39 | 118268 |
| ***S. rosea (S. purpurea ssp. venosa var. burkii)*** | | | | | | |
| | RoseaFL2_j002 | SAMN03354640 | J. D. Stephens | UGA5 | 54 | 121220 |
| | RoseaFL1_j009 | SAMN03354637 | J. D. Stephens | UGA16 | 83 | 105680 |
| | RoseaAL_j010 | SAMN03354638 | J. D. Stephens | UGA4 | 51 | 122458 |
| | RoseaMS_j080 | SAMN03354639 | J. D. Stephens | UGA7 | 28 | 127190 |
| ***S. rubra*** | | | | | | |
| | RubraGA1_j011 | SAMN03354641 | J. D. Stephens | UGA42 | 35 | 123685 |
| | RubraGA2_j012 | SAMN03354642 | J. D. Stephens | UGA58 | 17 | 130198 |
| | RubraGA3_j013 | SAMN03354643 | J. D. Stephens | UGA37 | 34 | 126171 |
| | RubraGA4_j014 | SAMN03354644 | J. D. Stephens | UGA36 | 22 | 128963 |
| | RubraGA5_j015 | SAMN03354645 | J. D. Stephens | UGA36 | 37 | 124041 |
| | RubraGA6_j016 | SAMN03354646 | J. D. Stephens | UGA14 | 45 | 114813 |
| | RubraSC_j017 | SAMN03354661 | J. D. Stephens | N/A | 1 | 154655 |
| | RubraSC_m001 | SAMN31020178 | E. Baldwin | 1001 | 1 | 155181 |
| | Rubra1_m005 | SAMN31020173 | E. Baldwin | 1005 | 1 | 155212 |
| | Rubra2_m006 | SAMN36416360 | E. Baldwin | 1006 | 8 | 128073 |
| ***S. rubra ssp. gulfensis*** | | | | | | |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | RubraGulfensisFL1_j020 | SAMN03354647 | J. D. Stephens | UGA3 | 65 | 115074 |
|  | RubraGulfensisFL2_j021 | SAMN03354648 | J. D. Stephens | UGA29 | 59 | 109430 |
|  | RubraGulfensisFL3_j022 | SAMN03354649 | J. D. Stephens | UGA25 | 23 | 125074 |
|  | RubraGulfensisFL1_m009 | SAMN31020174 | E. Baldwin | 1009 | 1 | 154989 |
|  | RubraGulfensisFL2_m010 | SAMN36416361 | E. Baldwin | 1010 | 9 | 127752 |
|  | RubraGulfensisFL3_m011 | SAMN31020175 | E. Baldwin | 1011 | 1 | 154974 |
| ***S. rubra ssp. Rubra*** |  |  |  |  |  |  |
|  | RubraRubraNC1_m012 | SAMN31020176 | E. Baldwin | 1012 | 1 | 155283 |
|  | RubraRubraNC2_m013 | SAMN31020177 | E. Baldwin | 1013 | 1 | 155302 |
| ***S. rubra ssp. viatorum*** |  |  |  |  |  |  |
|  | RubraViatorumGA1_m014 | SAMN36416362 | E. Baldwin | 1014 | 4 | 128950 |
|  | RubraViatorumGA2_m015 | SAMN36416363 | E. Baldwin | 1015 | 13 | 132536 |
|  | RubraViatorumGA3_m016 | SAMN31020179 | E. Baldwin | 1016 | 1 | 155157 |
|  | RubraViatorumGA4_m017 | SAMN31020180 | E. Baldwin | 1017 | 1 | 155185 |
| ***S. rubra ssp. wherryi*** |  |  |  |  |  |  |
|  | RubraWherryiAL_j027 | SAMN03354650 | J. D. Stephens | UGA38 | 38 | 123432 |
| ***S.leucophylla*** |  |  |  |  |  |  |
|  | LeucophyllaFL1_j049 | SAMN03354603 | J. D. Stephens | UGA57 | 11 | 129896 |
|  | LeucophyllaAL1_j050 | SAMN03354604 | J. D. Stephens | UGA40 | 19 | 127213 |
|  | LeucophyllaGA_j051 | SAMN03354605 | J. D. Stephens | UGA17 | 19 | 129788 |
|  | LeucophyllaFL2_j052 | SAMN03354606 | J. D. Stephens | UGA56 | 24 | 132845 |
|  | LeucophyllaAL2_j053 | SAMN03354607 | J. D. Stephens | UGA52 | 12 | 132508 |
|  | LeucophyllaFL3_j054 | SAMN03354608 | J. D. Stephens | UGA6 | 20 | 126675 |

# CHAPTER 3

# GENOME EXPANSION AND GENE LOSS: ELUCIDATING CARNIVORY'S FOOTPRINT IN SARRACENIA [2]

---

[2] Baldwin, E., Rogers, W., and Leebens-Mack, J. To be submitted to American Journal of Botany.

# INTRODUCTION

The pitcher plant genus *Sarracenia* L. (Sarraceniaceae, Ericales) comprises charismatic carnivorous plants native to eastern North America. Each of the 10 species of *Sarracenia* produces tube-shaped leaves that attract, capture, and digest insect prey, providing the plant with mineral nutrients. *Sarracenia* species perform vital ecosystem services, providing moist microenvironments for numerous inquiline arthropod species (Harvey and Miller 1996) and serving as host for moths in the genus *Exyra*, which spend most of their lives in *Sarracenia* pitchers (Stephens and Folkerts 2012). Due to widespread loss of *Sarracenia*'s habitat, almost all species are of conservation concern, including two taxa that are federally endangered (*S. jonesii* and *S. oreophila*).

Carnivory in plants is an evolutionary innovation that at arisen at least 10 times across the angiosperm tree of life (Fleischmann et al. 2017), and is therefore an important trait for studying the convergent evolution of complex traits. However, only four of the ten carnivorous lineages have published genome assemblies (Fukushima et al. 2017;Lan et al. 2017;Hartmann et al. 2020;Palfalvi et al. 2020), with the Lentibulariaceae and Caryophylales representing the sole carnivorous plant lineages with long-read genome assemblies. Sequencing the genomes of species from the Sarraceniaceae (which includes *Sarracenia*, *Heliamphora* and *Darlingtonia*) is an important step in closing this gap.

Here, we exploit recent technological advances to generate fully phased chromosome-level genome assemblies from two *Sarracenia* species from PacBio HiFi and Omni-C reads derived from an F1 hybrid. We perform comparative analyses to identify ancestral shifts in gene content that we hypothesize are associated with the evolution of carnivory. In addition, these assemblies will serve as invaluable resources for conservation practitioners working to safeguard the remaining populations of rare *Sarracenia* species. They will also be integral to broader comparative analyses seeking to understand the mechanisms underlying the repeated evolution of carnivory in angiosperms.

## MATERIALS AND METHODS

### DNA and RNA extraction and sequencing

We have sequenced an F1 hybrid of *S. psittacina* and *S. rosea* in order to generate two genome assemblies for the price of one and leverage the genomes for future analyses of an F2 mapping population (Malmberg et al. 2018). *S. rosea* has pitchers with wide openings that are typical for *Sarracenia* species and other pitcher plants with pitfall traps, while *S. psittacina* has pitchers with extremely narrow, funnel-shaped openings which act as “lobster pot traps” (Fig. 1A). *S. psittacina*, *S. rosea*, and their F1 hybrid were obtained from a greenhouse at UGA, where clones have been maintained for nearly two decades (Malmberg et al. 2018). High molecular weight DNA was obtained from the F1 by first isolating nuclei according to the “Isolating nuclei from plant tissue using TissueRuptor disruption” protocol available from PacBio, and then extracting the DNA from the nuclei using the Nanobind plant nuclei kit. The high molecular weight DNA was sent to Hudson Alpha where SMRTbell libraries were prepared and sequenced on four SMRT cells on a Revio. Tissue from the F1 was sent to Hudson Alpha where an Omni-C library was prepared and sequenced on an S4 flow cell on a NovaSeq 6000.

Illumina shotgun sequencing was performed on the parents of the sequenced F1 genotype for trio-binning (Koren et al. 2018). DNA was isolated from fresh or dry tissue using Qiagen DNeasy Plant Pro kits. Libraries were constructed with Kapa HyperPlus library kits using custom adapters and iTru primers from Adapterama I (Glenn et al. 2019). Sequencing was done on 10B flow cells on a NovaSeq X at SeqCenter in Pittsburgh, PA.

RNA was extracted from *S. psittacina* and *S. rosea* young pitchers, mature pitchers, and roots using Zymo Direct-zol RNA kits, substituting Invitrogren Plant RNA reagent for the TRI reagent. Sequencing libraries were constructed using Kapa Stranded mRNA-Seq kits using custom adapters and iTru primers from Adapterama I (Glenn et al. 2019). Sequencing was done on 10B flow cells on a NovaSeq X at SeqCenter in Pittsburgh, PA.

**Genome assembly and scaffolding**

Assembly of the F1 HiFi reads was performed with the trio binning (Koren et al. 2018) method in hifiasm (Cheng et al. 2021). Parental 31-mers were counted from the shotgun sequencing reads using yak (https://github.com/lh3/yak) and used as input for hifiasm.

Scaffolding was done with Omni-C sequencing data derived from the F1. The Omni-C reads were mapped to each assembly using BWA-MEM(Li and Durbin 2009) with the -5SP flag. PCR duplicates  were removed using SAMBLASTER(Faust and Hall 2014) and reads were sorted using SAMtools (Li et al. 2009). Initial scaffolding was done with YaHS (Zhou et al. 2023), and final scaffolding was manually performed using Juicebox Assembly Tools (Durand et al. 2016;Dudchenko et al. 2018).

**Annotation**

Repeat sequences in the genome assemblies were identified by using RepeatModeler to produce TE libraries (http://www.repeatmasker.org/RepeatModeler/).

The TE libraries were then used to annotate TEs and mask repeat regions using RepeatMasker (http://www.repeatmasker.org/RepeatMasker/). The BRAKER3 pipeline was used to annotate protein coding genes using RNA-seq and protein homology. RNA-seq reads from *S. psittacina* and *S. rosea* were mapped to their respective genomes using HISAT v. 2.2.1 (Kim et al. 2019). Eudicot protein sequences were downloaded from OrthoDB v. 11 (Kuznetsov et al. 2022).

### Gene loss and gene family evolution analysis

Gene models for both parental haplotype assembles were assigned to orthogroups using OrthoFinder 2.5.4 (Emms and Kelly 2019) in an analysis including proteomes from nine other angiosperm species selected to represent major flowering plant clades and include multiple close relatives of *Sarracenia*. The additional species include *Amborella trichopoda*, the sister lineage to all other extant angiosperms, *Oryza sativa*, a monocot, *Arabidopsis thaliana*, the model asterids *Mimulus guttatus* and *Solanum lycopersicum,* and four species more closely related to *Sarracenia* within *Ericales – Vaccinium darrowii, Camellia sinensis, Actinidia eriantha,* and *Actinidia chinensis*. After orthogroup assignment, we analyzed the functional profile of orthogroups that were lost in *Sarracenia*. Lost orthogroups were defined as those where both *Sarracenia* species had no orthologs and at least four of the six rosid species contained orthologs. GO term enrichment of missing orthogroups was done using the clusterProfiler R package (Yu et al. 2012). An *Arabidopsis* ortholog from each of the missing orthogroups was used as the foreground set, and an *Arabidopsis* ortholog from each orthogroup was used as the background. P values were adjusted using false discovery rate, and a cut off of 0.05 was used.

To model gene family evolution, we first generated a species tree with orthogroups identified as being conserved in single copy by OrthoFinder. To ensure accurate gene trees, OrthoFinder was run with the multiple sequence alignment option, and IQ-TREE2 was used as the gene tree estimation software (Nguyen et al. 2015). The species tree output from OrthoFinder was made ultrametric using the make_ultrametric.py command in OrthoFinder with a root age of 140 million years ago based on the divergence time between *Amborella* and all other angiosperms (Magallón et al. 2015). To reduce the complexity of the model, the species tree was pruned to include only the Ericales genomes and *Solanum* as an outgroup. CAFE5 was used to model gene family expansion and contraction along this reduced species tree under the gamma model with the number of rate categories (k) varying from 1-4 to determine which k value best fit the data (Mendes et al. 2020). A k value of 3 was found to have the highest likelihood, so all results presented are from the k=3 model. Orthogroups with more than 100 orthologs in any given species were removed from the analysis using the clade_and_size_filter.py script from CAFÉ. The functional enrichment of expanding and contracting gene families on the branch leading to *Sarracenia* were analyzed for GO term enrichment as described above.

## RESULTS

### Two chromosome scale reference genomes for *Sarracenia*

We assembled *S. rosea* and *S. psittacina* genomes using 337gb of Pacbio Hifi reads from their F1 (~48X coverage per haplotype) in addition to Illumina WGS reads from both parents for trio binning. The resulting genome assemblies are 3488mb and 3594mb long for S. rosea and S. psittacina respectively, coinciding closely with published estimates of genome sizes across Sarracenia based on flow cytometry (Veleba et al. 2020). After scaffolding using Omni-C reads generated from the F1, 96.6% of *S. psittacina* and 98.2% of *S. rosea* assemblies were assigned to psuedochromosomes (Table 1). The assemblies have embryophyte BUSCO scores of 97.8% and 98.9% for *S. psittacina* and *S. rosea* respectively (Fig. 1B), and ~22000 gene models were annotated in each of the assemblies (Supplementary Table 1).

