MOLECULAR AND MORPHOLOGICAL TESTS OF SPECIES DELIMITATION IN *ROBINIA* OF THE SOUTHERN APPALACHIANS

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Biology

By
Brandon Thomas Wheeler

Director: Dr. Katherine Mathews

Professor of Botany, Biology Department

Committee Members: Dr. Beverly Collins, Biology

Dr. Paul Manos, Biology

March 2023

ACKNOWLEDGMENTS

I've always been told that science takes a village, but only now do I wholeheartedly understand that every study is truly the work of so many people. I would like to thank my committee members for their assistance and encouragement. This project is a testament to their patience, support, and mentorship as it would have never been possible without their guidance. They have allowed me to explore my curiosities while providing the necessary structure to know when it was time to stop coding and start writing.

I'd also like to extend sincere thanks to Dr. Alan Weakley, who allowed me to have a seat in Local Flora and successively nurture a passing interest into a blazing obsession that became a career in botany. I'd also like to thank Mike Kunz for providing sage advice and levity where it was always needed. In addition, I'd like to thank fellow graduate students who have allowed me to kvetch about computer issues and provided insight and laughter where it was desperately needed: Tara Hall, Nicole Cook, Thomas Hennessey, and Austin Brenek. I'd also like to thank my mother, Tara Nye, for instilling me with a love of science from such a young age. Finally, I'd like to thank my partner, Kayla Ebert, for being by my side through all of this and offering me support, laughter, and love at every possible turn.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

MOLECULAR AND MORPHOLOGICAL TESTS OF SPECIES DELIMITATION IN *ROBINIA* OF THE SOUTHERN APPALACHAINS

Brandon Wheeler, M.S.

Western Carolina University (March 2023)

Director: Dr. Katherine Mathews

Robinia L. (Fabaceae) is a genus of trees and small shrubs native to North America, with several species that are of conservation concern within the Southern Appalachians. The genus has been re-circumscribed utilizing morphologically based taxonomic treatments several times in the last century, though no molecular study has focused solely on the genus. Two taxa within *Robinia,* both narrowly endemic to the Southern Appalachians, have variously been treated as species or varieties: *Robinia viscosa* and *Robinia hartwigii.* I studied the phylogenetic relationships of these species in the context of other members of the genus in the Southern Appalachians utilizing a robust Restriction Site Associated Sequencing (RAD-Seq) dataset which was then applied to both species group discovery (STRUCTURE, DAPC) and species delimitation methods (SNAPPER) in order to examine the taxa within this genus. In addition, I examined the commonly held hypothesis that *Robinia hartwigii* is a species of hybrid origin using phylogenetic networks and ABBA-BABA tests to identify potential admixture between taxa in *Robinia*. Additional morphological work utilizing Principal Component Analyses of both herbarium and field observations showed weak divergence in quantitative traits but many

qualitative differences in morphology between the focal taxa. Results from several phylogenetic computational methods were mixed, though several support the recognition of *Robinia hartwigii* at the species level and fail to support a hybrid origin of the species, though they indicate another taxon as highly intermixed. This work represents the first major molecular study of *Robinia* using next generation sequencing and supports the continued recognition of *Robinia hartwigii* as a distinct species.

CHAPTER ONE: INTRODUCTION

Robinia L. (Fabaceae) is a genus comprised of five to eight species of shrubs and small trees native to eastern and southwestern North America. Previous authors have recognized different numbers of species within the genus in the Southeastern United States ranging from as few as four to as many as twelve (Ahles et al. 1968; Isely and Peabody 1984). The most recent treatment recognizes five species in the genus in the Southeastern United States- *Robinia pseudoacacia* L., *Robinia hispida* L., *Robinia viscosa* Vent., *Robinia hartwigii* Koehne, and *Robinia nana* Elliot as well as one species in Southwestern North America, *Robinia neomexicana* A. Gray (Weakley 2020). A lack of consistent circumscription within the genus among authors indicates that this group is an excellent candidate for modern study to verify the current treatment. I will be particularly focusing on delimitation between *Robinia viscosa* and *Robinia hartwigii*, which have been variously lumped and split by different authors since their original description. This group is also an excellent model to study the potential of reticulate evolution in a genus with many wild hybrids reported (Weakley 2020). Additionally, this study will utilize methods that will determine if *Robinia hartwigii* is of hybrid origin, arising from an interbreeding of *Robinia hispida* and *Robinia viscosa* as has been highly hypothesized since the original description of *Robinia hartwigii* (Koehne 1913).

TAXONOMIC BACKGROUND

Robinia viscosa was first described by Ventenant in 1799 from a specimen from 'the headwaters of the Savannah River' in the 'Allegheny' mountains of South Carolina (Ventenat 1799). Several years later, in 1802, Simms described *Robinia glutinosa*, which was later

synonymized with *R. viscosa* (Peabody 1984). *Robinia viscosa* is diploid (2n=20; see discussion of ploidy below) and putative hybrids have been reported with all of the other diploid taxa within the genus (Whittaker 1934). *Robinia viscosa* is listed as S1 (~5 or fewer occurrences) for the state of North Carolina (Wichmann 2021). While rare in its native range, *Robinia viscosa* is an invasive species north of Virginia and in Eastern Europe, thriving in disturbed habitats that mirror the native niche of the species (Burda and Koniakin 2019). It should be noted that while *R. viscosa* is thought to occur in the vicinity of Highlands, NC, I have been unable to substantiate any records in both field and herbarium study (Isely and Peabody 1984). I believe these records are the result of outdated taxonomy that fails to recognize the distinction of *R. hartwigii* at either the species or the variety level. In exhaustive field and herbarium study, I found no evidence of *R. viscosa* in the proposed vicinity of *R. hartwigii,* the closest individuals occurring approximately 120 kilometers to the northeast.

Robinia hartwigii is endemic to forests on the edges of granitic rock outcrops in a roughly 25-mile radius from Highlands, NC, often earning the species a secondary common name of the “Highlands Clammy Locust" (Isely and Peabody 1984). The species was originally discovered in Macon County, North Carolina, by Karl Theodor Hartweg and transported to Germany, where it was officially described by Koehne who named it in honor of Hartweg (Koehne 1913). Due to morphological similarities, *R. hartwigii* has been lumped into *Robinia viscosa* by many authors within the last century (Ahles et al. 1968; Isely and Peabody 1984). Common orthographic variants are known for *R. hartwigii* including *R. viscosa* var. *hardwegii* which Ashe generated when demoting *R. hartwigii* to a variety (Ashe 1922). *Robinia hartwegii* persists as an orthographic variant (likely due to the name of the individual being honored with the specific epithet, Karl Hartweg) but the original description by Koehne is spelled *R. hartwigii*

(Koehne 1913). *Robinia hartwigii* is classified as an S1 species in North Carolina and potentially exotic in Arkansas, though these records appear to need further examination to determine their validity (Wichmann 2021).

Difficulties of identification and inconsistent nomenclature have made study and record keeping of *Robinia hartwigii* and *R. viscosa* difficult and riddled with inaccuracies. Both species are commonly referred to as clammy locusts due to sticky glands that can be found on their twigs and rachises (the central stalk of the leaf, which bears the leaflets). In *Robinia viscosa* these glands are sessile, giving wood a varnished, lacquered appearance. In *Robinia hartwigii*, these glands are on short stalks, persisting on twigs 1-2 years after the original season of growth (Weakley 2020). The secretions from the glands are a supposed defense against herbivory and pathogens due to their composition of a mixture of mucilage, fats, flavonoids, proteins, and alkaloids (Konarska and Łotocka 2020). Beyond these sticky glands, both *Robinia viscosa* and *R. hartwigii* are quite similar in appearance: both species seem to have similar leaflet morphology, growth habit, and overlapping floral characters. Further confounding quick and reliable identification, *R. viscosa* is even known to produce stalked glands along the twigs of the inflorescence**.** This difficulty in identification, combined with their rarity and variable classification in taxonomy, has created a number of problems in the herbarium and scientific record and further exacerbates confusion surrounding accurate taxonomic treatment and the potential ranges of these species.

Robinia hartwigii is often differentiated from *Robinia viscosa* by an ability to abundantly set fruit, while *R. viscosa* is thought to rarely set fruit. A previous study proposed this difference is due to high rates of pollen sterility in *R. viscosa*, as 45% of the pollen grains sampled in *R. viscosa* were sterile, compared to only 10% in *R. hartwigii* (Whittaker 1934). These high rates of

pollen sterility may explain the rarity and poor distribution of the species as well as the apparent rarity of viable fruit on herbarium sheets and in field observations.

The natural ranges of these species are further complicated by anthropogenic introductions and an insidious tendency to escape cultivation. Both species were lauded in the horticultural trade for their blooms and stately foliage, leading to widespread planting throughout Europe and North America (Whittaker 1934). *Robinia viscosa* and *R. hartwigii* are considered invasive species outside of their native range, particularly in northeastern North America and Central Europe (Isely and Peabody 1984; Burda and Koniakin 2019). This combination of anthropogenic establishment and natural persistence outside of their natural range further complicates the understanding of the original range and ecological niche in the Southeastern United States for both *R. hartwigii* and *R. viscosa.*

PLOIDY

Previous studies of the genus have reported (Whittaker 1934) and verified ploidy (Peabody 1984) of *Robinia viscosa* and *Robinia hartwigii* as well as the ploidy of the named varieties and species of *Robinia* available to the authors at the time of publication. Most species within the genus are diploid, excluding most varieties of *Robinia hispida,* which are triploid. The varieties of *Robinia hispida* that are triploid are *Robinia hispida* var. *hispida* and *Robinia hispida* var. *rosea* (synonymized with *Robinia hispida* var. *boyntonii*). Diploid varieties within *Robinia hispida* include *Robinia hispida* var. *kelseyi* (Cowell ex Hutchinson) Isely and *Robinia hispida* var. *fertilis* (Ashe) Clausen. As no tetraploid species of *Robinia* has been identified, it is thought the triploid varieties are the result of a hybridization event between *Robinia hispida* var. *fertilis* and another species within the genus, but there has been little physical evidence to support this

claim (Whittaker 1934). In the remainder of the genus, all species are diploid and readily cross with each other in the wild and in cultivation (Isely and Peabody 1984).