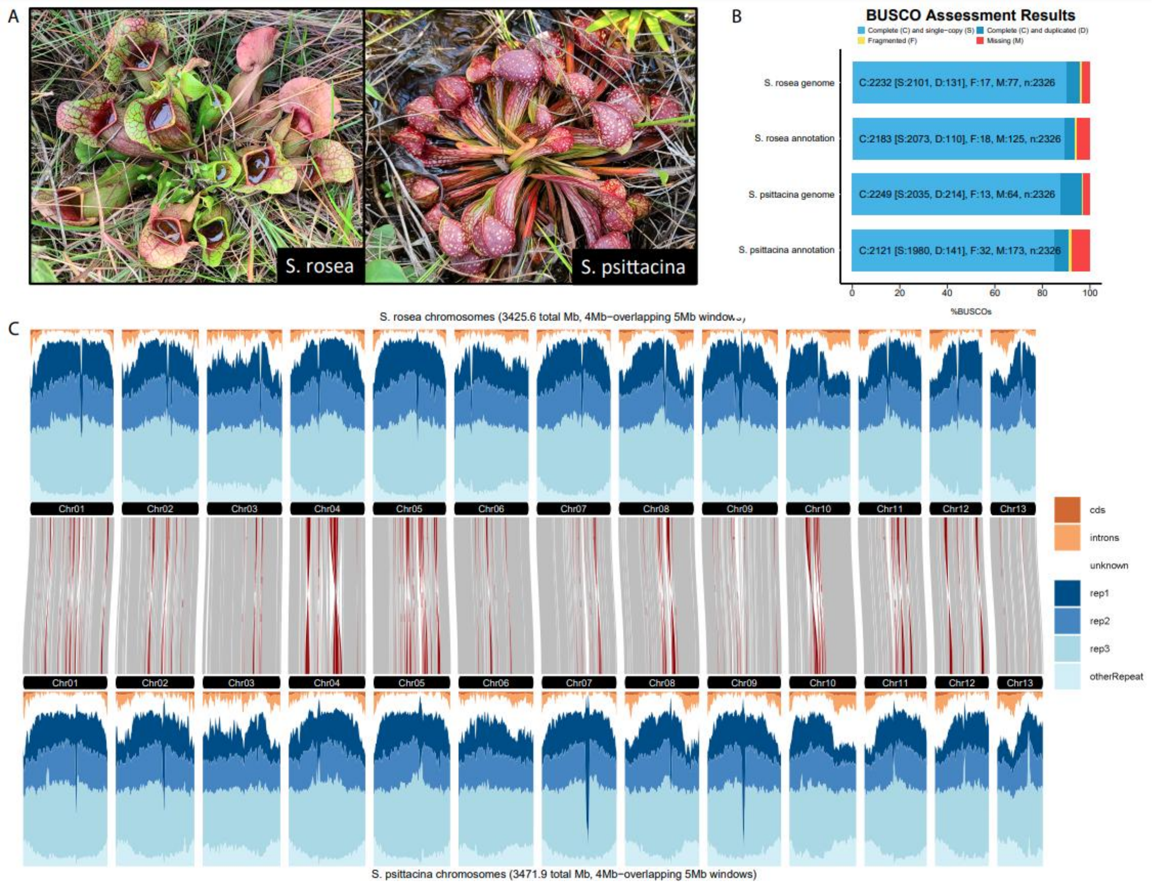


**Figure 3.1**. *Sarracenia* and its genome assembly. (A) In-situ photographs of *S. rosea* and *S. psittacina*. (B) BUSCO results. (C) Comparison of genome gene and repeat content between *S. rosea* and *S. psittacina*. Repeats are as follows: Rep1 = Ty3, Rep2 = Ty1-copia, Rep3 = unknown LTR. The blocks connecting the two genomes in the middle represent nucleotide alignments. Grey blocks represent colinear alignments, while red blocks represent inversions.

Both *Sarracenia* genomes were found to be highly repetitive, with annotated repeats comprising ~87% of the assemblies, which is considerably higher than any other sequenced genome in the Ericales (*Diospyros oleifera* at 64.96%) (Zhu et al. 2019). The most abundant repetitive elements are LTR retrotransposons, half of which belong to unknown LTR families and the remainder approximately equal proportions of Ty1 and Ty3 elements. These elements are concentrated in pericentromeric regions, while genes are concentrated in the chromosome arms (Fig. 1C), which is a typical configuration for angiosperm genomes (Neumann et al. 2011;Sigman and Slotkin 2016). *S. rosea* and *S. psittacina* genomes are highly colinear across gene-rich regions, with many small structural rearrangements in the pericentromeric regions (Fig. 1C).

### Significant gene loss in *Sarracenia*

Despite its large genome size, *Sarracenia* has significantly fewer annotated gene models than its relatives in the Ericales, which in part is due to experiencing widespread gene family contractions, including complete loss of many gene families during its evolution (Fig 2). We found that 3654 orthogroups (i.e. gene families) have contracted on the branch leading to the last common ancestor of the two *Sarracenia* species while only 751 gene families have expanded. This is by far the largest number of contracted gene families in any of the internal branches in this analysis, and is only smaller than the terminal branch leading to *Camellia*, which also has a significantly large number of expansions. In addition, 934 orthogroups are conserved in the six other asterids included in our analysis but orthologs are completely absent in the *Sarracenia* genomes.

### Genes lost in *Sarracenia* enriched in photosynthesis and immune response

GO term enrichment of the contracted and missing gene families in *Sarracenia* genomes revealed that *Sarracenia* has lost genes related to several key biological functions. Whereas only one GO functional annotation term (translation) was

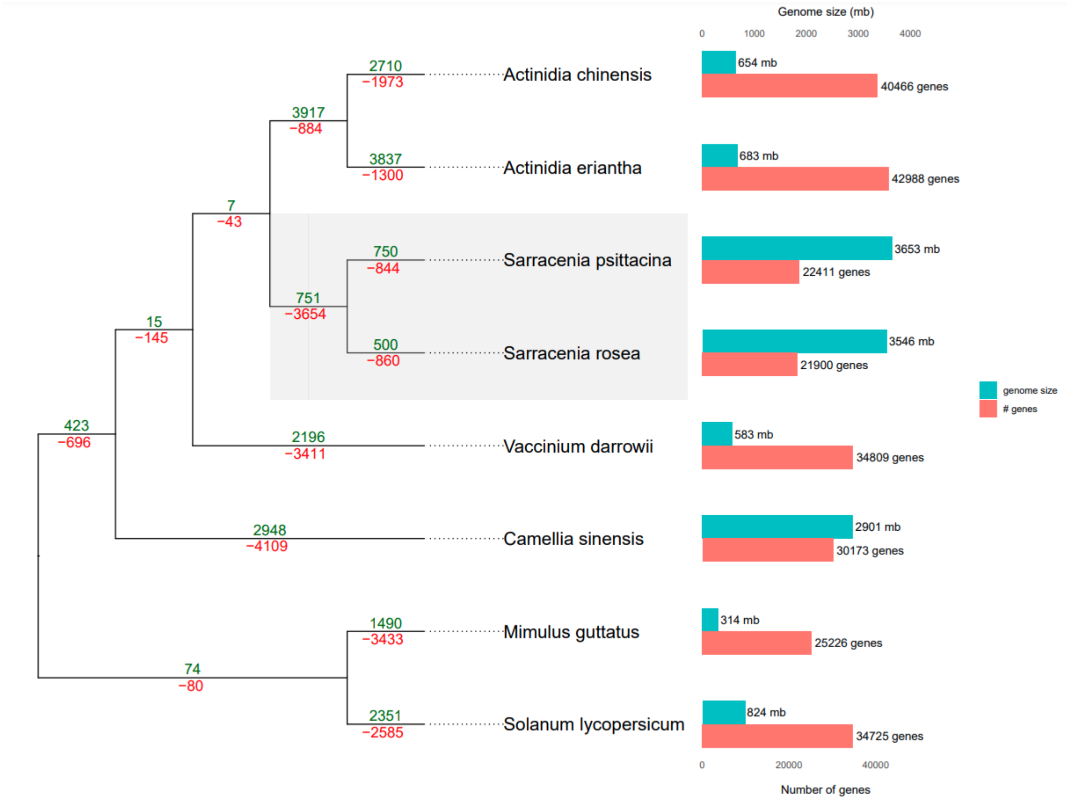


**Figure 3.2**. Summary of orthogroup expansion and contraction. Branches are labeled with the number of orthogroups that are either expanding (green, above) or contracting (red, below) on that branch. Aligned with the tips of the tree are bar plots of genome size and number of annotated gene models in each genome.

enriched for gene families that expanded on the branch leading to *Sarracenia*, 3654 contracted and 429 lost gene families were enriched in several key biological functions relating to photosynthesis (Fig. 3). The enrichment for loss of photosynthesis-related genes is primarily driven by the absence of the majority of genes involved in the NADH dehydrogenase (*Ndh*) complex assembly (Fig. 3 B,D). While *Ndh* genes are not strictly necessary for photosynthetic function, the *Ndh* complex plays a role in photosynthetic electron transport and is important for maintaining photosynthetic efficiency under certain environmental stresses (Graham et al. 2017). Many of the plastome-encoded *Ndh* subunits are absent or psuedogenized in *Sarracenia* plastomes (Baldwin et al. 2023) and other carnivorous plant plastomes (Fu et al. 2023), but this is the first case where the loss of nuclear-encoded *Ndh* complex and related genes has been shown. The loss of photosynthesis genes is specific to *Sarracenia* and not seen in the other Ericales genomes (Fig. 3E).

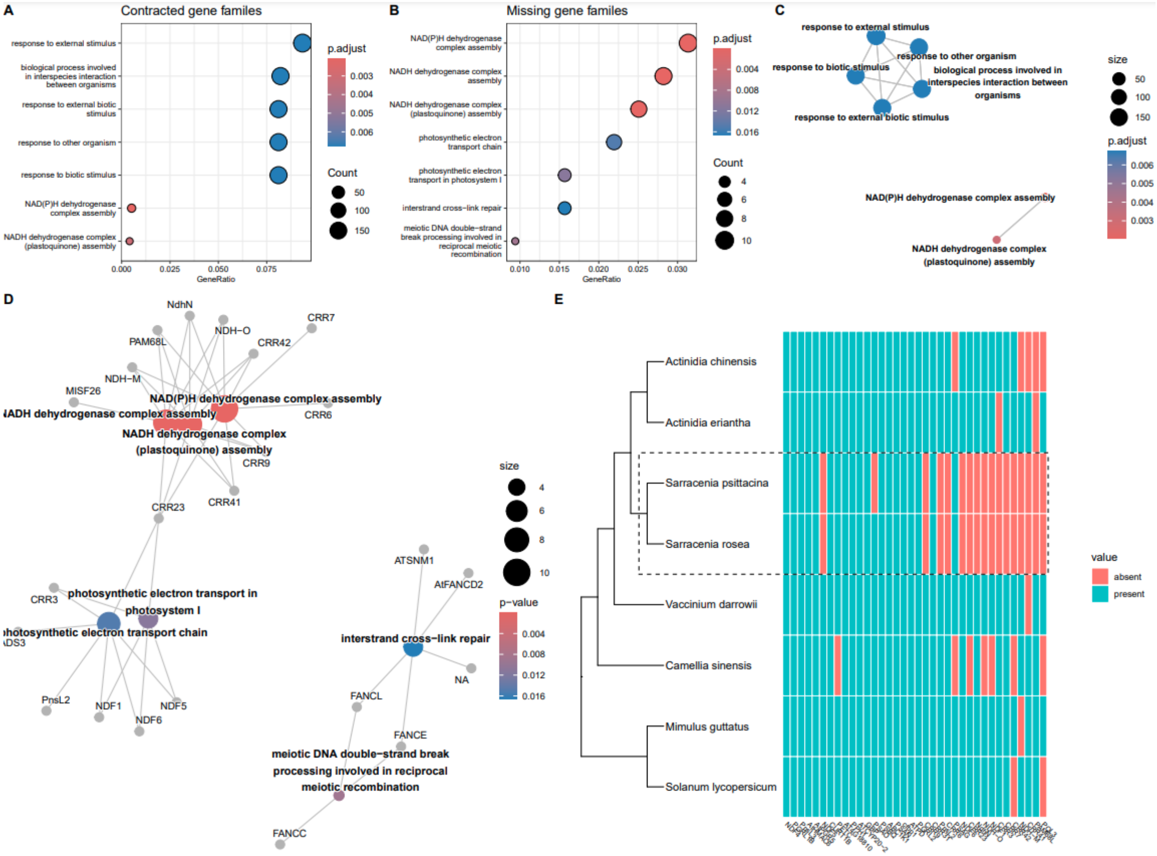


**Figure 3.3**. Functional enrichment of missing genes in *Sarracenia* genomes. (A) GO term enrichment of contracted gene families and (B) missing gene families. (C) Enrichment map of contracted gene families. GO terms are connected by edges when gene sets are overlapping. (D) Gene concept network for missing gene families. Enriched terms are hubs and the genes annotated with those terms are connected to them by edges. (E) Presence/absence of all genes in the 5 enriched photosynthesis-related GO terms.

# DISCUSSION

By leveraging a combination of state-of-the-art sequencing technologies and the trio-binning technique, we have produced chromosome-level genome assemblies from two species of the carnivorous pitcher plant genus *Sarracenia*. These are the first published genome assemblies in the Sarraceniaceae and the first long-read assemblies published for any carnivorous member of the Ericales, although a short read assembly has been published for *Roridula* (Hartmann et al. 2020). These high-quality assemblies have broadened the representation of carnivorous plant lineages with available genomic resources, and will be critical resources for identifying genomic changes associated with the evolution of carnivory.

*Sarracenia* has a large genome filled with repetitive content, but is relatively gene-poor. This architecture closely parallels that of the venus flytrap's genome, while its close carnivorous relatives exhibit less extreme genome size and repeat content but similar gene reductions (Palfalvi et al. 2020). *Utricularia gibba*, another carnivorous plant, exhibits an inverse pattern, with a more typical number of genes but one of the smallest genome sizes known in vascular plants (Ibarra-Laclette et al. 2013). Carnivory thus does not seem to have a strong impact on plant genome size, repeat content, or gene number, despite carnivorous plants containing several outliers.

Genes involved in photosynthesis—namely those that code for subunits of the *Ndh* complex—have been lost in *Sarracenia* and other carnivorous plant plastid genomes (Fu et al. 2023). For the first time, we identified the loss of *Ndh* genes in the nuclear genome of a carnivorous plant. While carnivorous plants are well-known for their ability to obtain nitrogen and phosphorous from their digested prey, the dispensability of photosynthesis-related genes could hint at the assimilation of carbon from prey as well, relaxing the need for efficient photosynthetic machinery. A few studies have indicated that carnivorous plants are obtaining some carbon from their prey (Fasbender et al. 2017), but the loss of *Ndh* genes in *Sarracenia* warrants further study on the functional impact of prey-derived carbon.

While the loss of *Ndh* genes represents the dispensability of a function, we identify the loss of immune response genes which we hypothesize is a necessary adaptation to *Sarracenia*'s carnivorous mode of nutrition. *Sarracenia*'s pitchers contain decomposing insect prey in addition to an assortment of bacterial and fungal symbionts that assist with digestion. An immune response to this constant contact with other organisms is unnecessary and likely detrimental in *Sarracenia* and other carnivorous plants, making the loss of immune response genes adaptive to the digestion of prey in *Sarracenia*.

## CONCLUSION

The reference genomes presented here will facilitate future research into the unique evolutionary adaptations in *Sarracenia.* Furthermore, these genomic resources will enable more powerful and accurate techniques for conservation geneticists working to conserve this ecologically significant group. Furthermore, by identifying the loss of genes related to photosynthesis and immune response, we provide a novel perspective into the molecular basis of carnivory in plants.

**Table 3.1**. Genome assembly statistics.

| | **S. rosea** | **S. psittacina** |
|---|---|---|
| Total assembly length | 3488 mb | 3594 mb |
| Chromosome length | 3426 mb | 3472 mb |
| Small scaffold length | 62 mb | 122 mb |
| Scaffold N50 | 280 mb | 281 mb |
| Contigs N50 | 9 mb | 7 mb |
| Number of scaffolds | 896 | 1161 |
| Number of contigs | 1505 | 1847 |
| Number of genes | 21900 | 22411 |
| Complete BUSCOs | 1564 | 1579 |
| Complete and single-copy BUSCOs | 1494 | 1451 |
| Complete and duplicated BUSCOs | 70 | 128 |
| Fragmented BUSCOs | 14 | 8 |
| Missing BUSCOs | 36 | 27 |
| Total BUSCO groups searched | 1614 | 1614 |

# CHAPTER 4

# PHYLOGENOMICS REVEALS RAPID DIVERSIFICATION AND WIDESPREAD INTROGRESSION IN *SARRACENIA*

## INTRODUCTION

*Sarracenia* is a genus of carnivorous plants in the Sarraceniaceae endemic to eastern North America, with most of its 8-11 species being distributed throughout the southeastern United States (McPherson and Schnell 2011). The common name, “pitcher plants”, references their funnel-shaped leaves which capture and digest insects. All species in the pitcher plant family Sarraceniaceae, including *Darlingtonia californica* and South American *Heliamphora* species in addition to those in *Sarracenia* are carnivorous, but Sarraceniaceae is not the only carnivorous pitcher plant lineage. Their unique and charismatic pitcher plant form has led to widespread horticultural interest, with numerous species and hybrids being available in the horticultural trade.

Hybridization is common in *Sarracenia*—a hybrid has been found in the wild for almost every sympatric pair of species (Bell 1952). Although *Sarracenia* has high potential for interspecific gene flow, the extent of introgression among Sarracenia species remains largely unexplored. Two studies exist that examine gene flow between *Sarracenia* species at specific sites using microsatellite loci (Furches et al. 2013;Rentsch and Holland 2020). Furches et al. (2013) found that interspecific gene flow between *S. leucophylla, S. alabamensis ssp. wherryi,* and *S. alata* at their study location was closely tied to geographic proximity. Rentsch and Holland (2020) did not detect gene flow between *S. minor* and *S. flava*. Both studies are limited by a small geographic focus and the use of only eight microsatellite loci. The plastome analysis in Chapter 2 revealed that gene flow is occurring across the genus, but it is difficult to determine the amount of gene flow and which species are involved since the plastome acts as a single locus. A genus-wide investigation of nuclear gene flow is therefore necessary to better understand the extent and directionality of interspecific genetic exchange.

Molecular evidence suggest that *Sarracenia* experienced a recent and rapid radiation, with most of its diversification occurring in the past 1-3 million years (Ellison et al. 2012;Stephens et al. 2015). Phylogenetic relationships within rapid radiations are notoriously difficult to resolve due to high amounts of incomplete lineage sorting (ILS) which results in gene tree discordance (Whitfield and Lockhart 2007). Hybridization and gene flow provide an additional source of gene tree heterogeneity. *Sarracenia*'s rapid radiation and potential for gene flow provide dual challenges for estimating accurate phylogenies. In the face of these challenges, Stephens et al. (2015) succeeded in producing a phylogeny with high support for several key nodes by using 199 nuclear loci. However, there remain many relationships with tenuous support despite thorough sampling.

The goals of this study are to (1) construct an accurate species-level phylogeny for *Sarracenia* in the presence of high levels of ILS and (2) explore the landscape of interspecific nuclear gene flow in the genus. We use a phylogenomic approach for both goals, leveraging the reference genomes sequenced in chapter 3 to identify every single-copy gene in the genome that will use to estimate a phylogeny and analyze gene flow.

This large set of genes will provide a complete genomic picture of diversification in Sarracenia in the presence of ILS and gene flow.

## METHODS

### Taxon sampling

Silica-dried tissue for most *Sarracenia* samples were obtained from the Atlanta Botanical Garden’s Conservation DNA Biorepository. *Sarracenia minor* was collected from private property in Nassau County, Florida. Additionally, *S. psittacina*, *S. flava*, and *S. rosea* were obtained from a greenhouse at UGA. Sequencing reads from two *Heliamphora* species were obtained from SRA (*H. ciliata*: SRR24877724, *H. pulchella*: SRR25244091) and used as outgroups.

### DNA isolation and sequencing

DNA was isolated from fresh or dry tissue using Qiagen DNeasy Plant Pro kits. Libraries were constructed with Kapa HyperPlus library kits using custom adapters and iTru primers from Adapterama I (Glenn et al. 2019). Sequencing was done on a 10B flow cell on a NovaSeq X at SeqCenter in Pittsburgh, PA.

### ***Sarracenia* phylogeny**

We used orthogroups identified by OrthoFinder as conserved in single copy across the Ericales (chapter 3) for phylogenomic analysis. Genes were assembled from short reads using the HybPiper v. 2.1.6 (Johnson et al. 2016). After assembly, “supercontig” sequences, which include coding sequences and any intron or flanking non-coding sequences that were assembled, were used for further analysis. For each single copy locus, sequences from all species were aligned using MAFFT v.7.505 (Katoh and Standley 2013), and alignments were trimmed using trimal v.1.4.1 (Capella-Gutiérrez et al. 2009). Maximum-likelihood (ML) trees were estimated for each orthogroup with IQ-TREE v.2.2.6 (Nguyen et al. 2015) using the ModelFinder option to identify the optimal substitution model (Kalyaanamoorthy et al. 2017) and 1000 ultrafast bootstrap replicates (Minh et al. 2013). A species tree was estimated with the multi-species coalescence approach implemented in ASTRAL-IV v1.16 (Zhang and Mirarab 2022), using 3189 gene trees with nodes collapsed if they had bootstrap scores below 10. To evaluate the fit of a fully bifurcating species tree topology, a polytomy test as implemented in ASTRAL3 was run (Zhang et al. 2018). In addition to the ASTRAL species tree, a maximum-likelihood (ML) partitioned analysis tree was estimated with IQ-TREE using the ModelFinder option and 1000 ultrafast bootstraps.

### Phylogenetic network estimation

To estimate a phylogenetic network, SNaQ (Solís-Lemus and Ané 2016) was used as implemented in the Julia package PhyloNetworks (Solís-Lemus et al. 2017). Concordance factors were calculated from the gene trees using the function countquartetsintrees(). We estimated networks with the maximum number of hybridization edges (h) between 1-5. The ASTRAL species tree (h=0) was used as a starting tree for the h=1 network estimation, and we used the best network from the h-1 runs as a starting network for the rest of the runs. The optimal h value was determined by the point where the pseudo-likelihood score stopped improving drastically.

### Detecting gene flow using gene tree heterogeneity

In addition to the network analysis, the possibility of introgression was also assessed by examining patterns of gene tree discordance using a quartet-based approach. Given any quartet of tips from a species tree, there are three possible unrooted trees: one that is concordant with the species tree, and two that are discordant with the species tree. Under a null coalescent model (i.e. with only ILS contributing to gene tree discordance), the two discordant topologies will be produced in equal proportions. Introgression between two of the taxa in the quartet will cause one of the discordant topologies to exceed the frequency of the other. Using a framework based on D-statistics (Green et al. 2010), one can test for introgression between any two taxa by examining the frequencies of discordant topologies of certain quartets that include those two taxa.

Here, we only consider quartets with one of the *Heliamphora* sequences as an outgroup tip in order to root the relationships of the three *Sarracenia* sequences and assess the pattern of discordance with the species tree. Therefore, the quartets consisted of the two taxa we are testing for introgression between, the outgroup, and an additional taxon that is more closely related to one of the test taxa. As such, this approach does not test for introgression between sister taxa. In most other cases, there is more than one quartet that can be used to test for introgression for a given taxon pair. Here, we test all quartets and report, the average of a D-like statistic (ABBA-BABA/ABBA+BABA, referred to as D here for simplicity) for each taxon pair. For each quartet, a chi squared test is used to test if the proportion of either discordant topology (ABBA or BABA) is significantly greater than the other. P values for all tests are adjusted using false discovery rate. This method is being compiled into an R package (https://github.com/ethan-baldwin/quaint).

## RESULTS

### Highly resolved *Sarracenia* species phylogeny

We identified 3189 conserved single copy ortholog sets across the *Sarracenia* and other Ericales genomes (chapter 3) to use for phylogenomic analysis. Fifteen *Sarracenia* taxa were sequenced to a depth of at least 10x, and at least 95% of genes were recovered for every sample. Despite high levels of incomplete lineage sorting (as evidenced by significant gene tree-species tree discordance (see pie charts on node in Fig. 1 tree)), the ASTRAL species tree is highly resolved, with posterior probabilities of over 0.99 at all but two nodes. The polytomy null hypothesis could not be rejected at two adjacent nodes in the fully bifurcating species tree estimation, resulting in a polytomy with four daughter lineages (Fig. 1). Notably, the polytomy occurs within the *rubra* complex clade (i.e. *rubra*, *alabamensis*, and *jonesii*), a group of closely related taxa where taxonomic changes have been rampant and species delimitation is tenuous (Case and Case 1976).

Consistent with the most recent *Sarracenia* phylogeny (Stephens et al. 2015), two major clades were recovered: the "*purpurea* clade", consisting of *purpurea, rosea*, *psittacina*, and *flava*, and the "*oreophila* clade", consisting of the *rubra* complex, *oreophila, leucophylla,* and *alata*. Within the *oreophila* clade, *oreophila* and *leucophylla* are successive sister lineages to the polytomy including *rubra ssp. gulfensis, alata, alabamensis,* and a clade containing *rubra ssp. rubra* and *jonesii*. The *purpurea* clade retains the same relationships as Stephens et al. (2015), with *purpurea* (including *rosea*) being sister to a clade that contains *minor, flava*, and *psittacina*.

The partitioned ML analysis was done on concatenated matrix including the same 3189 gene alignments used for gene tree and species tree estimation in the ASTRAL analysis. In total, the alignment had 7,228,445 columns, 475,845 of which were parsimony informative. The resulting ML tree is highly supported, with all nodes having a bootstrap score of 100 except for one which was 88. The ML tree had significant topological differences in comparison with the ASTRAL tree (Fig. 1). Within the

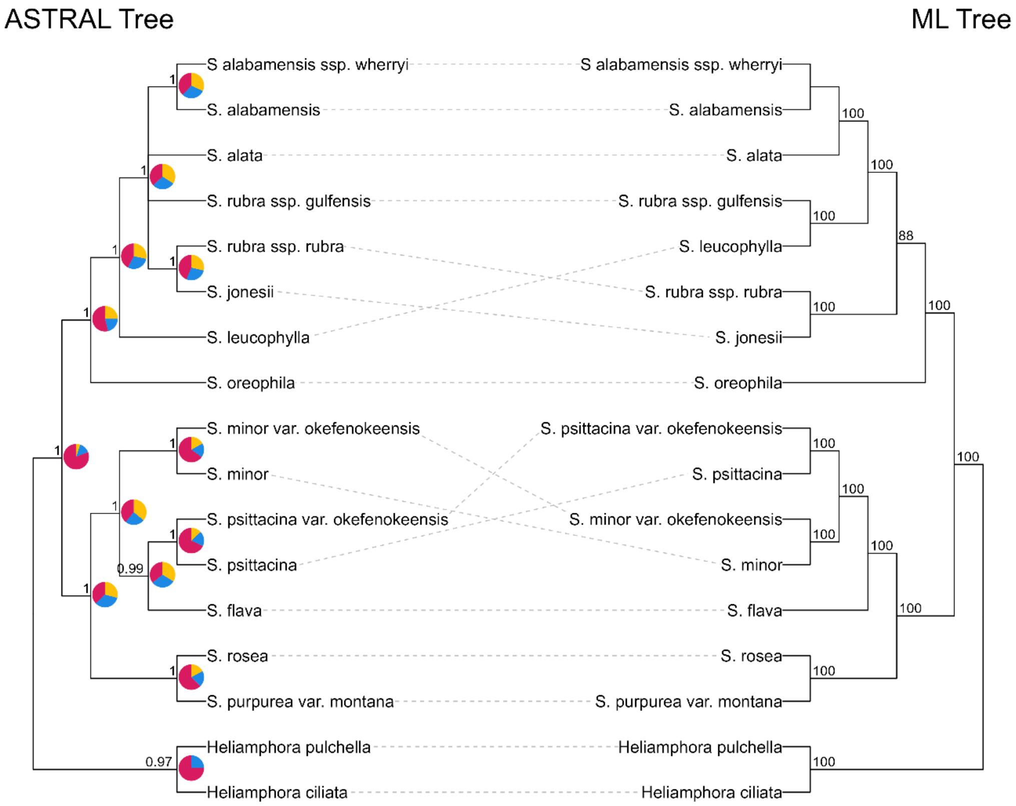


**Figure 4.1**. ASTRAL tree (left) and concatenated ML tree (right). Node support values are posterior probabilities on the ASTRAL tree and bootstrap score on the ML tree. The pie charts on the nodes in the ASTRAL tree show the proportion of gene trees that match the topology in the tree (red) or are one of the two alternate quartet topologies at that node (yellow and blue).

*purpurea* clade, *minor* and *psittacina* are sister to one another in the ML tree while *flava* and *psittacina* are sister to one another in the ASTRAL tree. The relationships within the *oreophila* clade have more drastic differences between the two trees, with the polytomy in the ASTRAL tree fully resolved in the ML tree and *leucophylla* placed sister to *rubra ssp. gulfensis* in the ML tree. While the ML tree has higher support values and does not include polytomies, the assumption that all genes share the same underlying topology is not supported by the individual gene tree estimates, so we treat the ASTRAL tree as a more likely estimation of *Sarracenia* species relationships and use it as the backbone tree for gene tree analyses. Rapid diversification and rapid ILS (as seen the *Sarracenia* species tree / gene tree discordance levels show in the lefthand tree of Figure 1), can cause ML analyses of concatenated gene sequence alignments to produce the wrong trees with high support (Warnow 2015). The multispecies coalescent analyses implemented in ASTRAL and other coalescence analysis programs account for ILS and are thus statistically consistent in scenarios with high ILS (Degnan and Rosenberg 2009).

### Gene flow in *Sarracenia*

The optimal number of hybridization edges in the phylogenetic networks estimated by SNaQ was two (-log pseudolikelihood = 130.9997). The two hybrid edges in the h=2 network are between *rosea* and the branch leading to *flava* and *psittacina* and between *rubra ssp. gulfensis* and a branch leading to *alata* and *alabamensis* (Fig. 2). Interestingly*, rubra ssp. gulfensis* is sister to *leucophylla* in the network instead of in a clade with *rubra*, jonesii, and *alabamensis*. This is similar to the topology found in the ML tree (Fig. 1), where *rubra ssp. gulfensis* and *leucophylla* are sister taxa, but are both placed within the *rubra* complex rather than sister to it.