PHYLOGENETIC RELATIONSHIPS

A previous study by Lavin et al. (2003) estimated a broad-scale phylogeny of all species within the Robinoid tribe of Fabaceae in order to investigate relationships of genera within the clade and confirm a morphologically based classification of the tribe. This study analyzed the group using low-throughput Sanger sequencing, specifically the chloroplast intron *trnL*, *matK* loci, and ITS regions of nuclear rDNA and found *R. viscosa* to be sister to *R. hispida* (Figure 1) but did not include *R. hartwigii* or *R. nana*. Lavin et al. (2003) provides the backbone of the genus and showed *Robinia neomexicana* to be sister to the other species within *Robinia* and appropriate as an outgroup for this study. I will build upon the scaffolding of previous work by Lavin et al. to shed light on the evolutionary relationships within the genus *Robinia* with an updated high-throughput sequencing methodology as well as an emphasis on species delimitation between *R. hartwigii* and *R. viscosa*.

CHALLENGES TO SPECIES DELIMITATION AND CAUSES FOR PHYLOGENETIC DISCORDANCE

As many natural hybrids have been described between species of *Robinia*, it was necessary to investigate potential admixture in the phylogeny of the genus. By sharing genetic information across species boundaries, admixture can weaken phylogenetic accuracy and species delimitation by weakening statistical signals based on those shared genes. Introgression is often detected with ABBA-BABA tests, also known as Patterson's D-statistic, as well as f-statistics (Durand et al. 2011; Patterson et al. 2012). As has been seen in previous studies of *Viburnum*, these statistics have been utilized to great success to detect and quantify hypothesized introgression in botanical taxa that can be clouding efforts to delineate species (Spriggs et al.

2019). The D-statistic tests for introgression by swapping the location of potentially introgressed taxa and measuring the signal that indicates where taxon A and B would occur in the phylogeny. Typically, if a D-score is greater than zero this is evidence for introgression due to discordant elements within the phylogeny as genetic material is not following a directly phylogenetic pattern and is instead being found in non-sister taxa (Patterson et al. 2012).

An additional process that can cloud phylogenetic inference from multi-locus datasets is incomplete lineage sorting (ILS) (Maddison and Knowles 2006). ILS is the result of using genes and specific regions of DNA to infer phylogenies of species, as genes occasionally fail to have the same phylogeny of the species, muddying potential phylogenies. RAD-Seq approaches are known to actively combat this by using a wealth of loci from across the genome, though additional methods may be necessary in recently diverged groups (Wagner et al. 2013). In order to infer phylogenies of recently or rapidly radiated groups, it is necessary to employ coalescent methods which account for these potential difficulties. Many of these methods require a reference genome for accurate phylogenetic estimation, but a number have been developed specifically for *de novo* RAD-Seq assemblies using SNPs like 'SVDQuartets' (Chifman and Kubatko 2014).

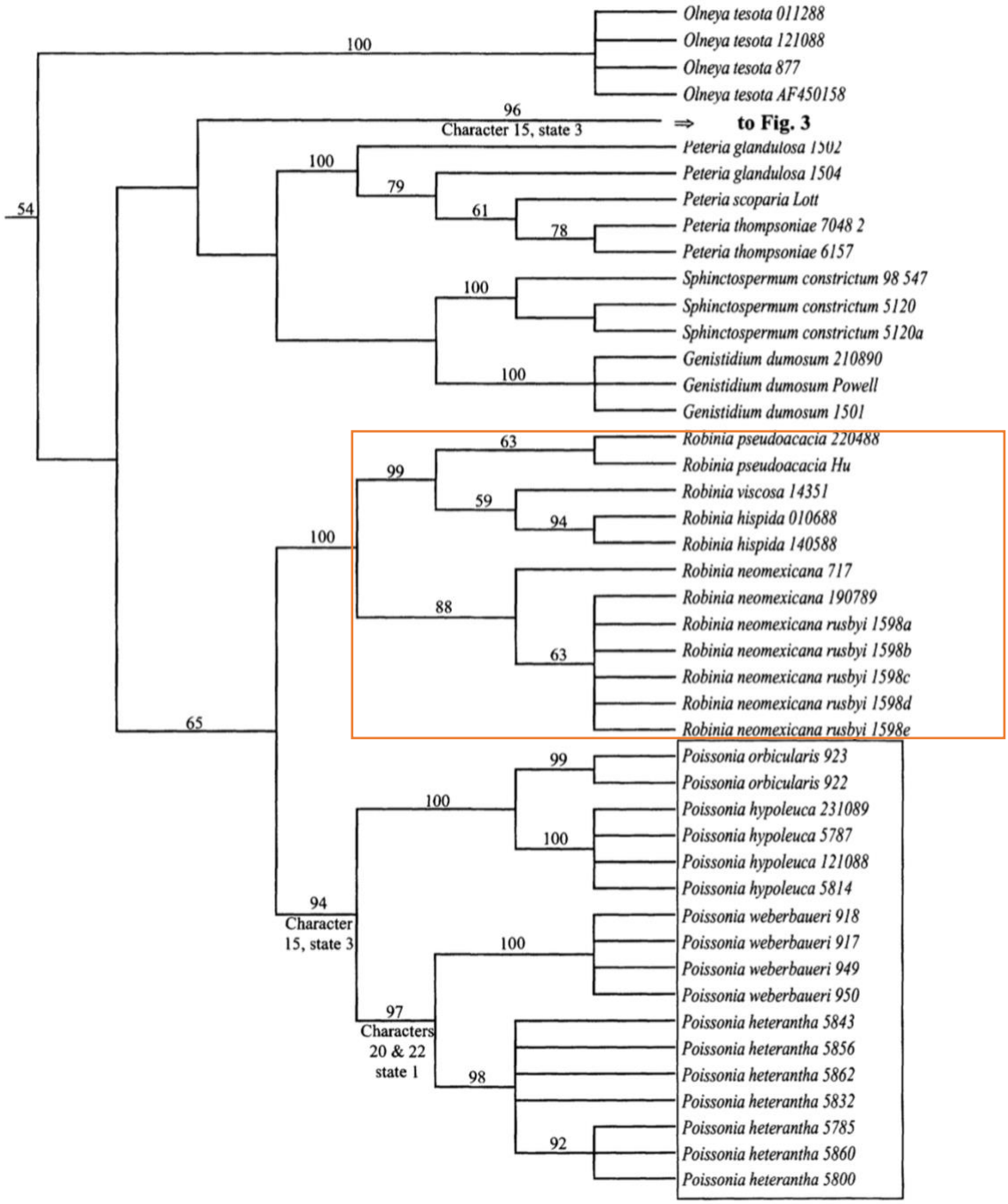

FIGURE 1. Robinoid phylogeny as reported by Lavin et al. 2003. *Robinia* is highlighted by an orange box. This study represents the extent of the current understanding of this group, though several taxa as they are presently treated are not sampled. Thanks to this work I am able to use *R. neomexicana* as a suitable outgroup to root phylogenetic trees.

SPECIES CONCEPTS AND DELIMITATION

Prior to delimiting species, it is necessary to identify a concept by which I will define a species. Over the course of the last century, many species concepts have been used, to varying levels of success and agreement. There are a number of species concepts regularly applied to biological systems, most notably the biological species concept, which defines a species as a group of organisms whose natural breeding produces viable and fertile offspring (Mayr 1942). By this definition, if naturally occurring populations of *Robinia hartwigii* and *Robinia viscosa* are producing viable offspring, they should be treated as a single species. While the prevailing species concept for much of the 20th century, this concept is difficult to when applied to plants, many of which are able to generate viable hybrids across considerable taxonomic distance (Mallet 2005).

Beyond the biological species concept, there are purely phylogenetic species concepts, which require that a species must contain an ancestor and all of its descendants and a diagnosable and definable synapomorphy for each species (Donoghue 1985). In *Robinia*, this would require *R. hartwigii* and *R. viscosa* to form separate clades in a phylogeny in order to recognize both species. Phylogenetic species concepts are quite popular and integral to many modern species delimitation efforts. Further still, there are phenetic species concepts, which require species to exhibit morphological differences that are distinct and quantifiable in order to be delimited (Michener and Sokal 1956). This concept fails to address many aspects of modern systematics and phylogenetics and has become something of an artifact of biological classification. Previous studies in *Robinia* have diagnosed *R. hartwigii* and *R. viscosa* as morphologically distinct, due to the different appearances of their glands, leading to their current recognition as distinct taxa.

More molecular work is necessary to either confirm this treatment or relegate *R. hartwigii* to a variety of *R. viscosa*.

OBJECTIVES AND SIGNIFICANCE

The main purpose of this work is to determine the correct treatment of *Robinia hartwigii* and determine if the taxon should be treated as either a variety of *Robinia viscosa* or deserving of the recognition of an individual species using a robust Restriction Site Associated Sequencing dataset. In addition, this work will identify the phylogenetic origins of *Robinia hartwigii* and if it is truly a species of hybrid origin, as has been widely hypothesized. I will also explore the reticulation throughout the phylogeny of the entirety of the genus *Robinia* in order to further understand the potential origins of *Robinia hartwigii*.

As both *Robinia hartwigii* and *Robinia viscosa* are extremely rare within their native range of the Southern Appalachians, accurate taxonomy is vital to record keeping and their continued existence. Both taxa are monitored by the Natural Heritage Program of North Carolina and confusion surrounding these taxa continues to thwart understanding of their range size. Herbarium specimens are often mislabeled and difficult to identify, as they can be missing key characters that differentiate the taxa.

CHAPTER TWO: METHODS

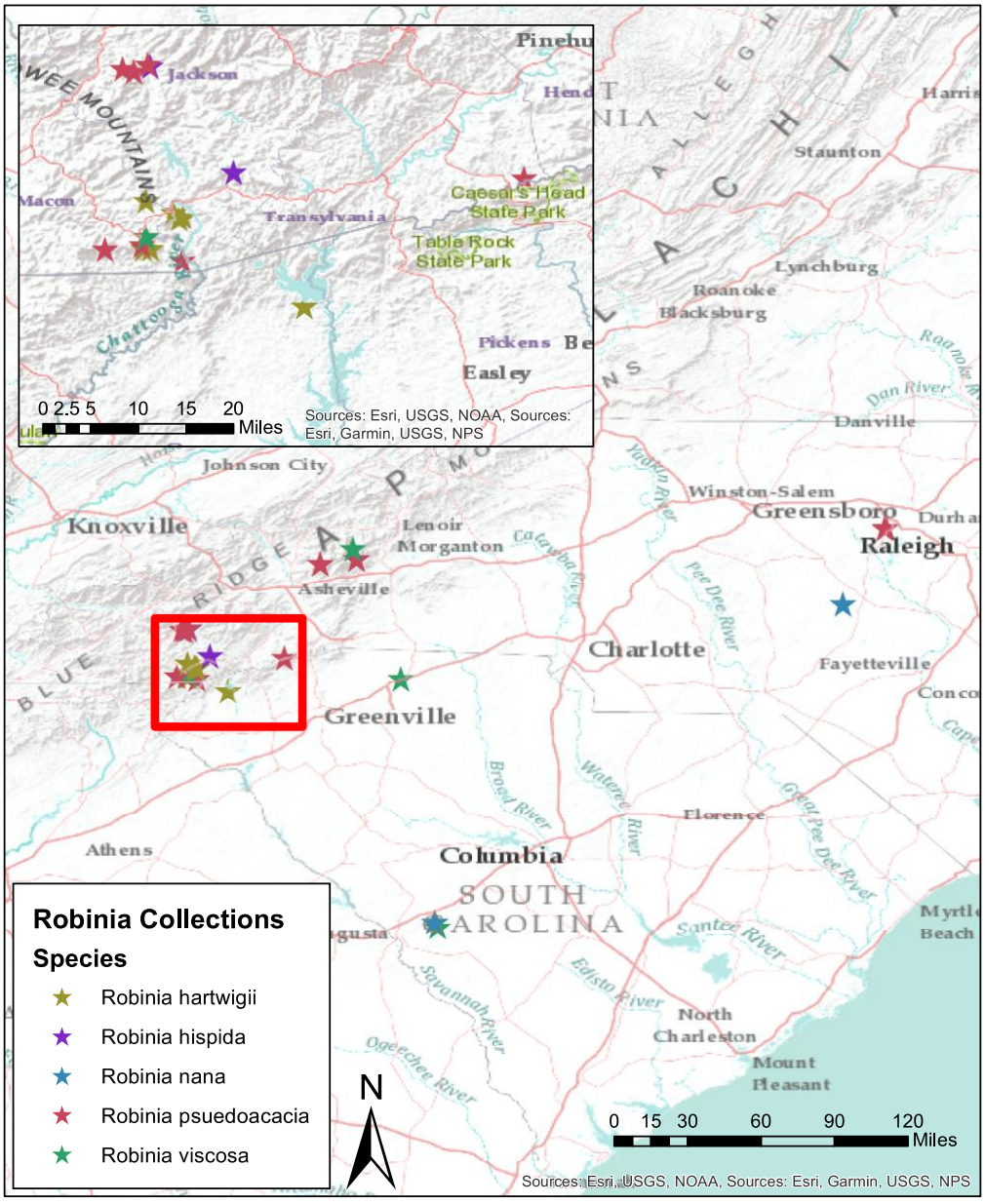

FIGURE 2. MAP OF COLLECTIONS. *Robinia* populations sampled for this analysis, with an inset centered on Highlands, North Carolina. DNA was not successfully extracted from all samples. The color of each symbol corresponds to the species each population was identified as in the field.

SAMPLE COLLECTIONS

To estimate the evolutionary relationships of taxa within *Robinia,* I sampled populations of all species that are known to occur in the southeastern United States as they are currently treated (Weakley 2020). Collection permits were granted by both the United States Forest Service and the Highlands-Cashiers Land Trust, who hold the land that contains the majority of known *Robinia viscosa* and *Robinia hartwigii* populations, respectively. Populations were sampled for both morphological and molecular study. Sampling events in 2022 consisted of preserving fresh, young leaf tissue in silica gel, as well as collecting representative herbarium vouchers for later study. The leaf tissue was then transported to Western Carolina University and stored in silica gel at room temperature prior to extraction. Samples collected in 2021 were stored on ice after collection and transferred to -80º C within 24 hours for storage until DNA extraction, but this method often failed to preserve enough high-quality DNA for sequencing. Each population received a four-letter identifier code, and each individual plant sampled received a number- the four-letter population code and number were then combined to create an identifiable name for each sample. Locations of individuals were recorded with GPS coordinates in order to limit resampling of the same clumps on return visits. Voucher specimens were deposited into the Western Carolina University Herbarium (WCUH) as a function of this project and for future study.

Flowers and fruit were collected from every population that was surveyed for leaf material. Several populations failed to produce flowers or fruit and were therefore not sampled. Fresh flowers were deposited into a Formaldehyde Alcohol Acetic Acid (10%:50%:5% + 35% water) for two days before being transferred to 100% ethanol for microscopy analysis. Fruits

were collected from several populations of both *Robinia viscosa* and *R. hartwigii* populations that produced fruit in the summer of 2022 in order to compare legume morphology.

TABLE 1. Molecular Samples and Corresponding Information. The name of each sample corresponds to the extraction number, while the Pop indicates the population surveyed. The collection refers to the herbarium specimen collected for each population; BW refers to the authors collections as they were deposited into WCUH. Species indicates the morphological field identification of the population.

Name	**Pop**	**Species**	**County**	**State**	**Latitude**	**Longitude**	**Collection**
BW16-har	HIBS2	*hartwigii*	Macon	NC	35.0543400	-83.1895250	BW 143
BW32_psu	LISA1	*pseudoacacia*	Jackson	NC	35.3078460	-83.2084080	BW 122
BW37_har	FODE7	*hartwigii*	Macon	NC	35.0356260	-83.1805000	BW 201
BW39_vis	STFR2	*viscosa*	Buncombe	NC	35.7047950	-82.4024590	BW 181
BW51_har	BIVI6	*hartwigii*	Jackson	NC	35.0922010	-83.1463660	BW 202
BW52_vis	STFR3	*viscosa*	Buncombe	NC	35.7062390	-82.4019350	BW 203
BW60_har	HIBS1	*hartwigii*	Macon	NC	35.0532070	-83.1881980	BW 145
BW72_vis	AISC3	*viscosa*	Aiken	SC	33.5820530	-81.7309070	BW 140
BW73_psu	MOFE2	*pseudoacacia*	Jackson	NC	35.308922	-83.223363	BW 121
BW75_psu	CHMO4	*pseudoacacia*	Macon	NC	35.034948	-83.250097	BW 133
BW111_har	BIVI22	*hartwigii*	Jackson	NC	35.0922010	-83.1463660	BW 205
BW113_har	BIVI23	*hartwigii*	Jackson	NC	35.0922010	-83.1463660	BW 206
BW115_har	WTSD24	*hartwigii*	Macon	NC	35.081359	-83.138049	BW 125
BW116_har	WTSD22	*hartwigii*	Macon	NC	35.08507	-83.132442	BW 207
BW117_har	DEVO1	*hartwigii*	Macon	NC	35.087889	-83.134618	BW 142
BW118_har	SATU23	*hartwigii*	Macon	NC	35.035981	-83.192047	BW 127
BW119_har	STOF22	*hartwigii*	Macon	NC	35.10864	-83.185193	BW 134
BW120_har	SATU22	*hartwigii*	Macon	NC	35.036151	-83.192255	BW 128
BW121_vis	MODO22	*viscosa*	McCormick	SC	33.729193	-82.183977	BW 182
BW124_har	DEFO22	*hartwigii*	Oconee	SC	34.948305	-82.947879	BW 142

BW125_visc	BSSC22	*viscosa*	Spartanburg	SC	35.018749	-81.928615	BW 138
BW126_psu	ROCR22	*pseudoacacia*	Yancey	NC	35.791496	-82.212164	BW 208
BW127_vis	ROCR23	*viscosa*	Yancey	NC	35.79186	-82.212152	BW 190
BW129_psu	STFR25	*pseudoacacia*	Buncombe	NC	35.7047950	-82.4024590	BW 209
BW132_vis	CUCR26	*viscosa*	McDowell	NC	35.730405	-82.190232	BW 135
BW133_har	WTSD23	*hartwigii*	Macon	NC	35.082197	-83.137371	BW 210
BW134_psu	CUCR26	*pseudoacacia*	McDowell	NC	35.731942	-82.190558	BW 211
BW135_vis	CUCR27	*viscosa*	McDowell	NC	35.72937	-82.190592	BW 135
BW112_his	WECA22	*hispida var. hispida*	Jackson	NC	35.312518	-83.18142	BW 212
BW114_his	ZIRC22	*hispida var. kelseyi*	Henderson	NC	35.236215	-82.395937	BW 184
BW130_his	CUCR25	*hispida var. hispida*	McDowell	NC	35.729376	-82.190478	BW 213
BW136_his	CUCR23	*hispida var. hispida*	McDowell	NC	35.728602	-82.185247	BW 214

RESTRICTION SITE ASSOCIATED DNA SEQUENCING

Molecular approaches to generating an evolutionary phylogeny offer a wealth of information in the context of species delimitation and conservation, ranging from the resolution of cryptic taxa to revealing historical rates of introgression between species (Fitz-Gibbon et al. 2017; Burge et al. 2018). High-throughput sequencing and molecular analysis are rapidly increasing the resolution of genetic analysis available to researchers to inform taxonomy in groups of organisms with high morphological variation with previously unclear taxonomic boundaries (Massatti et al. 2016). As the cost of these methods continue to drop, the ability to sequence the genome of an organism quickly and accurately has revolutionized the field of phylogenetics. Restriction Site Associated DNA Sequencing (RAD-Seq) has become a common

tool to construct evolutionary phylogenies, proving to be more accurate than previous methods of DNA Sequencing in species of potential hybrid origin or recent divergence (Perkins et al. 2021).

RAD-Seq is a popular method of sampling an entire genome of an individual to obtain short sequence reads by utilizing restriction enzymes to generate short reads from thousands of potentially homologous loci, across multiple individuals (Narum et al. 2013). This method of high-throughput sequencing allows researchers to easily analyze small portions of the genome when there is no available reference genome for the species. Researchers are able directly compare the genomes of several individuals to each other and scores these differences at sites across the genome. Each sample is ligated with a pre-determined sequence that allows the researcher to identify the sample in downstream analysis (Davey and Blaxter 2010).These differences in the DNA between individuals are referred to as Single Nucleotide Polymorphisms (SNPs) and are scored across the genome to estimate phylogenetic trees, delineate species, and estimate a wealth of conservationally relevant statistics (Evans et al. 2014; Candy et al. 2015). In direct comparisons, RAD-Seq has produced phylogenetic trees that cover larger portions of the genome and offer more parsimony informative sites, leading to phylogenetic inferences with higher statistical support than previous methods.