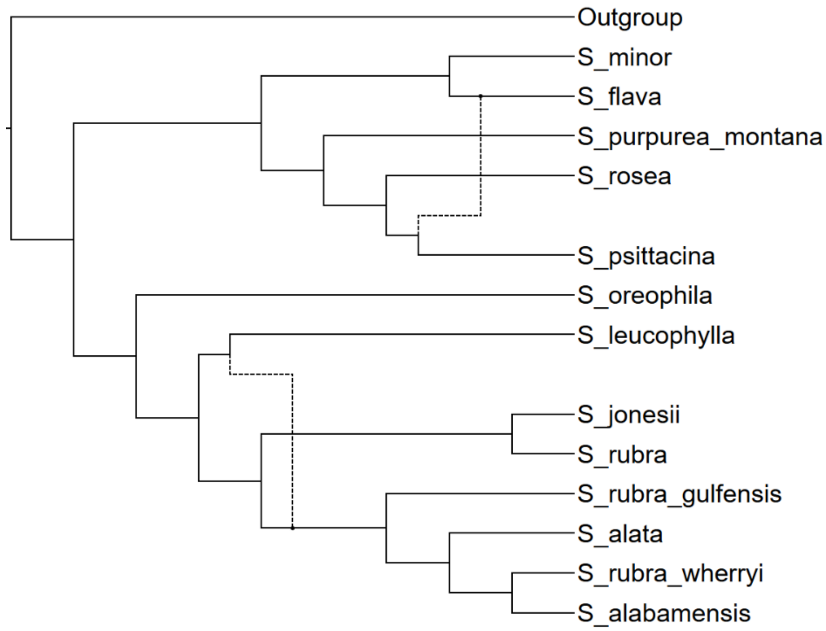


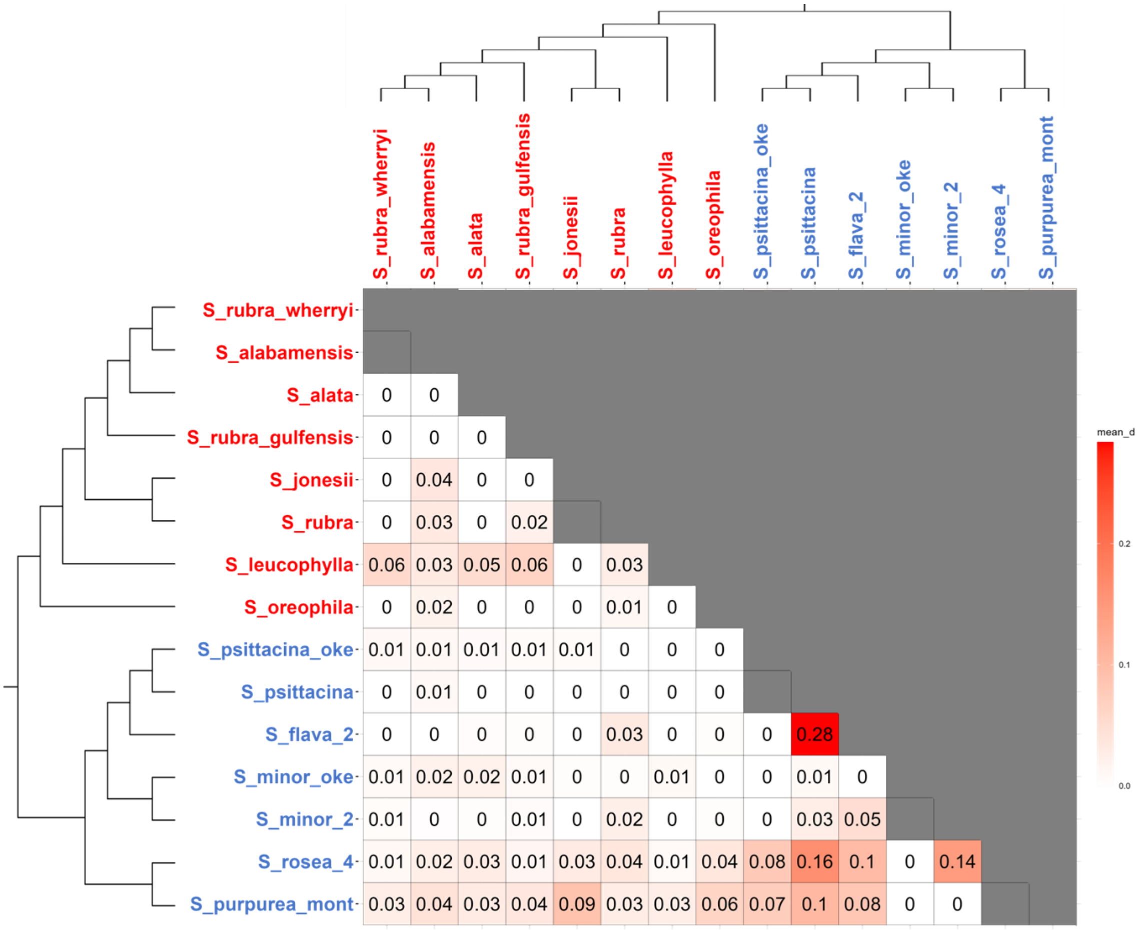
**Figure 4.2**. Best network from the SNaQ phylogenetic network estimation.

The quartet analysis revealed some amount of support for gene exchange between more than half of the tested species pairs (Table 1). The species pair with the strongest support for gene flow was *flava* and *psittacina,* with an average D of 0.26 (Fig. 3; D=0 expected in the absence of gene flow). Due to the topology of the species tree, gene flow between these species can only be tested using one quartet (Outgroup, *psittacina, psittacina var. okefenokeensis, flava*).Morphological hybrids between these two species are rare (Bell 1952) and so it is surprising that the strongest signal of gene flow came from this species pair, and it may be due to sample-specific histories.

This analysis also indicates that *rosea* has exchanged genes with both *psittacina* samples and with *flava*. This is congruent with the phylogenetic network (Fig. 2), although with the quartet analysis it is difficult to determine if this gene flow occurred before the divergence of *flava* and *psittacina* or if it represents parallel gene flow since divergence. Introgression was also detected between *rosea* and *minor*. With the exception of *psittacina var. okefenokeensis*, all of these species occur in sympatry with *rosea*. *S. purpurea var. montana* is closely related to *rosea*, and is considered the same species in some taxonomic treatments. This analysis also indicates gene flow between *purpurea var. montana* and both *psittacina* and *flava*, however *purpurea var. montana* is not sympatric with these taxa.

*S. leucophylla* shows gene flow between it and most of the *oreophila* clade. The strongest signal of gene flow was detected between it and *rubra ssp. gulfensis*, where each of the ten quartet tests between them were significant (Table 1). Inferred geneflow between *leucophylla* and *rubra ssp. gulfensis* is notable as these taxa were placed together in the ML tree and the phylogenetic network (Fig. 1,2). *S. leucophylla* is sympatric with *rubra ssp. gulfensis* and the other taxa it has evidence for gene flow with, except *alabamensis*.

The plastome analysis in chapter 2 indicated potential gene flow between the *Sarracenia* taxa that are endemic to the southern Appalachians: *oreophila, jonesii,* and *purpurea var. montana*. Here, we infer gene flow between *purpurea var. montana* and both *jonesii* and *oreophila*, but not between *oreophila* and *jonesii*.



| | S_rubra_wherryi | S_alabamensis | S_alata | S_rubra_gulfensis | S_jonesii | S_rubra | S_leucophylla | S_oreophila | S_psittacina_oke | S_psittacina | S_flava_2 | S_minor_oke | S_minor_2 | S_rosea_4 | S_purpurea_mont |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S_rubra_wherryi | | | | | | | | | | | | | | | |
| S_alabamensis | | | | | | | | | | | | | | | |
| S_alata | 0 | 0 | | | | | | | | | | | | | |
| S_rubra_gulfensis | 0 | 0 | 0 | | | | | | | | | | | | |
| S_jonesii | 0 | 0.04 | 0 | 0 | | | | | | | | | | | |
| S_rubra | 0 | 0.03 | 0 | 0.02 | | | | | | | | | | | |
| S_leucophylla | 0.06 | 0.03 | 0.05 | 0.06 | 0 | 0.03 | | | | | | | | | |
| S_oreophila | 0 | 0.02 | 0 | 0 | 0 | 0.01 | 0 | | | | | | | | |
| S_psittacina_oke | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | | | | | | | |
| S_psittacina | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | |
| S_flava_2 | 0 | 0 | 0 | 0 | 0 | 0.03 | 0 | 0 | 0 | 0.28 | | | | | |
| S_minor_oke | 0.01 | 0.02 | 0.02 | 0.01 | 0 | 0 | 0.01 | 0 | 0 | 0.01 | 0 | | | | |
| S_minor_2 | 0.01 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0 | 0.03 | 0.05 | | | | |
| S_rosea_4 | 0.01 | 0.02 | 0.03 | 0.01 | 0.03 | 0.04 | 0.01 | 0.04 | 0.08 | 0.16 | 0.1 | 0 | 0.14 | | |
| S_purpurea_mont | 0.03 | 0.04 | 0.03 | 0.04 | 0.09 | 0.03 | 0.03 | 0.06 | 0.07 | 0.1 | 0.08 | 0 | 0 | | |

**Figure 4.3**. Results of the quartet-based analysis of gene flow.

## DISCUSSION

### Relationships in *Sarracenia*

Stephens et al. (2015) resolved many of the relationships within Sarracenia using 199 nuclear loci across 75 samples, however, many of the relationships within the *oreophila* clade were not highly supported. Here, using significantly more loci (3189) but fewer samples (18), we present a phylogeny with high support (posterior probability >= 0.99) for all but two nodes in the *oreophila* clade. Nevertheless, many important relationships within the taxonomically unstable *rubra* complex were uncovered. Our samples for both *alabamensis* subspecies are sister to one another in every analysis, supporting the adoption of the *wherryi* taxon as a subspecies of *alabamensis*. Consistent with Stephens et al. (2015), *rubra ssp. rubra* and *jonesii* are sister to one another. *S. rubra ssp. gulfensis*, on the other hand, is in a polytomy with members of the *rubra* complex and determining its true relationship may be complicated by hybridization with *leucophylla*. *S. alata* is also part of this polytomy, although its inclusion in the clade with all of the *rubra* complex members is highly supported. It is clear that the *rubra* complex as currently circumscribed is not monophyletic, and future taxonomic treatments should be aware of the close relationship between *alata* and its members.

With a set of >3000 loci comprised of all of the single copy genes in the genome, we would expect to recover accurate species relationships for most species. However, high levels of gene tree heterogeneity in this data set resulted in inference of a polytomy in the ASTRAL species tree estimate. Rather than interpreting the polytomy as simply unresolved bifurcations resulting from too little data (a soft polytomy), our analyses support the inference of a true polytomy with one ancestral species spawning four daughter lineages. Specifically, we infer the most common ancestor for *alata, alabamensis, rubra,* and *jonesii* as a single ancestral species/population. This scenario could explain why various authors have proposed widely varying taxonomic treatment of these taxa (Case and Case 1976).

### Gene flow in *Sarracenia*

Our phylogenetic network and quartet-based analyses show that gene flow has significantly contributed to the complex evolutionary history of *Sarracenia*. While the SNaQ network implicated only two hybrid edges, the quartet analysis revealed a much more complex landscape of reticulation. While SNaQ is able to estimate phylogenetic networks with simple reticulations, more complex reticulations like those involving hybrid edges are not resolvable (Solís-Lemus and Ané 2016). Nonetheless, the network topology estimated by SNaQ was more or less corroborated by the quartet analysis.

Hybridization is frequent between sympatric species in *Sarracenia* and our data suggest extensive gene flow results from this. Nevertheless, species seem to remain phenotypically isolated from one another. Because of this, we propose that *Sarracenia* should be considered a syngameon (Buck and Flores-Rentería 2022), where ongoing and historical gene flow result in a loosely connected network of distinct taxa. Future studies of *Sarracenia* should consider how introgression could affect the evolution of traits. This also has implications for the conservation of imperiled taxa in *Sarracenia*. Conservation practitioners should be aware that hybridization has been an integral factor in *Sarracenia* evolution. It seems that in their evolutionary history, *Sarracenia* species have exchange genes while maintaining phenotypic differences. On the other hand, hybridization and genetic swamping, a process wherein one or both parental populations in a hybrid zone are replaced by hybrids, may occur in the future if rates of hybridization increase. Under such a scenario hybridization may result in population decline and even extirpation (Todesco et al. 2016).

Although we observe clear evidence of gene flow, the genomic distribution of introgressed loci and timing of gene flow events is yet to be uncovered. Similarly, quantifying the strength and directionality of gene flow requires further exploration, as the quartet-based analysis we present here functions primarily to detect gene flow rather than quantify it. The inclusion of more samples per species would open up more powerful techniques for elucidating these aspects of introgression. For example, F-statistics

(Patterson et al. 2012) and their derivatives, e.g. F-branch (Malinsky et al. 2018), are able to estimate the timing and magnitude of introgression, but both approaches require at least two phased haplotypes per taxon. Moreover, sampling multiple genotypes within a taxon will allow a broader interpretation of gene flow in the context of evolutionary time by lessening the likelihood of population-specific gene flow dominating the signal.

Taken together, our phylogenomic approach has confirmed inference drawn from whole plastome analyses (see chapter 2) that hybridization and gene flow occur frequently in *Sarracenia*, potentially influencing the evolutionary trajectory of populations within the genus. These findings underscore the need to consider reticulate evolution when reconstructing species relationships in rapidly radiating plant lineages and have important implications for both taxonomy and conservation in *Sarracenia*.

**Table 4.1**. Quartet-based test of gene flow results.

| Taxon 1 | Taxon 2 | Number of tests | Number of positive tests | Number of positive significant tests | Average D | Combined D |
|---|---|---|---|---|---|---|
| S_alabamensis | S_alata_m003 | 2 | 2 | 0 | 0.000 | 0.031 |
| S_alabamensis | S_rubra_gulfensis | 4 | 4 | 0 | 0.000 | 0.028 |
| S_alabamensis | S_rubra_RVCAGA001 | 8 | 8 | 3 | 0.032 | 0.056 |
| S_alabamensis | S_jonesii | 8 | 6 | 4 | 0.036 | 0.046 |
| S_alabamensis | S_leucophylla | 10 | 5 | 3 | 0.033 | 0.026 |
| S_alabamensis | S_oreophila_m002 | 12 | 10 | 3 | 0.018 | 0.041 |
| S_alabamensis | S_purpurea_montana | 26 | 19 | 10 | 0.043 | 0.052 |
| S_alabamensis | S_rosea_4 | 26 | 16 | 6 | 0.024 | 0.022 |
| S_alabamensis | S_minor_2 | 26 | 16 | 1 | 0.003 | 0.004 |
| S_alabamensis | S_minor_okefenokeensis | 26 | 18 | 6 | 0.020 | 0.033 |
| S_alabamensis | S_flava_2 | 26 | 3 | 0 | 0.000 | -0.054 |
| S_alabamensis | S_psittacina_P008 | 26 | 6 | 2 | 0.010 | -0.046 |
| S_alabamensis | S_psittacina_okefenokeensis | 26 | 17 | 2 | 0.011 | -0.005 |
| S_alata_m003 | S_rubra_wherryi | 2 | 0 | 0 | 0.000 | -0.031 |
| S_rubra_gulfensis | S_rubra_wherryi | 4 | 1 | 0 | 0.000 | -0.008 |
| S_rubra_RVCAGA001 | S_rubra_wherryi | 8 | 1 | 0 | 0.000 | -0.043 |
| S_jonesii | S_rubra_wherryi | 8 | 4 | 0 | 0.000 | -0.011 |
| S_leucophylla | S_rubra_wherryi | 10 | 8 | 4 | 0.056 | 0.066 |
| S_oreophila_m002 | S_rubra_wherryi | 12 | 4 | 0 | 0.000 | -0.014 |
| S_purpurea_montana | S_rubra_wherryi | 26 | 15 | 6 | 0.026 | 0.023 |
| S_rosea_4 | S_rubra_wherryi | 26 | 10 | 3 | 0.011 | -0.008 |
| S_minor_2 | S_rubra_wherryi | 26 | 12 | 2 | 0.005 | -0.002 |
| S_minor_okefenokeensis | S_rubra_wherryi | 26 | 9 | 3 | 0.008 | -0.003 |
| S_flava_2 | S_rubra_wherryi | 26 | 1 | 0 | 0.000 | -0.055 |
| S_psittacina_P008 | S_rubra_wherryi | 26 | 0 | 0 | 0.000 | -0.060 |
| S_psittacina_okefenokeensis | S_rubra_wherryi | 26 | 9 | 2 | 0.010 | -0.012 |
| S_alata_m003 | S_rubra_gulfensis | 4 | 1 | 0 | 0.000 | -0.019 |
| S_alata_m003 | S_rubra_RVCAGA001 | 8 | 5 | 0 | 0.000 | 0.007 |
| S_alata_m003 | S_jonesii | 8 | 1 | 0 | 0.000 | -0.028 |
| S_alata_m003 | S_leucophylla | 10 | 5 | 4 | 0.052 | 0.028 |
| S_alata_m003 | S_oreophila_m002 | 12 | 10 | 0 | 0.000 | 0.025 |
| S_alata_m003 | S_purpurea_montana | 26 | 19 | 8 | 0.030 | 0.037 |
| S_alata_m003 | S_rosea_4 | 26 | 18 | 8 | 0.028 | 0.035 |
| S_alata_m003 | S_minor_2 | 26 | 14 | 1 | 0.003 | -0.010 |
| S_alata_m003 | S_minor_okefenokeensis | 26 | 19 | 6 | 0.020 | 0.026 |
| S_alata_m003 | S_flava_2 | 26 | 11 | 1 | 0.005 | -0.022 |
| S_alata_m003 | S_psittacina_P008 | 26 | 9 | 0 | 0.000 | -0.039 |
| S_alata_m003 | S_psittacina_okefenokeensis | 26 | 13 | 1 | 0.007 | -0.013 |
| S_rubra_gulfensis | S_rubra_RVCAGA001 | 8 | 5 | 2 | 0.022 | 0.009 |