DNA Isolation and RAD Library Generation

To prepare the leaves for DNA extraction, the silica-dried material was ground in a mortar and pestle over liquid nitrogen immediately prior to extraction. I attempted to extract DNA using both the CTAB method as described in (Doyle 1991) as well as the EZNA kit (Omega Bio-Tek, Norcross, GA). The only extraction method that provided reliable success was the DNeasy Plant mini kit following the manufacturer's protocol (DNeasy, Qiagen, Valencia, Calif). 5 µg of extraction product was run on a 1% agarose gel to confirm the quality and

quantity of DNA obtained from each sample prior to quantifying and sequencing. The concentration of DNA in the extraction was quantified using Broad Range Qubit Fluorometric Quantification Kit (Thermo-Fisher Scientific, Waltham, MA) on a Qubit 2.0 fluorometer (Thermo-Fisher Scientific, Waltham, MA) according to manufacturer's protocol. All samples were then diluted to 10 ng/μL using excess extraction buffer prior to RAD Library preparation. All RAD libraries were prepared by Floragenex Inc. (Portland, Oregon, USA) following the protocol of Baird et al. 2008, which follows. Genomic DNA (0.1–1 μg; from either individual or pooled samples) was digested for 15 min at 37°C in a 50 μL reaction with 20 units (U) of PstI (New England Biolabs [NEB]). Samples were heat-inactivated for 20 min at 65°C. 2.5 μL of 100 nM P1 Adapter, a modified Solexa© adapter (2006 Illumina, Inc., all rights reserved) were added to the sample along with 1 μL of 100 mM rATP (Promega), 1 μL 10× PstI buffer, 0.5 μL (1000 U) T4 DNA Ligase (high concentration, NEB), 5 μL H2O and incubated at room temperature (RT) for 20 min. Samples were again heat-inactivated for 20 min at 65°C, pooled, and randomly sheared (Bioruptor or Branson sonicator 450) to an average size of 500 bp. Samples were then run out on a 1% agarose (Sigma), 0.5× TBE gel and DNA 300 bp to 700 bp was isolated using a MinElute Gel Extraction Kit (Qiagen). The Quick Blunting Kit (NEB) was used to polish the ends of the DNA. Samples were then purified using a Quick Spin column (Qiagen) and 15 U of Klenow exo− (NEB) was used to add adenine (Fermentas) overhangs on the 3′ end of the DNA at 37°C. After another purification, 1 μL of 10 μM P2 Adapter, a divergent modified Solexa© adapter (2006 Illumina, Inc., all rights reserved; top: 5′-Phos-CTCAGGCATCACTCGATTCCTCCGAGAACAA-3′, bottom: 5′-CAAGCAGAAGACGGCATACGACGGAGGAATCGAGTGATGCCTGAGT-3′), was ligated to the DNA fragments at RT. Samples were again purified and eluted in 50 μL. 5 μL of this

product was used in a PCR amplification with 50 µL Phusion Master Mix (NEB), 5 µL of 10 µM modified Solexa© Amplification primer mix (2006 Illumina, Inc., all rights reserved), and 40 µL H2O. Phusion PCR settings followed product guidelines (NEB) for a total of 18 cycles. Samples were gel purified, excising DNA 300–700 bp, and diluted to 10 nM. Barcodes were ligated to sequences prior to digestion followed by sonication and size selection for a mean fragment length of 400 bp. PSTI is a methylation-sensitive enzyme and hence can considerably reduce the fraction of repetitive elements that is otherwise very high in plants, targeting mostly nuclear genes and a few organelle sites (Fellers 2008). RAD libraries were barcoded by individual and multiplexed on NovaSeq 6000. Quality of the reads was checked with 'FQC Dashboard' prior to de-multiplexing and assembly (Brown et al. 2017).

SEQUENCE ASSEMBLY, DATA QUALITY FILTERING AND DATASET GENERATION

There are several programs that convert raw high-throughput sequences to formats useful for downstream phylogenetic and population genetic analyses. *'*ipyrad*'* has emerged as a popular program for demultiplexing and assembling RAD sequences due to the combination of user-friendly tutorials and powerful tools offered to researchers (Eaton and Overcast 2020). Many commonly used downstream analyses have been incorporated into the program under the 'ipyrad.analysis' package, which offers a streamlined workflow to generate reproducible results using Python scripts.

Demultiplexing, assembly, and preparation were performed in 'ipyrad' (Eaton and Overcast 2020). Due to a high volume of reads and to ensure high quality loci for downstream analyses, I used a strict phred Qscore of 43 (i.e., a high quality read of the nucleotide) varied the cluster threshold between 80 and 90 between runs (sequences must be 80 or 90 % identical in order to cluster together in assembly) in the ipyrad assembly parameters. Otherwise, I utilized the

recommended settings in the ipyrad documentation for all assemblies. Reads were assembled *de novo* due to the lack of a high-quality reference genome for *Robinia*. I generated several datasets from the ipyrad pipeline following a similar procedure to Cohen and Schenk 2022, using different assembly parameters to find a final dataset with the highest bootstrap and likelihood support in 'RAxML-NG.' Datasets varied as follows: 1) minimum samples per locus, requiring at least 20, 23, or 24 individuals to have a locus present in the sequence in order for the locus to be retained (min20, 23, and 24, respectively); 2) putative hybrid individuals retained or removed. Using the min samples per locus parameter reduces the percentage of missing data in the alignment at the expense of retaining fewer informative SNPs (Cohen and Schenk 2022); however, large amounts of missing data are known to impact phylogenetic estimation in addition to other downstream applications, like STRUCTURE (Chattopadhyay et al. 2014; Eaton et al. 2017). Additionally, including F1 hybrids in a phylogenetic study could potentially cloud phylogenetic signals and reduce bootstrap supports at crown nodes that correspond to the parental species (Eaton & Ree, 2013). Hybrids were identified by cursory maximum likelihood analysis and confirmed by identifying individuals heterozygous at sites that were homozygous for each species as visualized in 'MEGA 11' (Tamura et al., 2021). All possible dataset formats were generated with the Output Formats parameter in ipyrad, including full sequence alignments (.phy), SNPs only, and unlinked SNPs (one SNP per locus). Different datasets were used in downstream analyses as indicated below.

PHYLOGENETIC ANALYSES

To fully investigate potential phylogenies of *Robinia*, I utilized maximum likelihood method of phylogeny estimation. The most accurate model of molecular evolution for maximum likelihood analysis was TPM1uf+I+G4 as determined using AICc in 'ModelTest-NG' (Darriba et

al., 2020). Phylogenies were generated from the dataset utilizing the 'RAxML Graphical User Interface' (Edler et al., 2021; Stamatakis, 2014). Final maximum likelihood analyses were run for 1000 bootstraps using the rapid hill-climbing tree-search algorithm in 'RAxML-NG'. As input, I used the *.phy ipyrad output files (each individual characterized by one sequence [majority-rule base calling], all loci concatenated into a supermatrix). Felsenstein bootstrap proportion (FBP) and transfer bootstrap expectation (TBE) values were calculated in RAxML-NG. TBE is more appropriate for large phylogenies (>300 samples) and for phylogenies with conflicted branches compared to FBP (i.e., hybridization events) (Lemoine et al., 2018). FBP and TBE values were mapped by RAxML-NG onto the best-scoring maximum likelihood trees. Owing to the size and complexity of the dataset, the required computational resources proved prohibitive for Bayesian phylogeny estimation using 'MrBayes' (Ronquist et al., 2012), as all attempts failed to reach concordance after tens of millions of generations.

I further addressed potential conflict in phylogenetic signal by using concatenated analyses with 'Quartet Sampling' (QS) vers. 1.3.1 (Pease et al., 2018; Weisrock et al., 2012). The quartet concordance score (QC) is defined as the ratio of concordant to both discordant quartets (1: all concordant, > 0: more concordant patterns, < 0: even more discordant patterns), the quartet differential score (QD) indicates the skewness of both discordant patterns (1: equal, 0.3: skewed, 0: all topologies 1 or 2), and the quartet informativeness score (QI) describes the proportion of informative replicates (1: all informative, 0: none informative; see Pease et al. 2018). QD values around 1 indicate ILS (presence of both discordant topologies) whereas QD values towards 0 hint at directional introgression (presence of one alternative topology; Pease et al. 2018; see also (Karbstein et al., 2021). I set 100 replicates per branch and log-likelihood

threshold cutoff to 2, as recommended by the Quartet Sampling documentation (Pease et al., 2018).

Coalescent phylogenetic analyses were performed using the unlinked SNP dataset generated by the ipyrad assembly in both 'SVDQuartets' (Chifman & Kubatko, 2014). SVDQuartets was used within 'PAUP*' v 4.0a169 and *a priori* species assignments based on morphology (Wilgenbusch & Swofford, 2003). SVDQuartets was run for 1000 bootstraps and trees were exported for visualization in 'FigTree' (FigTree)**.** For analyses, putative hybrids were removed prior to running the analysis to reduce potential clouding of phylogenetic signals and interference with species tree generation.

SPECIES GROUP DISCOVERY

To estimate the number of genetic clusters within the *Robinia viscosa* complex without *a priori* assumptions, I utilized several statistical clustering methods. In order to explore potential genetic clustering within the hypothesized species groups, I ran a Principal Components Analysis of the SNP dataset generated by ipyrad in the Assembly and Demultiplexing workflow. These analyses were run in ipyrad.analysis API pipeline (Eaton & Overcast, 2020).

Due to the size and breadth of my molecular dataset, combined with the high potential of sampling clones in populations of *Robinia*, I also utilized a Discriminant Analysis of Principal Components to visualize genetic clustering of all sampled individuals, as recommended for highly clonal populations by (Jombart et al., 2010). I used the R package 'adgenet' and the R Statistical Programming Language for this analysis (Jombart & Bateman, 2008;R Core Programming Team, 2020). Additionally, I ran STRUCTURE in the ipyrad.analysis pipeline to find ancestral groupings indicative of genetically independent groups that could be putative species (Eaton & Overcast, 2020; Pritchard et al., 2000). For these analyses in STRUCTURE, I

removed potential clones and outgroups from the dataset as both can skew potential results from the program (Manos, pers. comm). STRUCTURE was run for 10,000 generations for burn in and 100,000 generations post burn-in. The k values (number of assumed *a priori* ancestral groups) ranged from 2 to 6 on a dataset including all samples of *R. hispida, R. pseudoacacia, R. hartwigii, and R. viscosa* as well as a dataset containing only samples of *R. hartwigii* and *viscosa.* Results from all runs were compiled and visualized using the R Shiny Program 'pophelper' (Francis, 2017). Optimal K values were chosen by selecting the number of clusters for which the slope of the log probability was highest as described by (Evanno et al., 2005).