| | | | | | | |
|---|---|---|---|---|---|---|
| S_jonesii | S_rubra_gulfensis | 8 | 2 | 0 | 0.000 | -0.037 |
| S_leucophylla | S_rubra_gulfensis | 10 | 10 | 5 | 0.064 | 0.081 |
| S_oreophila_m002 | S_rubra_gulfensis | 12 | 2 | 0 | 0.000 | -0.031 |
| S_purpurea_montana | S_rubra_gulfensis | 26 | 17 | 9 | 0.035 | 0.040 |
| S_rosea_4 | S_rubra_gulfensis | 26 | 14 | 4 | 0.014 | 0.025 |
| S_minor_2 | S_rubra_gulfensis | 26 | 11 | 2 | 0.006 | -0.012 |
| S_minor_okefenokeensis | S_rubra_gulfensis | 26 | 10 | 3 | 0.008 | 0.000 |
| S_flava_2 | S_rubra_gulfensis | 26 | 9 | 1 | 0.004 | -0.023 |
| S_psittacina_P008 | S_rubra_gulfensis | 26 | 3 | 0 | 0.000 | -0.052 |
| S_psittacina_okefenokeensis | S_rubra_gulfensis | 26 | 6 | 1 | 0.006 | -0.027 |
| S_leucophylla | S_rubra_RVCAGA001 | 10 | 2 | 1 | 0.026 | -0.078 |
| S_oreophila_m002 | S_rubra_RVCAGA001 | 12 | 8 | 2 | 0.012 | 0.017 |
| S_purpurea_montana | S_rubra_RVCAGA001 | 26 | 18 | 7 | 0.027 | 0.031 |
| S_rosea_4 | S_rubra_RVCAGA001 | 26 | 23 | 10 | 0.036 | 0.058 |
| S_minor_2 | S_rubra_RVCAGA001 | 26 | 23 | 7 | 0.024 | 0.038 |
| S_minor_okefenokeensis | S_rubra_RVCAGA001 | 26 | 11 | 1 | 0.003 | -0.010 |
| S_flava_2 | S_rubra_RVCAGA001 | 26 | 19 | 7 | 0.032 | 0.022 |
| S_psittacina_P008 | S_rubra_RVCAGA001 | 26 | 14 | 0 | 0.000 | -0.029 |
| S_psittacina_okefenokeensis | S_rubra_RVCAGA001 | 26 | 4 | 0 | 0.000 | -0.048 |
| S_jonesii | S_leucophylla | 10 | 0 | 0 | 0.000 | -0.134 |
| S_jonesii | S_oreophila_m002 | 12 | 7 | 0 | 0.000 | 0.002 |
| S_jonesii | S_purpurea_montana | 26 | 25 | 21 | 0.089 | 0.101 |
| S_jonesii | S_rosea_4 | 26 | 20 | 8 | 0.029 | 0.035 |
| S_jonesii | S_minor_2 | 26 | 8 | 0 | 0.000 | -0.025 |
| S_jonesii | S_minor_okefenokeensis | 26 | 16 | 1 | 0.003 | 0.000 |
| S_flava_2 | S_jonesii | 26 | 10 | 0 | 0.000 | -0.021 |
| S_jonesii | S_psittacina_P008 | 26 | 9 | 0 | 0.000 | -0.038 |
| S_jonesii | S_psittacina_okefenokeensis | 26 | 9 | 1 | 0.006 | -0.022 |
| S_leucophylla | S_oreophila_m002 | 12 | 1 | 0 | 0.000 | -0.036 |
| S_leucophylla | S_purpurea_montana | 26 | 12 | 6 | 0.026 | -0.004 |
| S_leucophylla | S_rosea_4 | 26 | 10 | 2 | 0.007 | -0.014 |
| S_leucophylla | S_minor_2 | 26 | 8 | 0 | 0.000 | -0.018 |
| S_leucophylla | S_minor_okefenokeensis | 26 | 17 | 3 | 0.009 | 0.014 |
| S_flava_2 | S_leucophylla | 26 | 18 | 0 | 0.000 | 0.006 |
| S_leucophylla | S_psittacina_P008 | 26 | 4 | 0 | 0.000 | -0.032 |
| S_leucophylla | S_psittacina_okefenokeensis | 26 | 11 | 0 | 0.000 | -0.015 |
| S_oreophila_m002 | S_purpurea_montana | 26 | 25 | 15 | 0.062 | 0.085 |
| S_oreophila_m002 | S_rosea_4 | 26 | 22 | 9 | 0.036 | 0.052 |
| S_minor_2 | S_oreophila_m002 | 26 | 9 | 1 | 0.004 | -0.014 |
| S_minor_okefenokeensis | S_oreophila_m002 | 26 | 13 | 0 | 0.000 | -0.009 |
| S_flava_2 | S_oreophila_m002 | 26 | 14 | 1 | 0.003 | -0.012 |
| S_oreophila_m002 | S_psittacina_P008 | 26 | 13 | 0 | 0.000 | -0.025 |
| S_oreophila_m002 | S_psittacina_okefenokeensis | 26 | 7 | 0 | 0.000 | -0.031 |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| S_minor_2 | S_purpurea_montana | 10 | 1 | 0 | 0.000 | -0.148 |
| S_minor_okefenokeensis | S_purpurea_montana | 10 | 2 | 0 | 0.000 | -0.132 |
| S_flava_2 | S_purpurea_montana | 10 | 7 | 4 | 0.084 | 0.062 |
| S_psittacina_P008 | S_purpurea_montana | 10 | 7 | 5 | 0.103 | 0.078 |
| S_psittacina_okefenokeensis | S_purpurea_montana | 10 | 4 | 4 | 0.074 | 0.064 |
| S_minor_2 | S_rosea_4 | 10 | 4 | 4 | 0.144 | -0.089 |
| S_minor_okefenokeensis | S_rosea_4 | 10 | 1 | 0 | 0.000 | -0.185 |
| S_flava_2 | S_rosea_4 | 10 | 8 | 6 | 0.100 | 0.097 |
| S_psittacina_P008 | S_rosea_4 | 10 | 10 | 8 | 0.162 | 0.159 |
| S_psittacina_okefenokeensis | S_rosea_4 | 10 | 6 | 4 | 0.080 | 0.064 |
| S_flava_2 | S_minor_2 | 6 | 4 | 2 | 0.050 | 0.029 |
| S_minor_2 | S_psittacina_P008 | 6 | 5 | 1 | 0.026 | 0.033 |
| S_minor_2 | S_psittacina_okefenokeensis | 6 | 1 | 0 | 0.000 | -0.035 |
| S_flava_2 | S_minor_okefenokeensis | 6 | 0 | 0 | 0.000 | -0.064 |
| S_minor_okefenokeensis | S_psittacina_P008 | 6 | 3 | 1 | 0.009 | 0.029 |
| S_minor_okefenokeensis | S_psittacina_okefenokeensis | 6 | 4 | 0 | 0.000 | 0.019 |
| S_flava_2 | S_psittacina_P008 | 2 | 2 | 2 | 0.284 | 0.285 |
| S_flava_2 | S_psittacina_okefenokeensis | 2 | 0 | 0 | 0.000 | -0.285 |

# CHAPTER 5

# GENETIC ARCHITECTURE OF PREY-CAPTURE TRAITS AND THEIR ROLE IN HYBRID FITNESS

## INTRODUCTION

Species can be defined as independently evolving populations, which often have more or less discrete phenotypic differences that are used to classify them as such  (De Queiroz 2007). When closely related species occur in sympatry, their evolutionary independence may be maintained by reproductive barriers. Barriers to reproduction can prevent hybrid zygotes from forming (pre-zygotic) or reaching sexual maturity to mate with the parent populations (post-zygotic). While pre-zygotic barriers are thought to be more effective in reducing or ceasing gene exchange (Martin and Willis 2007) than post-zygotic barriers, the persistence of discrete phenotypic differences between species that readily hybridize underscores the importance of post-zygotic barriers in speciation processes (Buck and Flores-Rentería 2022).

Hybrid unfitness is a critical component to the maintenance of species without strong pre-zygotic barriers since unfit hybrids are less likely to mate. Reduced fitness in hybrids can be a consequence of intrinsic hybrid unfitness due to mismatch gene interactions resulting in decreased fitness regardless of environment. Additionally, poor performance in parental environments can result from maladaptive trait combinations. For example, trait mismatch can cause unfitness when a hybrid inherits the phenotype of one parent for one trait and the other parent for another trait, especially if there is some functional interaction between them (Chhina et al. 2022). Additionally, intermediate phenotypes for functional traits can cause unfitness  (Thompson et al. 2021).

Carnivorous pitcher plants in the genus *Sarracenia* have tube-shaped leaves that attract, capture, and digest insect prey. Hybrids between almost all sympatric pairs of species have been reported in the wild and all hybrids are fertile (Bell 1952). Rampant chloroplast capture across the genus has been demonstrated (Baldwin et al. 2023), and nuclear gene flow has also been detected in some sympatric populations (Furches et al. 2013). Despite the apparent ease of exchanging genetic material across species boundaries, phenotypically distinct species persist in sympatry. One mechanism for the maintenance of species in the face of potentially homogenizing gene flow in *Sarracenia i*s premating isolation due to differences in pollinators or flower phenologies (Bell 1952). However, sympatric species pairs with large windows of overlapping flowering times are common (Furches et al. 2013), and *Sarracenia* species share pollinators (Folkerts 1982) .

Since *Sarracenia* relies on mineral nutrients derived from digesting captured prey (Ne'eman et al. 2006;Adamec and Pavlovič 2017), another explanation for the maintenance of species in the face of hybridization is post-zygotic hybrid unfitness due to an reduced prey capture efficiency in hybrids relative parental phenotypes. *Sarracenia* species have divergent pitcher morphologies that contribute to different prey capture strategies (McPherson and Schnell 2011). A striking example of this is in the species pair that this study will focus on: *S. rosea*, which has broad, open pitchers with a waxy lip that causes prey to slip and fall into the pitcher, and *S. psittacina*, which has lobster-pot traps with small openings and slender pitchers that prey can easily crawl into while searching for nectar and become trapped. In contrast, the pitchers of *S. rosea* are shorter but much wider, with large openings that allow rainwater to accumulate inside. *S. rosea*’s pitchers act as pitfall traps, like all other species in the genus. Hybrids between these two exhibit a mix of *S. psittacina*-like, *S. rosea*-like, and intermediate pitcher traits (Malmberg et al. 2018) . The mismatch of parental pitcher traits in hybrids or intermediacy of traits could decrease prey capture success, causing unfitness. For example, a pitcher with an intermediate sized pitcher opening would not have the benefits of a small opening (difficult for prey to escape) or the large opening (larger trap for flying insects). Mismatched hybrid pitchers, i.e. those with a wide opening (*S. rosea*-like) but a smaller and slender pitcher shape (*S. psittacina*-like), may not be as successful at prey capture or prey digestion relative to parental phenotypes, resulting in hybrid unfitness which contributes to the maintenance of reproductive isolation.

In such cases where multiple traits contribute to a shared function, the genetic architecture of those traits determines how effective they are as reproductive barriers. Functions that are determined by only a few loci (or sets of tightly linked loci) are more easily introgressed across species than traits that are more genetically complex and are therefore weaker reproductive barriers. Alternatively, functions that are controlled by many unlinked loci are stronger reproductive barriers because introgression of every locus is less likely (Lindtke and Buerkle 2015). Additionally, traits in tight genetic linkage make it more likely that species remain phenotypically distinct in the face of gene flow.

In this study, we investigate the genetic architecture of pitcher morphology in an F2 cross between *Sarracenia rosea* and *S. psittacina*. We aim to identify quantitative trait loci (QTL) associated with key functional traits that comprise the divergent prey capture strategies that differentiate the two species. This study builds on the previous work of Malmberg et al. (2018) by increasing marker density and associating markers and QTLs with physical locations in the genome. We also quantify prey capture success across parental species and hybrids, to determine if hybrid unfitness could be related to decreased prey capture. Taken together, our findings will shed light on the genetic and ecological mechanisms of reproductive isolation in a system where interspecific gene flow is occurring.

## METHODS

### Mapping population and pitcher phenotype measurement

An F2 mapping population was developed for a previous study (Malmberg et al. 2018) by crossing *S. psittacina* and *S. rosea*, and selfing the resulting F1 genotypes. Malmberg et al. measured 26 pitcher traits that contribute to the major differences between the prey capture strategies of *S. psittacina* and *S. rosea*.

Quantifying insect-derived nitrogen in a common garden

Clones of parents and progeny of the mapping population, including F1s, were planted outside in a common garden in 2020. The common garden is located in Athens, Georgia at the Mimsie Lanier Center for Native Plant studies within the State Botanical Gardens of Georgia. The common garden consists of raised beds in a mowed field, with a soil mixture of 50% peat and 50% sand. Plants were spaced evenly in a grid within the boxes.

The percentage of prey-derived N was estimated by comparing the N15/N14 isotopic ratio (δ15N) between a focal carnivorous plant, its insect prey, and a non-carnivorous reference plant. This method, adapted from (Schulze et al. 1991), provides a robust proxy for prey derived nitrogen as the δ15N in carnivorous plant tissue will be elevated relative to that of the reference plant due to the contribution of high-δ15N insect nitrogen. ANOVA was used to determine if there are any significant differences in insect-derived N between the parental species and hybrid progeny.

### Genotype-by-sequencing library preparation and sequencing

Young, unopened pitcher tissue was collected from greenhouse-grown progeny and parents. DNA was extracted either using a CTAB extraction protocol developed for grape (Lodhi et al. 1994) or Qiagen DNeasy Plant Pro kits. Genotype-by-sequencing (GBS) libraries were prepared using the protocol described in (Qi et al. 2018). The restriction enzymes MspI and PstI were used to digest the DNA during the GBS library preparation. Size selection was done using Sera-Mag SpeedBeads at 1X volume to remove fragments below 300bp. Progeny libraries were pooled at equal concentration while parents were pooled at double the concentration of the progeny. The pooled libraries were sequenced on a 10B flow cell on a NovaSeq X at SeqCenter in Pittsburgh, PA.

### Variant calling and linkage map construction

Reads were demultiplexed using the process_radtags function in STACKS (Rochette et al. 2019). Demultiplexed reads were mapped to the *Sarracenia psittacina* reference genome using BWA v. 0.7.17  (Li and Durbin 2009). Variants were called using BCFtools v. 1.15.1 (Danecek et al. 2021). SNPs were filtered to a minimum quality of 1000 and a maximum of 25% missing data. Filtered SNPs were used as markers to construct a linkage map using LepMap3 (Rastas 2017). Since the S. psittacina genome assembly is chromosome level, the physical order of the markers were used as the linkage map order. The markers in the F2 population were called as parental genotypes using the ParentCall2 module. Kosambi distances were calculated using the OrderMarkers2 module. Markers on the ends of chromosomes were manually trimmed if they were more than four centimorgan (cM) from the closest marker.

### QTL mapping

QTL mapping was performed with the R package qtl2 (Broman et al. 2019). Genotype probabilities were calculated with an assumed genotyping error probability of 0.002. QTL were mapped using Haley-Knott regression (Haley and Knott 1992), and LOD significance thresholds were established using a permutation test with 1000 permutations. The proportion of variance explained by each QTL was calculated with the following equation: 1 - 10^(-2 * LOD / number_of_individuals) based on (Broman and Sen 2009).

## RESULTS

### Decreased prey-derived nitrogen in hybrids

Prey-derived nitrogen was measured in a total of 137 individuals, including 6 *S. psittacina*, 6 *S. rosea*, 6 F1s and 119 F2s. ANOVA revealed significant variation in prey-derived nitrogen proportions (p=0.00887, df=3, F=4.027). Both F1s and F2s have significantly less prey-derived nitrogen than *S. rosea* (*S. rosea* vs. F1: p=0.0419, *S. rosea* vs. F2: p=0.0218), but not *S. psittacina*, based on a post-hoc Tukey test (Fig 1).

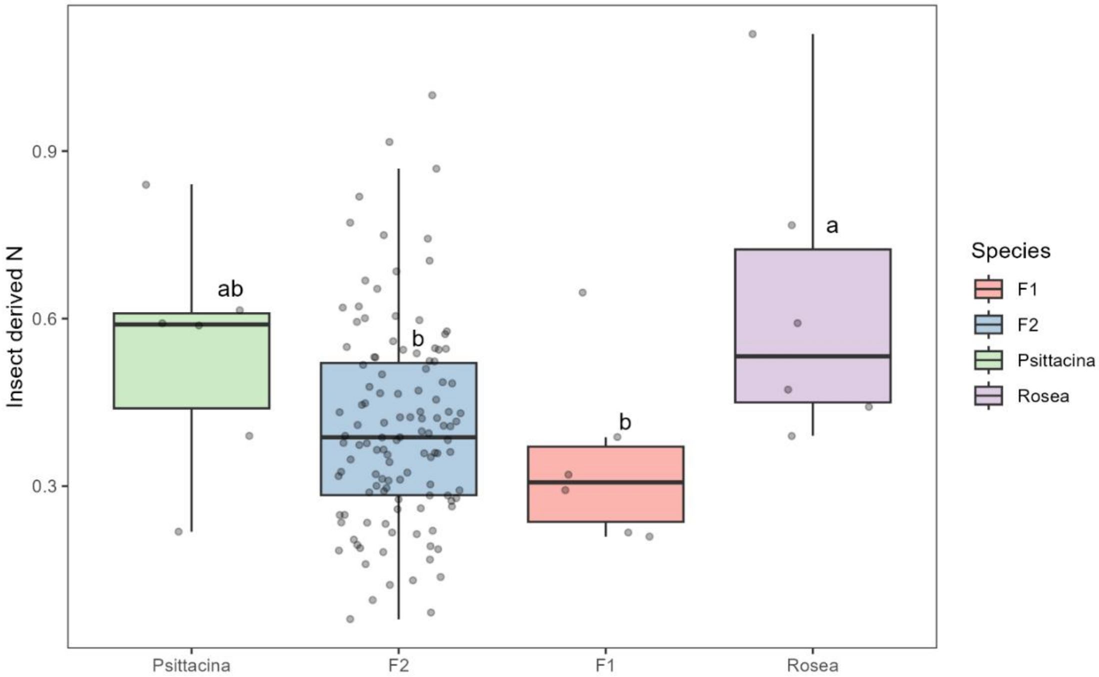


**Figure 5.1**. Proportion of plant nitrogen derived from insect prey in parental and hybrid groups.

## Linkage map

A 1600.38 cM linkage map was constructed from 162 F2s with 4060 parent-specific SNPs. There is a relatively low recombination rate of 0.46 cm/Mb across the entire genome (Table 1), which is typical for angiosperms with large chromosomes like *Sarracenia* (Brazier and Glémin 2022). Recombination is severely suppressed in the repeat-rich pericentromeric regions of the genome (Fig. 2). These regions are also where large structural rearrangements occur between the two parental species, however it is unlikely these are the cause of most of the recombination suppression in this cross as even chromosomes without large inversions have suppressed recombination in the repetitive regions (e.g. Chr 9, Fig.2).

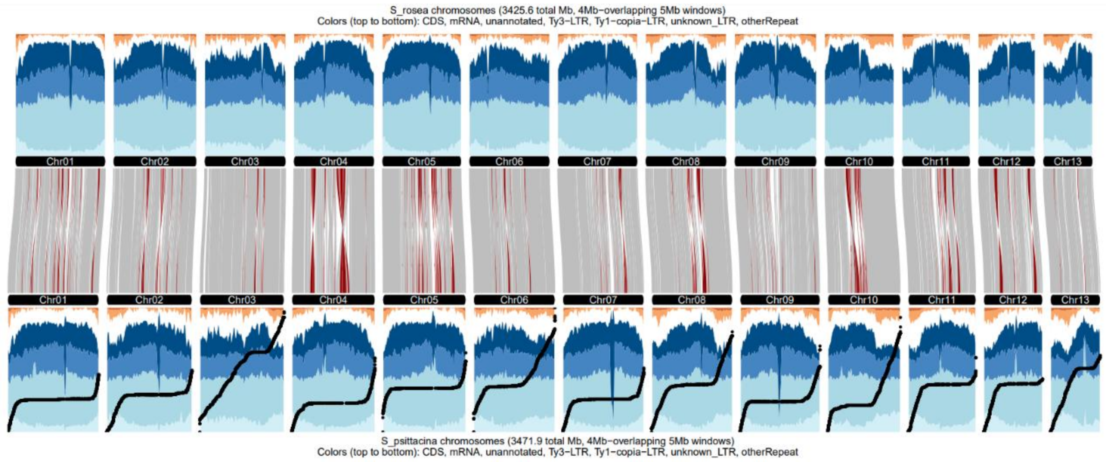


**Figure 5.2**. Linkage map vs. physical map. This plot is the same as Figure 3.1a, with the addition of the markers plotted as black dots along the *S. psittacina* genome at the bottom. Their position along the y-axis is proportional to their position on the linkage map in cM.

## QTL

Previously, Malmberg et al. (2018) mapped 64 QTL for 17 pitcher traits. We remapped these traits using the linkage map based on the newly-assembled genome (Chapter 2) enabling mapping of traits to their physical positions in the genomes. Using newly generated GBS data, we saw a drop to 13 QTL (Fig. 3, Table 1) across only 9 of the pitcher traits from Malmberg et al. (2018). The decreased number of statistically significant QTL may have to do with a drop in the number of F2s with both genotypes and phenotypes (138 vs 230 in Malmberg et al.). In any case, the proportion of variance explained (PVE) by individual QTLs is similar between the two studies.

*S. psittacina* has fenestrations on their pitchers, which are white or pale intravenous areas on the upper portion of the pitchers that may function to mimic the opening of the pitcher, confusing trapped insects (Wicke et al. 2013). QTLs were mapped for three traits related to pitcher fenestrations: *fenestra* - the degree to which a pitcher has fenestrations, *ptnwindw* - the color of the veins surrounding the fenestrations, and *veinprom* - the prominence of the veins compared to the intravenous regions. *Fenestra* and *veinprom* have QTL peaks within 0.5 cM of each other on chromosome 1 (Fig. 4A), while *fenestra* and *ptnwindw* have overlapping QTL intervals on chromosome 3. This colocalization of fenestration traits hints at some degree of pleiotropic control.

**Table 5.1**. QTLs mapped from the data produced in this study. Physical positions are determined by the closest markers. Number of genes within QTLs are shown based on the physical intervals.

| Trait | Chromosome | LOD | PVE | Peak (cM) | Interval (cM) | Peak (Mbp) | Interval (Mbp) | Length (Mbp) | Number of genes |
|---|---|---|---|---|---|---|---|---|---|
| openfrac | 1 | 4.2 | 0.09 | 41.5 | 35.2-65.0 | 23.9 | 21.2-303.9 | 282.7 | 715 |
| openness | 1 | 5.6 | 0.11 | 42.9 | 38.5-59.0 | 25.6 | 22.3-296.5 | 274.3 | 650 |
| fenestra | 1 | 6.5 | 0.12 | 30.1 | 16.9-32.2 | 17.4 | 9.8-19.7 | 9.8 | 100 |
| fenestra | 2 | 5.4 | 0.10 | 32.5 | 22.1-35.2 | 31.5 | 23.2-33.0 | 9.8 | 146 |
| fenestra | 3 | 4.4 | 0.09 | 57.2 | 50.4-105.1 | 57.3 | 45.6-122.8 | 77.2 | 655 |
| radalsym | 13 | 6.2 | 0.12 | 9.6 | 1.1-13.1 | 8.0 | 1.4-10.1 | 8.7 | 161 |
| veinprom | 1 | 5.6 | 0.11 | 30.6 | 25.4-42.9 | 17.9 | 14.0-25.6 | 11.6 | 124 |
| periotuk | 1 | 5.2 | 0.10 | 41.8 | 39.9-44.0 | 23.9 | 23.0-26.2 | 3.1 | 39 |
| ptnwindw | 3 | 4.8 | 0.09 | 33.7 | 32.1-59.4 | 20.0 | 18.7-59.6 | 40.9 | 377 |
| lengintr | 4 | 6.4 | 0.12 | 36.3 | 31.2-48.4 | 28.6 | 21.7-54.3 | 32.6 | 209 |
| smoothzn | 4 | 8.4 | 0.16 | 35.0 | 31.2-48.1 | 26.2 | 21.7-53.8 | 32.1 | 205 |
| smoothzn | 5 | 5.6 | 0.11 | 50.3 | 32.0-54.1 | 7.2 | 0.4-8.6 | 8.2 | 191 |
| smoothzn | 8 | 4.5 | 0.09 | 132.5 | 112.3-143.5 | 250.9 | 236.4-261.7 | 25.3 | 419 |

One of the major differences between the pitchers of *S. rosea* and *S. psittacina* are the openings—*S. rosea* has a wide pitcher opening with outward-curling peristomes (lips) while *S. psittacina* has inward-curling peristomes and a narrow opening, and a fused hood. Three traits related to the openings: *openness* - how open the pitcher mouth is, *periotuk* - the degree to which the peristome is curled inwards, and *openfrac* - the fraction of pitchers on a plant without fused hoods (the F2s typically have pitchers with both fused and unfused hoods). All three traits have only one QTL each, and they are located on chromosome 1 with their peaks existing in a small 1.4 cM interval (Table 1) and the LOD profiles are similar on chromosome 1 (Fig. 4b). Again, this suggests that there may be pleiotropy involved in pitcher opening traits.

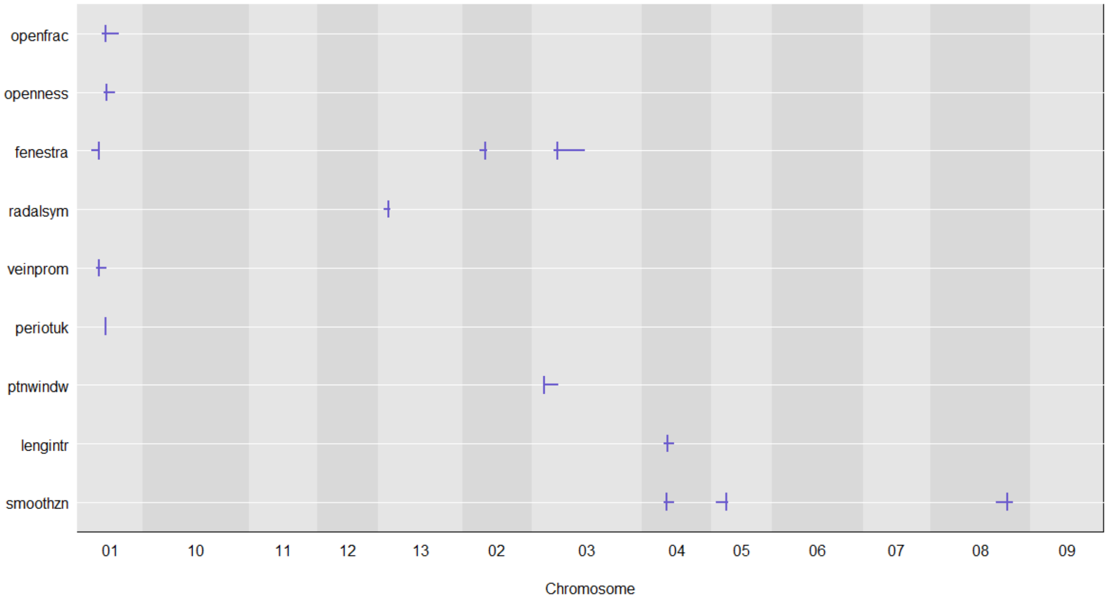


**Figure 5.3**. QTL peaks and 1.5 LOD drop intervals plotted on the linkage map.

In *S. psittacina*, stiff, downward pointing pubescence covers most of the internal pitcher surface, making it difficult for insects inside the pitcher to move up and out of the pitcher. In S. rosea, the majority of the inner pitcher surface is glabrous and waxy, creating a slippery surface that is difficult for insects trapped in the pitcher fluid to use to crawl out. QTL were mapped for two traits related to pubescence on the inner pitcher surface: *lengintr* - the length of the internal region of pubescence starting from the bottom of the pitcher, and *smoothzn* - the length of the internal smooth zone. *Smoothzn* had three QTL mapped to chromosomes 4, 5, and 8, with a combined PVE of 0.36. *Lengintr* had a single QTL mapped to chromosome 4, and its peak is only 1.3 cM away from the *smoothzn* QTL peak on the same chromosome.