SPECIES GROUP VALIDATION

Species group validation programs are necessary to validate both previously hypothesized species and groups identified in the group discovery analysis. I used 'SNAPPER,' a package added on to 'BEAST' 2.0 in order to generate coalescent species trees from the unlinked SNP dataset (.usnps) prepared by ipyrad (Bouckaert et al., 2014; Stoltz et al., 2021). SNAPPER generates a species tree for each individual SNP and calculates the likelihood of the species tree using *a priori* species assignments. As SNAPPER only accepts SNP data in binary input, the dataset was prepared for use with SNAPPER using the 'phytools' package in R (Revell, 2012). In order to find the delimitation that best suited the SNP dataset, I utilized a Bayes Factor Delimitation approach which identifies the SNAPPER species assignment that best fits the dataset (Grummer et al., 2014; Leaché et al., 2014). *A priori* species groups were modified between runs of SNAPPER (i.e., splitting and lumping of *R. hartwigii* and *R. viscosa*) and likelihoods were compared using Bayes Factor in order to find the delimitation which provided the lowest calculated likelihood. I utilized 20 chains in this analysis. Each chain consisted of 100,000 MCMC generations and were compiled and analyzed utilizing the Path Sampler

Analyser tool in BEAST 2.0 (Bouckaert et al., 2014). By comparing runs of the analysis that alternatively grouped *Robinia hartwigii* and *viscosa* as either one species or separate taxa, I can utilize Bayes Factor to determine which treatment has the highest posterior probability and therefore best fits the SNP dataset obtained in this study. Resulting coalescent phylogenies were visualized using 'DensiTree' (Bouckaert 2010).

INTROGRESSION DETECTION AND QUANTIFICATION

As hybridization and deeper introgression between species in *Robinia* is suspected based on intermediate morphology of many specimens, I chose to investigate and quantify potential admixture and introgression between hypothesized lineages using my RAD-Seq dataset (Isely & Peabody, 1984). This introgression could also potentially cloud phylogenetic signal and weaken statistical support of the phylogeny estimated using traditional methods. Beyond species group discovery, STRUCTURE analyses (mentioned above) were also utilized to visualize potential introgression and identify potential hybrids (Pritchard et al., 2000). D Statistics and other parameters to detect introgression were calculated using 'D Suite' (Malinsky et al., 2021).

As the ABBA-BABA tests indicated there was potential introgression in the phylogeny, I chose to investigate this introgression further using phylogenetic networks. To gain insight into topological conflict, uncertainty, and possible reticulation, I inferred networks using uncorrected pairwise distances that were calculated from the .usnps files (generated by ipyrad) from each alignment using SplitsTree (Bryant & Moulton, 2004). The .usnps output file randomly selects one SNP per RAD-Seq locus and concatenates them into a new SNP alignment for analysis. This approach was used to avoid linked SNPs, the inclusion of which violates molecular models of evolution. The uncorrected pairwise method estimates proportions of sites at which two sequences differ and uses this distance metric to infer a phylogenetic network.

MORPHOLOGICAL COMPARISONS

I examined 99 herbarium specimens of species within *Robinia* from NCU and WCUH (Appendix A). To identify potential characters to separate the species, I measured and recorded vegetative features in herbarium records as well as field collections and floral features in preserved flowers from field collections (Table 1). Characters targeted for measurement were informed by field observations, description of species in Weakley's 2020 Flora of the Southeastern United States, and other morphometric studies within the Fabaceae (Chandler & Crisp, 1998).

To identify which characters are associated with certain taxa, I performed a linear discriminant analysis (LDA) using the data matrix constructed from the morphological analysis using the 'MASS' package in R (Venables and Ripley 1997). As flowers for only two taxa were collected, the floral characters were analyzed in a PCA using the 'vegan' package in R and visualized using 'ggplot2' (Dixon 2003; Wickham 2016).

TABLE 2. **Morphological Characters Surveyed.** These characters were analyzed for qualitative and quantitative analysis of delimitation. These characters were determined from field/herbarium observations of the taxa in addition to other characters utilized in studies of Fabaceae.

Vegetative Characters	**Floral Characters**
Presence/Absence of Glands	Length of Calyx Lobes
Glands Raised	Calyx Vestiture
Length of Glandular Stalk	Banner Height
Ratio of Length/Width of Leaflets	Banner Width
Pubescence	Keel Width
Shape/Texture of Hairs	Ratio of Banner/Keel

CHAPTER THREE: RESULTS

DNA ISOLATION AND RAD LIBRARY GENERATION

RAD library construction generated between 1.6 million- 10 million reads per sample. After filtering and clustering in ipyrad about 45,400 putative loci were retained in the final min20 dataset. The min20 dataset contains 274,173 total SNPs, with a total of 8.31% missing data (Table 3).

TABLE 3. RAD-SEQ Assembly Statistics. Reported statistics of various assemblies and their parameters used in this work. clust=clustering parameter (i.e., what percentage of a read must be identical in order to cluster as a loci), min=minimum number of samples that must contain the loci for it to be retained, clam=only clammy locusts, nohybs= no putative hybrids, noclones= only one individual retained from each population in order to limit sampling of clones.

Matrix	***n***	**Number of loci**	**Number of unlinked SNPs**	**Consensus sequences (bp)**	**VAR (%)**	**PIS (%)**	**Missing data (%)**
clam_noclones_min16	16	19,588	55,390	2,845,106	1.94	1.14	0.32
clust80_nohybs_min20	26	44,892	228,562	6,538,377	3.50	1.75	6.28
clust80_nohybs_min22	26	41,477	206,564	6,039,615	3.42	1.70	5
clust80_nohybs_min23	26	39,001	191,173	5,678,157	3.37	1.67	4.29
clust80hisyes_min20	27	45,400	274,173	6,617,362	4.14	1.86	7.04
hisyes_no_outs_min20	23	15,484	75,951	2,252,411	3.37	1.47	0.45
clust90_nohybs_min20	26	47,096	235,326	7,091,968	3.32	1.66	5.28
clust90_nohybs_min23	26	42,016	206,210	6,115,746	3.37	1.67	4.18
clust90_nohybs_min22	26	44,625	222,029	6,495,879	3.42	1.70	4.88

PHYLOGENETIC ANALYSIS

As original attempts including all individuals were unsuccessful due to paralogs masquerading as orthologs and creating unrealistic branch lengths, they were removed from the analysis. For all diploid individuals sampled, I generated several RAD-Seq datasets consisting of ~ 40,000 putative loci retained for each individual. Both maximum likelihood and coalescent-based methods generated trees that supported four major clades (excluding the outgroups) that correspond to each of the species as they have recently been treated (*R. hispida* var. *kelseyi, R. hartwigii, R. viscosa, R. pseudoacacia)*. Bootstrap support of each of the nodes which corresponded to putative species were generally high (.90-1.00) and lower values were only seen within individuals of the same putative species (Figure 3). These clades were highly supported in all generated datasets and generally unaffected by the minimum loci or clustering parameters. In coalescent analysis using SVD Quartets, the species tree also generated nodes of high support as determined by bootstraps (Figure 4). Quartet values also indicate a high degree of confidence in the inferred phylogeny (Figure 5).

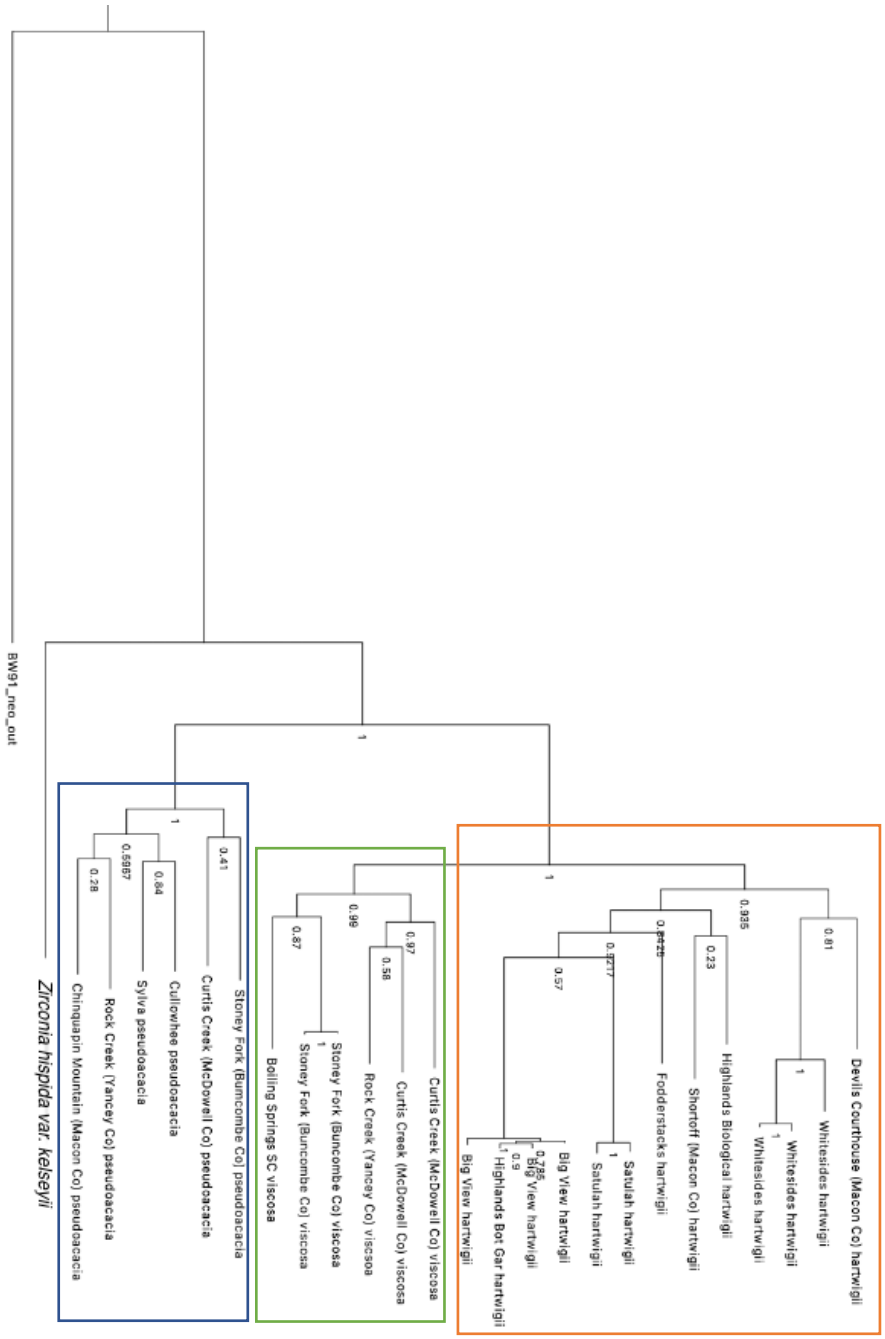

Figure 3. Maximum Likelihood Phylogeny. Generated by RAxML based on clust90_nohhybs_min20 dataset (-lnL= -9252301.137317). Transfer expectation bootstrap values reported at nodes. The orange box represents the *R. hartwigii* clade, the green represents the *R. viscosa* clade, and the blue represent the *R. pseudoacacia* clade.

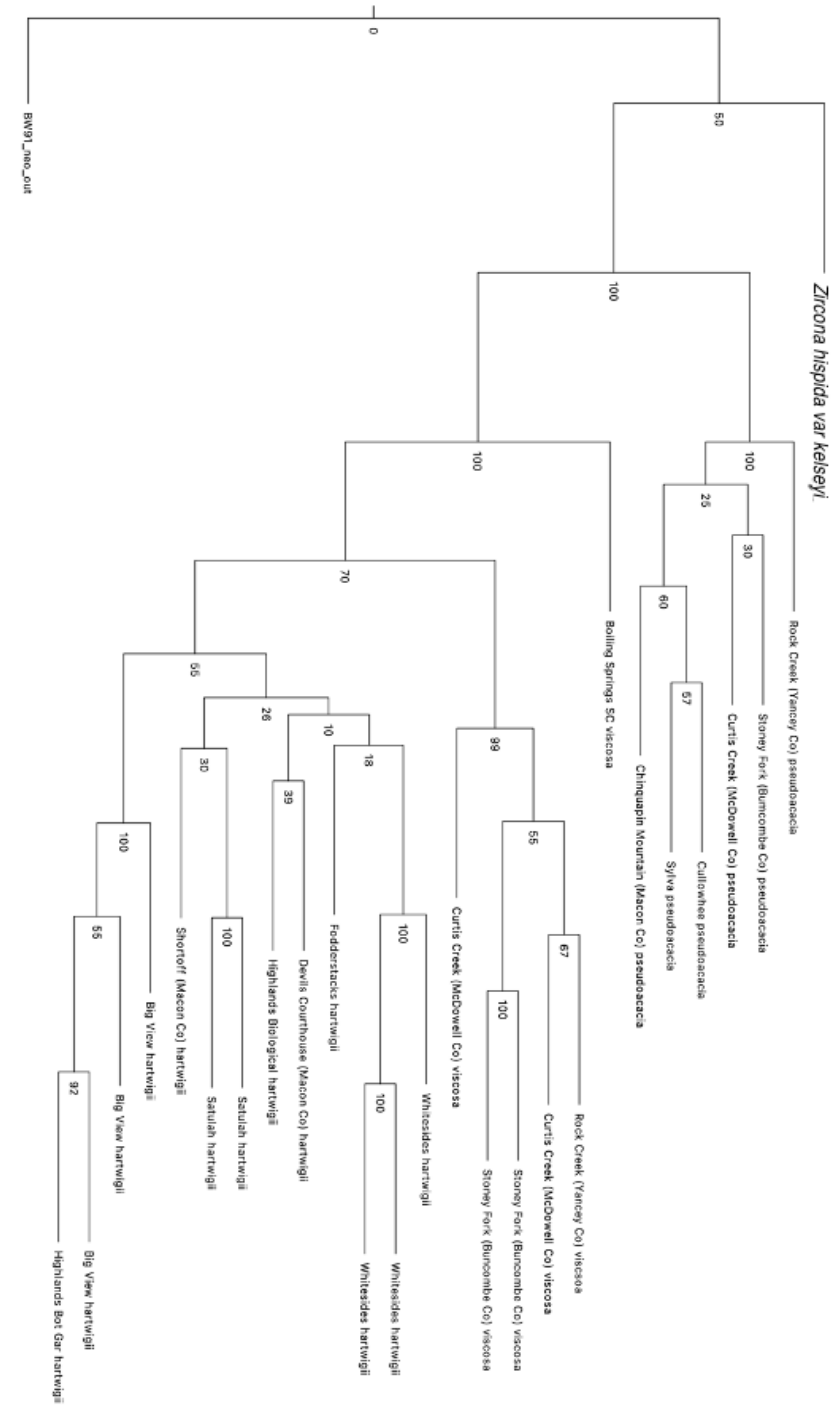

FIGURE 4. SVDQuartet Coalescent Phylogeny. Coalescent phylogeny based on the unlinked SNP dataset (only one SNP per RAD locus) estimated using SVDQuartets. This analysis is an attempt at combating potential incomplete lineage sorting in order to better understand the phylogeny of *Robinia*. Note bootstrap scores of nodes corresponding to the crown of each species group, which indicate high support of monophyly of both *R. hartwigii* and *R. viscosa* and their independence as taxa.

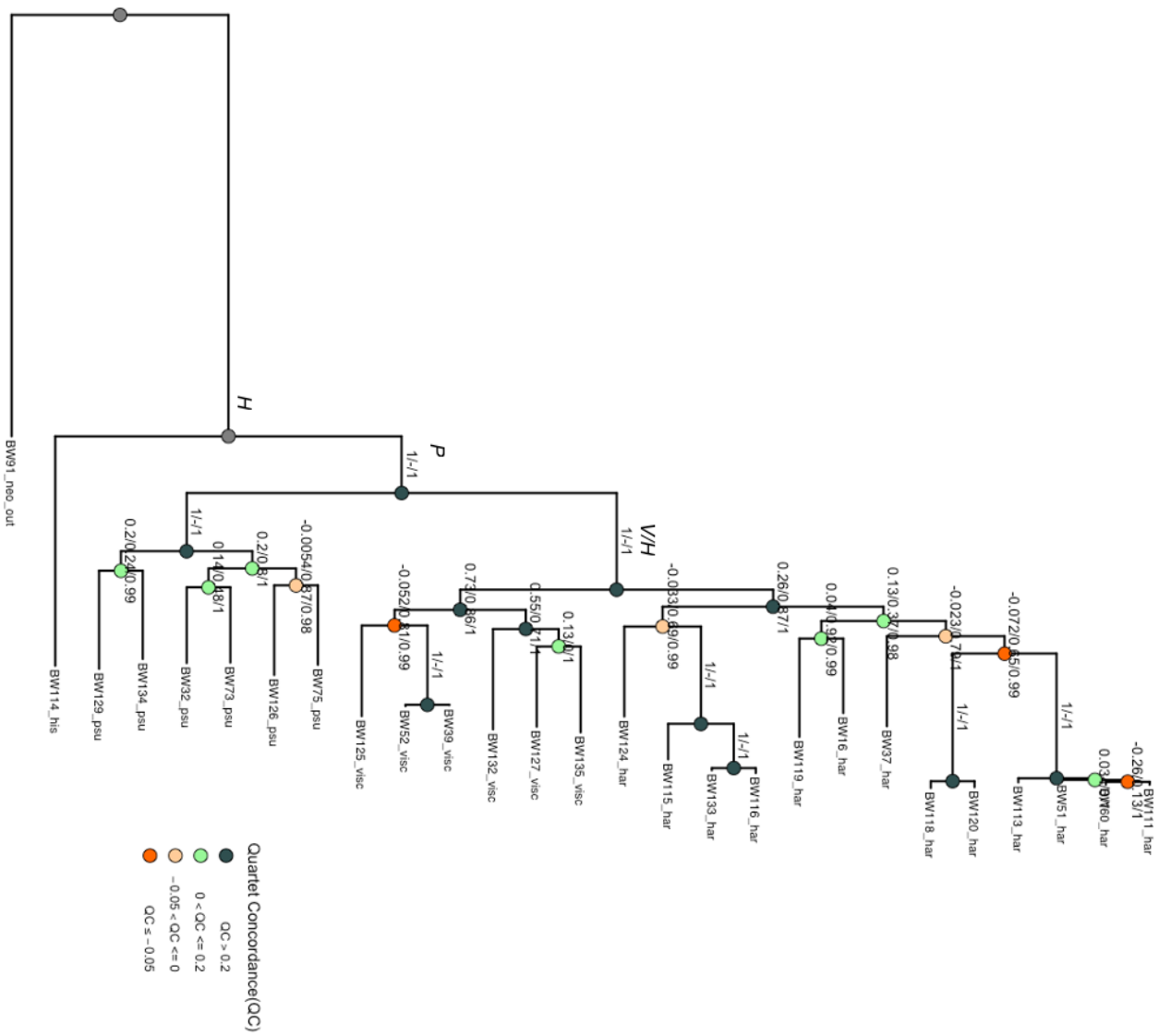

FIGURE 5. Quartet Sampling Phylogeny. Quartet sampling scores of phylogeny scores of each node are represented as quartet concordance/quartet differential/quartet informativeness. Color of node is depicted based on quartet concordance to emphasize potential introgression. The nodes corresponding to the taxa of interest appear to be highly supported and show little evidence of introgression due to Quartet Concordance scores. The crown node of each group is indicated by a capital letter that refers to the species group that is representative of: h for *R. hispida,* p for *R. pseudoacacia,* and v/h for *R. hartwigii/ R. viscosa.*

SPECIES GROUP DISCOVERY

Principal Component Analyses of the SNP dataset illustrate some overlap between *Robinia hartwigii* and *Robinia viscosa*, while indicating both are particularly distinct from other members of the genus that were sampled for this analysis (Figure 5). When other species were removed from the PCA, *Robinia hartwigii* and *Robinia viscosa* populations appear to be somewhat differentiated, though some populations of *R. hartwigii* seem to be highly divergent from the central cluster (Figure 6).