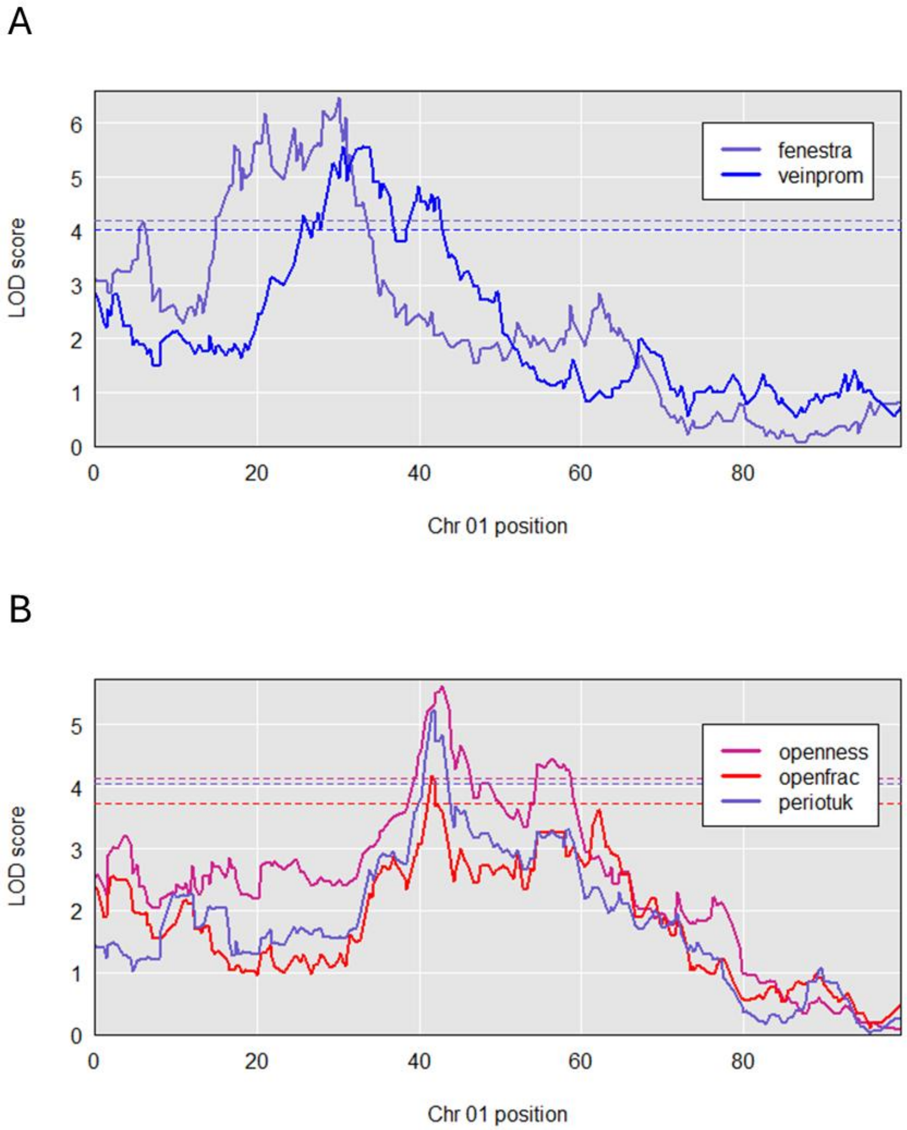


**Figure 5.4**. LOD curves from related traits on Chromosome 1. (A) Fenestration related traits. (B) Openness related traits.

# DISCUSSION

Our data suggest that hybrids between *S. psittacina* and *S. rosea* obtain less nitrogen from insect prey than *S. rosea* parents, but similar levels to that of *S. psittacina* (Fig. 1). This reduction in prey-derived N suggests that hybrid pitcher morphology—being intermediate to that of parents, having mismatched traits, or a combination of both—reduces prey capture efficacy. This reduced prey capture efficacy could in turn contribute to reproductive isolation via hybrid unfitness.

The genetic architecture of pitcher traits underlying the different prey capture strategies of *S. psittacina* and *S. rosea* are relatively simple, with 13 QTLs found across nine traits. Clusters of QTLs were detected on chromosome 1 for fenestration and pitcher opening related traits and chromosome 4 for pubescence traits. Such clustering indicates potential pleiotropy, or at least tight linkage of loci controlling multiple aspects of fenestration, pitcher opening architecture and internal pubescence. This colocalization of QTLs could facilitate coordinated inheritance of functionally coadapted traits.

This genetic architecture has important implications for the persistence of species in the face of gene flow. Traits controlled by few, tightly linked loci are less prone to be broken apart by recombination, allowing species-specific trait combinations to persist despite gene flow. While chromosomal inversions are often implicated in preserving coadapted traits in the face of gene flow (Todesco et al. 2020), it appears that other genomic mechanisms—namely the suppression of recombination in repeat-rich regions—may impede introgression of nuclear loci in *Sarracenia* if low-recombination chromosomal blocks include genes influencing multiple functional traits. In the face of hybridization (chapters 2 and 4), post-zygotic isolation may be enhanced by low recombination within chromosomes and selection for matching of functional traits with functional trait QTL on multiple chromosomes.  Selection against hybrid phenotypes—which have a disadvantage in prey capture efficacy—will work to reduce opportunities for recombination in later generation hybrids and thus reinforce species boundaries.

# CHAPTER 6

# CONCLUSION

Before I started my dissertation research, interspecific gene flow in *Sarracenia* was suspected based on the frequency of hybrids found in the wild (Bell 1952), but genetic investigations in the extent of gene flow were restricted to a small number of populations and used only a handful of loci. While these studies were foundational in our understanding of the consequences of hybridization in specific *Sarracenia* populations (Furches et al. 2013;Rentsch and Holland 2020), the restricted geographic, taxonomic, and genetic scope limited the conclusions we could make about how gene flow has played a role in the evolution of *Sarracenia*. Moreover, there were no genomic resources developed for this genus, rendering it difficult to provide a broader genomic insight into the evolutionary history of *Sarracenia*. In this dissertation I significantly advances our understanding of how hybridization and carnivory have shaped the evolution of pitcher plants throughout four focal chapters.

In **Chapter 2**, I sequence the first complete plastomes for *Sarracenia* and construct a phylogeny from them. The resulting phylogeny revealed extreme cytonuclear discordance—in other words, the plastome tree was vastly different than the previously published nuclear tree. Using coalescent simulations, I reject the null hypothesis that this discordance is cause by incomplete lineage sorting alone, meaning that chloroplast capture must be invoked. Due to the extreme nature of the discordance, I suggest that chloroplast capture is rampant across the genus. While it is difficult to ascertain which species are exchanging chloroplasts, I hypothesize about various scenarios of chloroplast capture by discussing the plastome tree in the context of geography.

**Chapter 3** details the sequencing of the first two *Sarracenia* nuclear genomes and the insights we gained from doing so. The 3.5gb genomes are unusually repeat-rich and gene-poor, which are characteristics shared by some but not all previously published carnivorous plant genomes. Comparative analyses revealed widespread loss of genes involved in photosynthesis and immune response. This suggests that prey-derived carbon may be relaxing selection for efficient photosynthetic machinery, while an intimate relationship with a digestive microbiome might necessitate the suppression of immune responses. The reference genomes I generated in this chapter play a foundational role in chapters 4 and 5.

Whereas chapter 2 describes rampant exchange of plastomes in *Sarracenia,* I explore gene flow in the nuclear genome in **chapter 4**. Utilizing the genomic resources developed in chapter 3, I identify over 3000 conserved single-copy nuclear loci, which I used for phylogenomic analysis. I first estimated a species tree using ASTRAL, revealing a well-resolved phylogeny punctuated by a hard polytomy in the rubra complex. I then investigated gene flow using this species tree as a backbone for phylogenetic network estimation and a quartet-based D-statistic analysis. These analyses revealed widespread gene flow.

Finally, in **chapter 5**, I explored the genetic basis of pitcher traits that contribute to the divergent prey capture strategies of *S. rosea* and *S. psittacina*. Building upon a previous QTL study (Malmberg et al. 2018), I added thousands of genetic markers and a physical context to the linkage map. With this enhanced foundation for QTL discovery, I uncovered a simple genetic architecture for most of the pitcher traits, with QTL held together in tight linkage by suppressed recombination in repeat-rich pericentromeric-regions of the genome. I also showed that hybrids obtain less nitrogen from their prey than *S. rosea*, indicating that hybrid pitcher phenotypes may be maladaptive for efficient prey capture.

My dissertation greatly expands our understanding of the evolutionary consequences of hybridization in *Sarracenia*. While previous work hinted that gene flow could occur between some sympatric species in *Sarracenia*, here I provide evidence that introgressive hybridization is a pervasive phenomenon throughout the evolutionary history of this group. I provide insights into the mechanisms that maintain species boundaries in the face of hybridization, showing that hybrid pitcher phenotypes can result in unfitness due to decreased prey capture. Simple genetic control and tight genetic linkage of these traits could help to maintain the more fit parental phenotypes in the face of gene flow. The reference genomes, phylogenies, and functional insights generated here provide foundational resources for future studies of adaptive evolution, speciation, taxonomy, and conservation in this ecologically and morphologically distinctive plant lineage.

# REFERENCES

Adamec, L. and A. Pavlovič. 2017. Mineral nutrition of terrestrial carnivorous plants. *In* Carnivorous Plants: Physiology, ecology, and evolution. Oxford University Press.

Albert, V. A., S. E. Williams and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. *Science* 257(5076): 1491-1495 DOI: 10.1126/science.1523408.

Baldwin, E., M. McNair and J. Leebens-Mack. 2023. Rampant chloroplast capture in Sarracenia revealed by plastome phylogeny. *Frontiers in Plant Science* 14 DOI: 10.3389/fpls.2023.1237749.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* 19(5): 455-477 DOI: 10.1089/cmb.2012.0021.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires and C. Santos. 2014. Investigating the Path of Plastid Genome Degradation in an Early-Transitional Clade of Heterotrophic Orchids, and Implications for Heterotrophic Angiosperms. *Molecular Biology and Evolution* 31(12): 3095-3112 DOI: 10.1093/molbev/msu252.

Bell, C. R. 1952. Natural Hybrids in the Genus Sarracenia: I. History, Distribution and Taxonomy. *Journal of the Elisha Mitchell Scientific Society* 68(1): 55-80.

Berthier, P., L. Excoffier and M. Ruedi. 2006. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species Myotis myotis and Myotis blythii. *Proceedings of the Royal Society B: Biological Sciences* 273(1605): 3101-3123 DOI: doi:10.1098/rspb.2006.3680.

Bolger, A. M., M. Lohse and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30(15): 2114-2120 DOI: 10.1093/bioinformatics/btu170.

Brazier, T. and S. Glémin. 2022. Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genet* 18(8): e1010141 DOI: 10.1371/journal.pgen.1010141.

Broman, K. W. and S. Sen. 2009. A Guide to QTL Mapping with R/qtl.

Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Ś. Sen, B. S. Yandell and G. A. Churchill. 2019. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics* 211(2): 495-502 DOI: 10.1534/genetics.118.301595.

Buck, R. and L. Flores-Rentería. 2022. The Syngameon Enigma. *Plants (Basel)* 11(7) DOI: 10.3390/plants11070895.

Cao, M., Z. Li, X. Dai, X. Wu, Y. Li and S. Wu. 2019. The complete plastid genome of carnivorous pitcher plant Cephalotus follicularis. *Mitochondrial DNA Part B* 4(1): 2025-2027 DOI: 10.1080/23802359.2019.1617054.

Capella-Gutiérrez, S., J. M. Silla-Martínez and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.

*Bioinformatics (Oxford, England)* 25(15): 1972-1973 DOI: 10.1093/bioinformatics/btp348.

Case, F. W. and R. B. Case. 1976. THE SARRACENIA RUBRA COMPLEX. *Rhodora* 78(814): 270-325.

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* 17(4): 540–552 DOI: 10.1093/oxfordjournals.molbev.a026334.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang and H. Li. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18(2): 170-175 DOI: 10.1038/s41592-020-01056-5.

Chhina, A. K., K. A. Thompson and D. Schluter. 2022. Adaptive divergence and the evolution of hybrid trait mismatch in threespine stickleback. *Evolution Letters* 6(1): 34-45 DOI: 10.1002/evl3.264.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10(2) DOI: 10.1093/gigascience/giab008.

De Queiroz, K. 2007. Species Concepts and Species Delimitation. *Systematic Biology* 56(6): 879-886 DOI: 10.1080/10635150701701083.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24(6): 332-340 DOI: 10.1016/j.tree.2009.01.009.

Dudchenko, O., M. S. Shamim, S. S. Batra, N. C. Durand, N. T. Musial, R. Mostofa, M. Pham, et al. 2018. The Juicebox Assembly Tools module facilitates <em>de novo</em> assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*: 254797 DOI: 10.1101/254797.

Durand, N. C., J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander and E. L. Aiden. 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* 3(1): 99-101 DOI: https://doi.org/10.1016/j.cels.2015.07.012.

Ellison, A. M., E. D. Butler, E. J. Hicks, R. F. C. Naczi, P. J. Calie, C. D. Bell and C. C. Davis. 2012. Phylogeny and biogeography of the carnivorous plant family Sarraceniaceae. *PloS one* 7(6): e39291-e39291 DOI: 10.1371/journal.pone.0039291.

Emms, D. M. and S. Kelly. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20(1): 238 DOI: 10.1186/s13059-019-1832-y.

Fasbender, L., D. Maurer, J. Kreuzwieser, I. Kreuzer, W. X. Schulze, J. Kruse, D. Becker, et al. 2017. The carnivorous Venus flytrap uses prey-derived amino acid carbon to fuel respiration. *New Phytologist* 214(2): 597-606 DOI: https://doi.org/10.1111/nph.14404.

Faust, G. G. and I. M. Hall. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics (Oxford, England)* 30(17): 2503-2505 DOI: 10.1093/bioinformatics/btu314.

Fleischmann, A., J. Schlauer, S. A. Smith and T. J. Givnish. 2017. Evolution of carnivory in angiosperms. *In* Carnivorous Plants: Physiology, ecology, and evolution. A. Ellison and L. Adamec. Oxford University Press**:** 0.

Folk, R. A., J. R. Mandel and J. V. Freudenstein. 2016. Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenomic Discord in a Lineage of Angiosperms. *Systematic Biology* 66(3): 320-337 DOI: 10.1093/sysbio/syw083.

Folk, R. A., P. S. Soltis, D. E. Soltis and R. Guralnick. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany* 105(3): 364-375 DOI: https://doi.org/10.1002/ajb2.1018.

Folkerts, D. R. 1992. Interactions of pitcher plants (Sarracenia: Sarraceniaceae) with their arthropod prey in the southeastern United States. Ph.D., University of Georgia.

Folkerts, G. W. 1982. The Gulf Coast Pitcher Plant Bogs: One of the continent's most unusual assemblages of organisms depends on an increasingly rare combination of saturated soil and frequent fires. *American Scientist* 70(3): 260-267.

Fu, C.-N., S. Wicke, A.-D. Zhu, D.-Z. Li and L.-M. Gao. 2023. Distinctive plastome evolution in carnivorous angiosperms. *BMC Plant Biology* 23(1): 660 DOI: 10.1186/s12870-023-04682-1.

Fukushima, K., X. Fang, D. Alvarez-Ponce, H. Cai, L. Carretero-Paulet, C. Chen, T.-H. Chang, et al. 2017. Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory. *Nature Ecology & Evolution* 1(3): 0059 DOI: 10.1038/s41559-016-0059.

Furches, M. S., R. L. Small and A. Furches. 2013. Hybridization leads to interspecific gene flow in Sarracenia (Sarraceniaceae). *Am J Bot* 100(10): 2085-2091 DOI: 10.3732/ajb.1300038.

Gernandt, D. S., X. Aguirre Dugua, A. Vázquez-Lobo, A. Willyard, A. Moreno Letelier, J. A. Pérez de la Rosa, D. Piñero and A. Liston. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in Pinus subsection Australes. *Am J Bot* 105(4): 711-725 DOI: 10.1002/ajb2.1052.

Glenn, T. C., R. A. Nilsen, T. J. Kieran, J. G. Sanders, N. J. Bayona-Vásquez, J. W. Finger, T. W. Pierson, et al. 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* 7: e7755 DOI: 10.7717/peerj.7755.

Good, J. M., D. Vanderpool, S. Keeble and K. Bi. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* 69(8): 1961-1972 DOI: 10.1111/evo.12712.

Graham, S. W., V. K. Y. Lam and V. S. F. T. Merckx. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist* 214(1): 48-55 DOI: https://doi.org/10.1111/nph.14398.

Grant, P. R. and B. R. Grant. 2019. Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences* 116(46): 23216-23224 DOI: doi:10.1073/pnas.1913534116.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979): 710-722 DOI: 10.1126/science.1188021.

Gruzdev, E. V., V. V. Kadnikov, A. V. Beletsky, E. Z. Kochieva, A. V. Mardanov, K. G. Skryabin and N. V. Ravin. 2019. Plastid Genomes of Carnivorous Plants Drosera rotundifolia and Nepenthes × ventrata Reveal Evolutionary Patterns Resembling

Those Observed in Parasitic Plants. *International Journal of Molecular Sciences* 20(17): 4107.

Haley, C. S. and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69(4): 315-324 DOI: 10.1038/hdy.1992.131.

Hartmann, S., M. Preick, S. Abelt, A. Scheffel and M. Hofreiter. 2020. Annotated genome sequences of the carnivorous plant Roridula gorgonias and a non-carnivorous relative, Clethra arborea. *BMC Res Notes* 13(1): 426 DOI: 10.1186/s13104-020-05254-4.

Harvey, E. and T. E. Miller. 1996. Variance in composition of inquiline communities in leaves of Sarracenia purpurea L. on multiple spatial scales. *Oecologia* 108(3): 562-566 DOI: 10.1007/BF00333734.

Ibarra-Laclette, E., E. Lyons, G. Hernández-Guzmán, C. A. Pérez-Torres, L. Carretero-Paulet, T.-H. Chang, T. Lan, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498(7452): 94-98 DOI: 10.1038/nature12132.

Jin, J.-J., W.-B. Yu, J.-B. Yang, Y. Song, C. W. dePamphilis, T.-S. Yi and D.-Z. Li. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21(1): 241 DOI: 10.1186/s13059-020-02154-5.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. Zerega and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl Plant Sci* 4(7) DOI: 10.3732/apps.1600016.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Syst Biol* 68(4): 594-606 DOI: 10.1093/sysbio/syy086.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14(6): 587-589 DOI: 10.1038/nmeth.4285.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4): 772-780 DOI: 10.1093/molbev/mst010.

Kim, D., J. M. Paggi, C. Park, C. Bennett and S. L. Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37(8): 907-915 DOI: 10.1038/s41587-019-0201-4.

Koren, S., A. Rhie, B. P. Walenz, A. T. Dilthey, D. M. Bickhart, S. B. Kingan, S. Hiendleder, et al. 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* 36(12): 1174-1182 DOI: 10.1038/nbt.4277.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5(2): R12 DOI: 10.1186/gb-2004-5-2-r12.

Kuznetsov, D., F. Tegenfeldt, M. Manni, M. Seppey, M. Berkeley, Evgenia V. Kriventseva and E. M. Zdobnov. 2022. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* 51(D1): D445-D451 DOI: 10.1093/nar/gkac998.

Lan, T., T. Renner, E. Ibarra-Laclette, K. M. Farr, T.-H. Chang, S. A. Cervantes-Pérez, C. Zheng, et al. 2017. Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proceedings of the National Academy of Sciences* 114(22): E4435-E4441 DOI: doi:10.1073/pnas.1702072114.

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* 34(18): 3094-3100 DOI: 10.1093/bioinformatics/bty191.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25(14): 1754-1760 DOI: 10.1093/bioinformatics/btp324.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25(16): 2078-2079 DOI: 10.1093/bioinformatics/btp352.

Lin, C.-S., J. J. W. Chen, C.-C. Chiu, H. C. W. Hsiao, C.-J. Yang, X.-H. Jin, J. Leebens-Mack, et al. 2017. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *The Plant Journal* 90(5): 994-1006 DOI: https://doi.org/10.1111/tpj.13525.

Lindtke, D. and C. A. Buerkle. 2015. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution* 69(8): 1987-2004 DOI: 10.1111/evo.12725.

Lodhi, M. A., G.-N. Ye, N. F. Weeden and B. I. Reisch. 1994. A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. *Plant Molecular Biology Reporter* 12(1): 6-13 DOI: 10.1007/BF02668658.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol* 207(2): 437-453 DOI: 10.1111/nph.13264.

Malinsky, M., H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner and R. Durbin. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution* 2(12): 1940-1955 DOI: 10.1038/s41559-018-0717-x.

Malmberg, R. L., W. L. Rogers and M. S. Alabady. 2018. A carnivorous plant genetic map: pitcher/insect-capture QTL on a genetic linkage map of Sarracenia. *Life Sci Alliance* 1(6): e201800146 DOI: 10.26508/lsa.201800146.

Martin, N. H. and J. H. Willis. 2007. ECOLOGICAL DIVERGENCE ASSOCIATED WITH MATING SYSTEM CAUSES NEARLY COMPLETE REPRODUCTIVE ISOLATION BETWEEN SYMPATRIC MIMULUS SPECIES. *Evolution* 61(1): 68-82 DOI: 10.1111/j.1558-5646.2007.00006.x.

McPherson, S. and D. E. Schnell. 2011. Sarraceniaceae of North America. Redfern Natural History Productions.

Meier, J. I., R. B. Stelkens, D. A. Joyce, S. Mwaiko, N. Phiri, U. K. Schliewen, O. M. Selz, et al. 2019. The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature Communications* 10(1): 5391 DOI: 10.1038/s41467-019-13278-z.

Mendes, F. K., D. Vanderpool, B. Fulton and M. W. Hahn. 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics (Oxford, England)* 36(22-23): 5516-5518 DOI: 10.1093/bioinformatics/btaa1022.

Minh, B. Q., M. A. Nguyen and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30(5): 1188-1195 DOI: 10.1093/molbev/mst024.

Mogensen, H. L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83(3): 383-404 DOI: https://doi.org/10.1002/j.1537-2197.1996.tb12718.x.

Ne'eman, G., R. Ne'eman and A. M. Ellison. 2006. Limits to reproductive success of Sarracenia purpurea (Sarraceniaceae). *Am J Bot* 93(11): 1660-1666 DOI: 10.3732/ajb.93.11.1660.

Neumann, P., A. Navrátilová, A. Koblížková, E. Kejnovský, E. Hřibová, R. Hobza, A. Widmer, et al. 2011. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA* 2(1): 4 DOI: 10.1186/1759-8753-2-4.

Nevill, P. G., K. A. Howell, A. T. Cross, A. V. Williams, X. Zhong, J. Tonti-Filippini, L. M. Boykin, et al. 2019. Plastome-Wide Rearrangements and Gene Losses in Carnivorous Droseraceae. *Genome Biology and Evolution* 11(2): 472-485 DOI: 10.1093/gbe/evz005.

Nguyen, L. T., H. A. Schmidt, A. von Haeseler and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1): 268-274 DOI: 10.1093/molbev/msu300.

Palfalvi, G., T. Hackl, N. Terhoeven, T. F. Shibata, T. Nishiyama, M. Ankenbrand, D. Becker, et al. 2020. Genomes of the Venus Flytrap and Close Relatives Unveil the Roots of Plant Carnivory. *Current Biology* 30(12): 2312-2320.e2315 DOI: https://doi.org/10.1016/j.cub.2020.04.051.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, et al. 2012. Ancient Admixture in Human History. *Genetics* 192(3): 1065-1093 DOI: 10.1534/genetics.112.145037.

Pease, J. B., D. C. Haak, M. W. Hahn and L. C. Moyle. 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biology* 14(2): e1002379 DOI: 10.1371/journal.pbio.1002379.

Qi, P., D. Gimode, D. Saha, S. Schröder, D. Chakraborty, X. Wang, M. M. Dida, et al. 2018. UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in polyploids without a reference genome: finger millet as a case study. *BMC Plant Biology* 18(1): 117 DOI: 10.1186/s12870-018-1316-3.

Qu, X. J., M. J. Moore, D. Z. Li and T. S. Yi. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15: 50 DOI: 10.1186/s13007-019-0435-7.

Rastas, P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics (Oxford, England)* 33(23): 3726-3732 DOI: 10.1093/bioinformatics/btx494.

Rentsch, J. D. and R. C. Holland. 2020. Population Genetic Structure and Natural Establishment of Hybrids Between Sarracenia flava and Sarracenia minor in Francis Marion National Forest. *Castanea* 85(1): 108-121, 114.

Rieseberg, L. H. and D. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants.

Rischer, H., A. Hamm and G. Bringmann. 2002. Nepenthes insignis uses a C2-portion of the carbon skeleton of l-alanine acquired via its carnivorous organs, to build up the allelochemical plumbagin. *Phytochemistry* 59(6): 603-609 DOI: https://doi.org/10.1016/S0031-9422(02)00003-1.

Rochette, N. C., A. G. Rivera-Colón and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* 28(21): 4737-4754 DOI: https://doi.org/10.1111/mec.15253.

Rose, J. P., C. A. P. Toledo, E. M. Lemmon, A. R. Lemmon and K. J. Sytsma. 2020. Out of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in the Flowering Plant Genus Polemonium (Polemoniaceae) Suggests Widespread Historical Gene Flow Despite Limited Nuclear Signal. *Systematic Biology* 70(1): 162-180 DOI: 10.1093/sysbio/syaa049.

Sabater, B. 2021. On the Edge of Dispensability, the Chloroplast ndh Genes. *Int J Mol Sci* 22(22) DOI: 10.3390/ijms222212505.

Schulze, E. D., G. Gebauer, W. Schulze and J. S. Pate. 1991. The utilization of nitrogen from insect capture by different growth forms of Drosera from Southwest Australia. *Oecologia* 87(2): 240-246 DOI: 10.1007/BF00325262.

Sigman, M. J. and R. K. Slotkin. 2016. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell* 28(2): 304-313 DOI: 10.1105/tpc.15.00869.

Smith, M. R. 2020. Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics (Oxford, England)* 36(20): 5007-5013 DOI: 10.1093/bioinformatics/btaa614.

Solís-Lemus, C. and C. Ané. 2016. Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. *PLOS Genetics* 12(3): e1005896 DOI: 10.1371/journal.pgen.1005896.

Solís-Lemus, C., P. Bastide and C. Ané. 2017. PhyloNetworks: A Package for Phylogenetic Networks. *Molecular Biology and Evolution* 34(12): 3292-3298 DOI: 10.1093/molbev/msx235.

Stephens, J. D. and D. R. Folkerts. 2012. Life History Aspects of <i>Exyra semicrocea</i> (Pitcher Plant Moth) (Lepidoptera: Noctuidae). *Southeastern Naturalist* 11(1): 111-126, 116.

Stephens, J. D., W. L. Rogers, K. Heyduk, J. M. Cruse-Sanders, R. O. Determann, T. C. Glenn and R. L. Malmberg. 2015. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus Sarracenia using target enrichment. *Molecular Phylogenetics and Evolution* 85: 76-87 DOI: 10.1016/j.ympev.2015.01.015.

Sukumaran, J. and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics (Oxford, England)* 26(12): 1569-1571 DOI: 10.1093/bioinformatics/btq228.

Thompson, K. A., M. Urquhart-Cronish, K. D. Whitney, L. H. Rieseberg and D. Schluter. 2021. Patterns, Predictors, and Consequences of Dominance in Hybrids. *The American Naturalist* 197(3): E72-E88 DOI: 10.1086/712603.

Todesco, M., M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hubner, S. M. Heredia, et al. 2016. Hybridization and extinction. *Evol Appl* 9(7): 892-908 DOI: 10.1111/eva.12367.

Todesco, M., G. L. Owens, N. Bercovich, J.-S. Légaré, S. Soudi, D. O. Burge, K. Huang, et al. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 584(7822): 602-607 DOI: 10.1038/s41586-020-2467-6.

Veleba, A., F. Zedek, L. Horová, P. Veselý, M. Srba, P. Šmarda and P. Bureš. 2020. Is the evolution of carnivory connected with genome size reduction? *American Journal of Botany* 107(9): 1253-1259 DOI: https://doi.org/10.1002/ajb2.1526.

Warnow, T. 2015. Concatenation Analyses in the Presence of Incomplete Lineage Sorting. *PLoS Curr* 7 DOI: 10.1371/currents.tol.8d41ac0f13d1abedf4c4a59f5d17b1f7.

Whitfield, J. B. and P. J. Lockhart. 2007. Deciphering ancient rapid radiations. *Trends in Ecology & Evolution* 22(5): 258-265 DOI: https://doi.org/10.1016/j.tree.2007.01.012.

Wicke, S., B. Schäferhoff, C. W. dePamphilis and K. F. Müller. 2013. Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. *Molecular Biology and Evolution* 31(3): 529-545 DOI: 10.1093/molbev/mst261.

Wicke, S., K. F. Müller, C. W. dePamphilis, D. Quandt, S. Bellot and G. M. Schneeweiss. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci U S A* 113(32): 9045-9050 DOI: 10.1073/pnas.1607576113.

Winkler, M., A. Tribsch, G. M. Schneeweiss, S. Brodbeck, F. Gugerli, R. Holderegger and P. Schönswetter. 2013. Strong nuclear differentiation contrasts with widespread sharing of plastid DNA haplotypes across taxa in European purple saxifrages (Saxifraga section Porphyrion subsection Oppositifoliae). *Botanical Journal of the Linnean Society* 173(4): 622-636 DOI: https://doi.org/10.1111/boj.12104.

Yu, G., L. G. Wang, Y. Han and Q. Y. He. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16(5): 284-287 DOI: 10.1089/omi.2011.0118.

Zhang, C. and S. Mirarab. 2022. Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees. *Molecular Biology and Evolution* 39(12) DOI: 10.1093/molbev/msac215.

Zhang, C., M. Rabiee, E. Sayyari and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19(6): 153 DOI: 10.1186/s12859-018-2129-y.

Zhou, C., S. A. McCarthy and R. Durbin. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics (Oxford, England)* 39(1): btac808 DOI: 10.1093/bioinformatics/btac808.

Zhu, Q.-g., Y. Xu, Y. Yang, C.-f. Guan, Q.-y. Zhang, J.-w. Huang, D. Grierson, et al. 2019. The persimmon (Diospyros oleifera Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution. *Horticulture Research* 6(1): 138 DOI: 10.1038/s41438-019-0227-2.