The discriminant analysis of principal components (DAPC) of the SNP dataset failed to differentiate between *R. hartwigii* and *R. viscosa* including attempts with *R. pseudoacacia* and attempts dropping *R. pseudoacacia* from the analysis (Figure 8). This method tended to split *R. hartwigii* between populations rather than differentiate directly from *R. viscosa*, echoing some of the earlier findings from the PCA. This could be due to high genetic differentiation between populations as a result of drift and small populations sizes in these rare species.

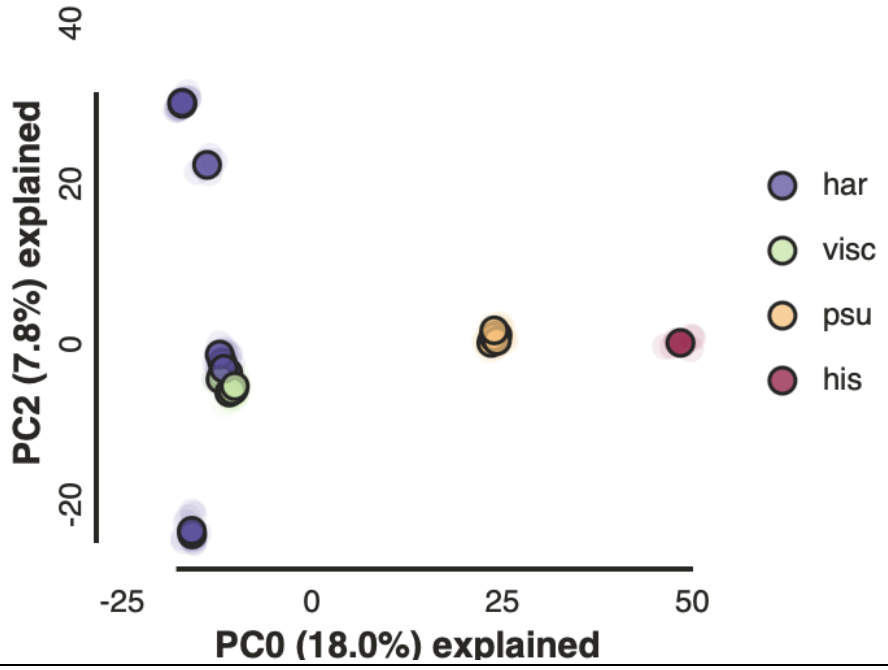

FIGURE 6. PCA of All Samples. Principal Components Analysis of the Full SNP dataset. This analysis attempts to identify grouping within the molecular dataset, with each species collected represented by a different color. *R. hispida* and *R. pseudoacacia* appear to be distinct, while *R. hartwigii* appears to be highly variable- some samples clustering with *R. viscosa*, and others broadly separated on the y-axis. These findings somewhat weaken the argument for the delimitation of *R. hartwigii*, as the samples appear to be tightly clustered.

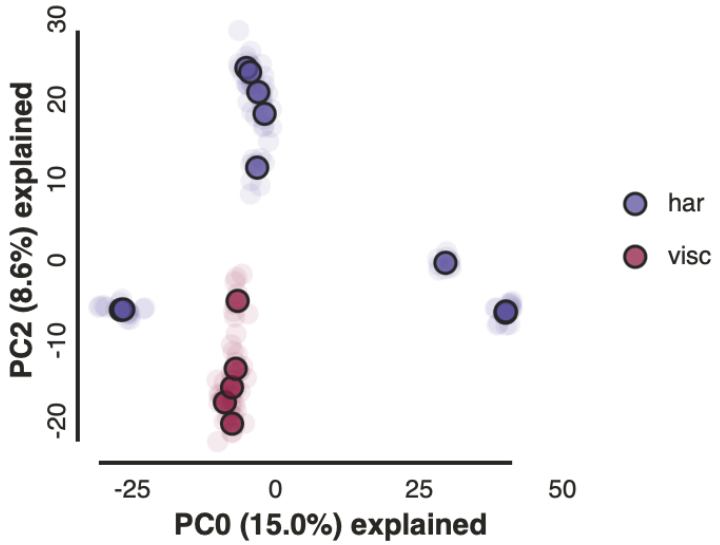

FIGURE 7. Principal Components Analysis of the Clammy Locusts. When removing the other taxa from the analysis, there appears to be further separation between *R. hartwigii* and *R. viscosa*, though *R. hartwigii* still appears to be highly variable, while *R. viscosa* appears to be more tightly clustered. This could be indicative of introgression or genetic drift within *R. hartwigii.*

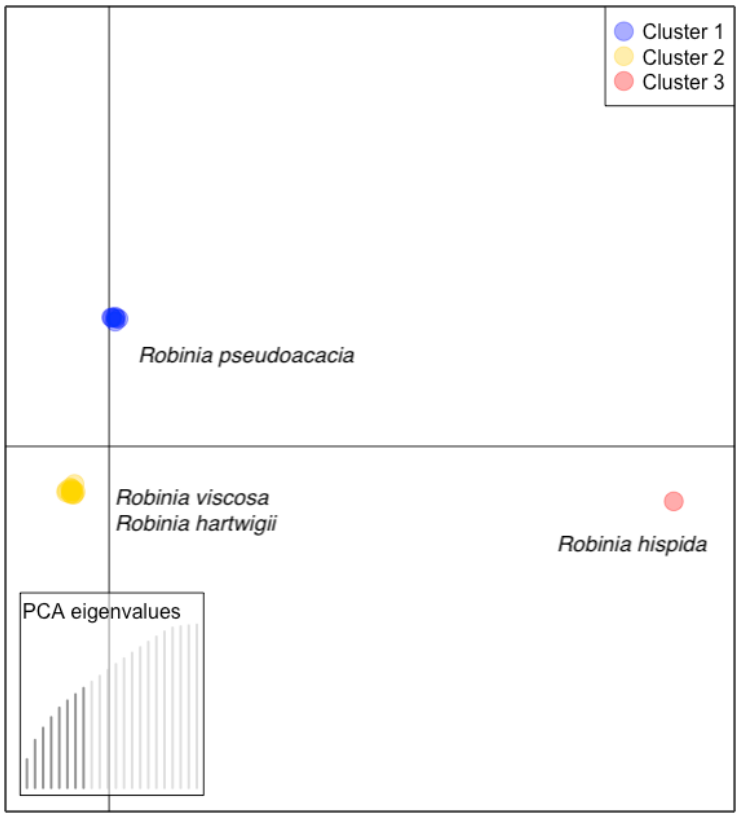

FIGURE 8. Discriminant Analysis of Principal Components for all Taxa Sampled. This analysis reduced the variation between *Robinia hartwigii* and *R. viscosa* causing them to tightly cluster together, another argument that these taxa fail to differentiate from each other.

Analysis using STRUCTURE was variable across values of k (i.e., different numbers of pre-assigned ancestral groups). When other groups were removed, *R. hartwigii* and *R. viscosa* were partially delineated under k=2 (Fig. 9), the k value that best fit this dataset using the Evanno 2015 Method of identifying changes in the log likelihood of each successive run of the analysis (Fig. 8) for this subset of the data. This pattern coordinates with the projected geographic range of each taxon, with *Robinia hartwigii* tightly clustered around Highlands, NC in Southern Jackson County while *Robinia viscosa* is broadly dispersed north of the French

Broad River in Buncombe County, NC. Individuals outside of these previously described regions appear to be admixed, either as a result of merging when reintroduced outside of their natural range or due to anthropogenic interbreeding for horticultural purposes (Figure 10). In runs of the analysis that included *Robinia pseudoacacia,* the populations fail to be delineated between *Robinia hartwigii* and *Robinia viscosa* (Figure 11).

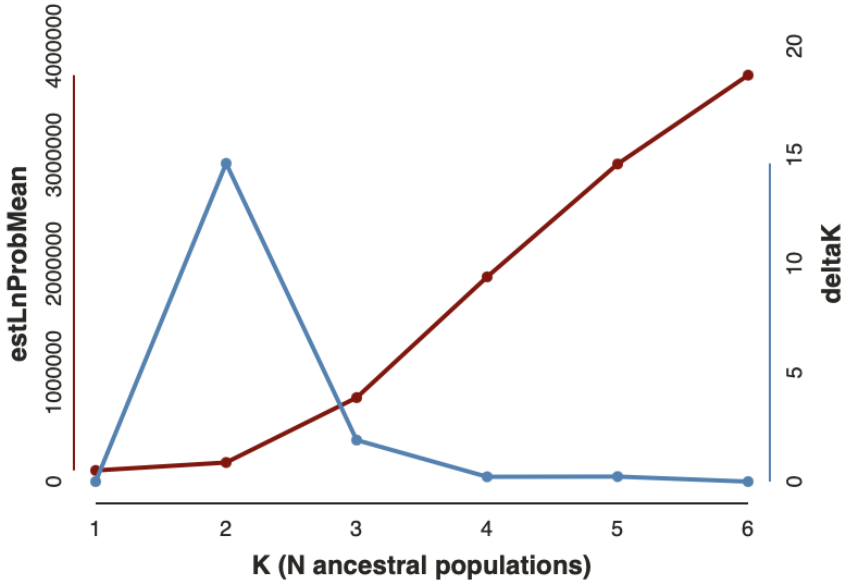

FIGURE 9. Evanno Estimation of K. Evaluation of the best value of k using the method proposed by Evanno 2005 on STRUCTURE analysis of only *R. hartwigii* and *R. viscosa* samples. The red line represents the mean log probability of the models in the analysis and the blue line represents rate of change in the log probability of data between successive k values (delta k). Both values indicate a k value of 2 to be optimal for this dataset.

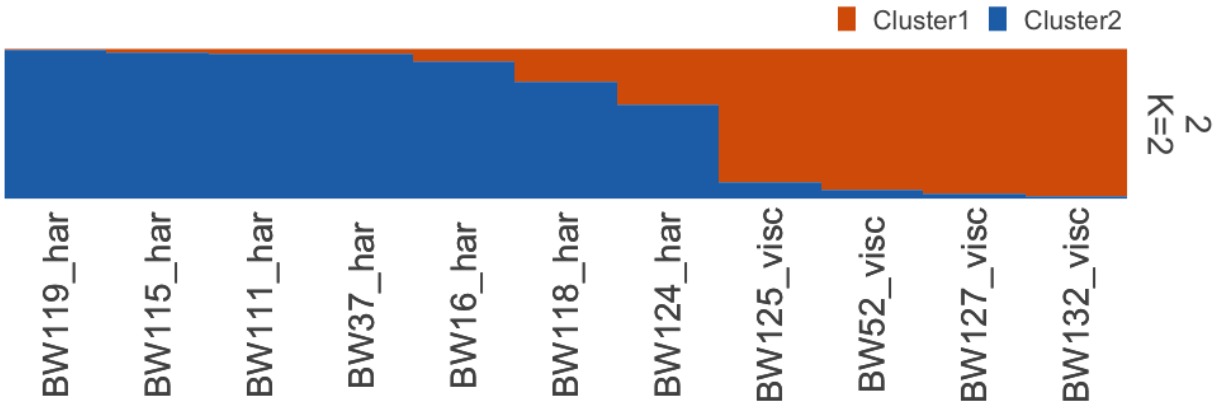

FIGURE 10. STRUCTURE Analysis of Only *R. hartwigii* and *R. viscosa.* STRUCTURE results when limiting the analysis to only *R. hartwigii* and *R. viscosa* samples. Here the algorithm clearly delineates between the two, though there is a gradient between the taxa. This could represent a geographic cline if they appear to be arranged geographically, or it could be the result of admixture when growing outside of the native range.

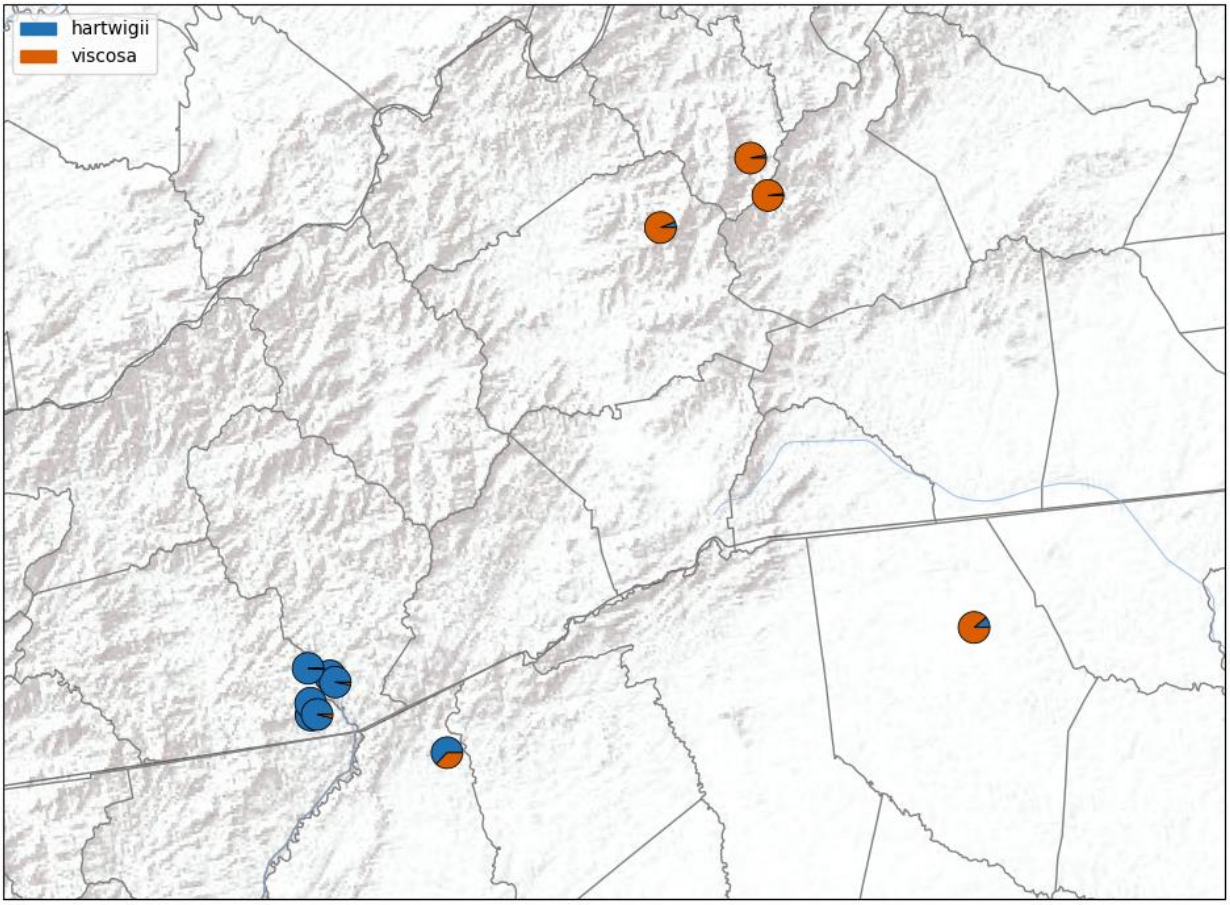

FIGURE 11. STRUCTURE Analysis Mapped to Population. STRUCTURE results for all viscosa and hartwigii populations mapped to the location of collection. *R. hartwigii* samples are centered around highlands, north carolina in Jackson and Macon counties and can be identified as the blue cluster. *R. viscosa* samples are more broadly distributed to the east. Both South Carolina populations appeared to be the result of anthropogenic cultivation due to their locations and are the most admixed populations.

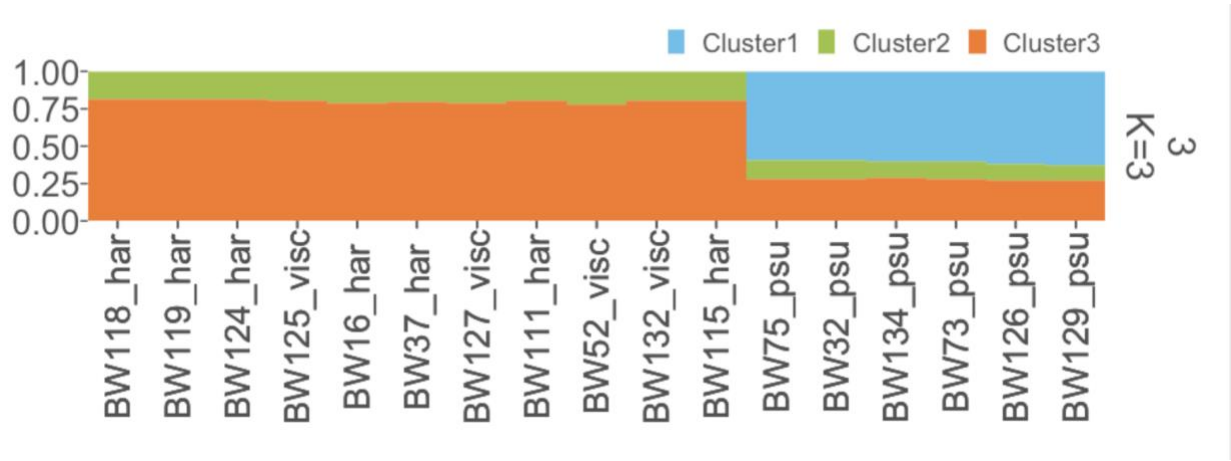

FIGURE 12. Structure Analysis Of All Populations Sampled. STRUCTURE diagram for k=3, including Robinia pseudoacacia populations sampled, indicated by the last 3 letters of the sample name. This analysis fails to delimit between the two focal species with the inclusion of the sister taxon, instead identifying the third entity as the random noise.

Though the STRUCTURE results seem to indicate the potential of a geographic cline between the two species (see above), there does not appear to be any isolation by distance between populations of *Robinia hartwigii* and *Robinia viscosa* (Figure 13, p=0.92), though this could potentially be skewed by anthropogenic spread of genotypes beyond their native range and highly segmented distributions of each species. Only one individual per population sampled was included in the analysis to prevent sampling of clones that would skew the resulting analysis.

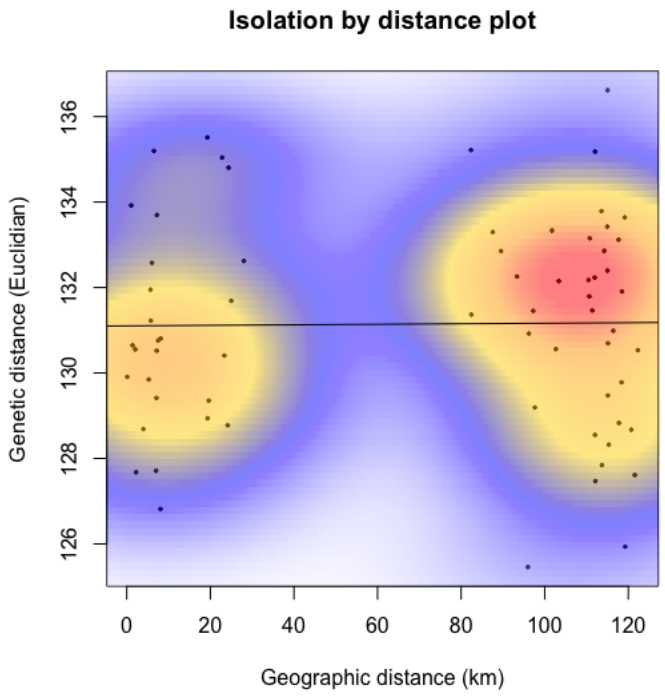

FIGURE 13. Analysis of potential isolation by distance. This is an analysis of only populations of *Robinia hartwigii* and *Robinia viscosa* with all potential clones removed. This indicates a lack of isolation by distance and negates that the STRUCTURE results are due to a geographic cline of variation between the two taxa.

INTROGRESSION DETECTION AND QUANTIFICATION

The phylogenetic network visualized from the unlinked SNPs file generated by ipyrad shows many potential introgressions in the phylogeny of *Robinia* in the Southern Appalachians, between all species groups, as seen by additional lines between each node (Figure 14). This reflects a highly reticulate pattern of evolution throughout the genus, which has often been assumed by previous study of this group.

Quartet Concordance scores did not detect alternative topologies among putative species groups, as each node corresponding to a species group showed few discordant patterns (Figure 5). This indicates few patterns of introgression that weaken the phylogenetic implication of this

work. It should be noted, however, that quartet values were not calculated for the *Robinia hispida* var. *kelseyi* node due to the nature of the program and lack of sampling.

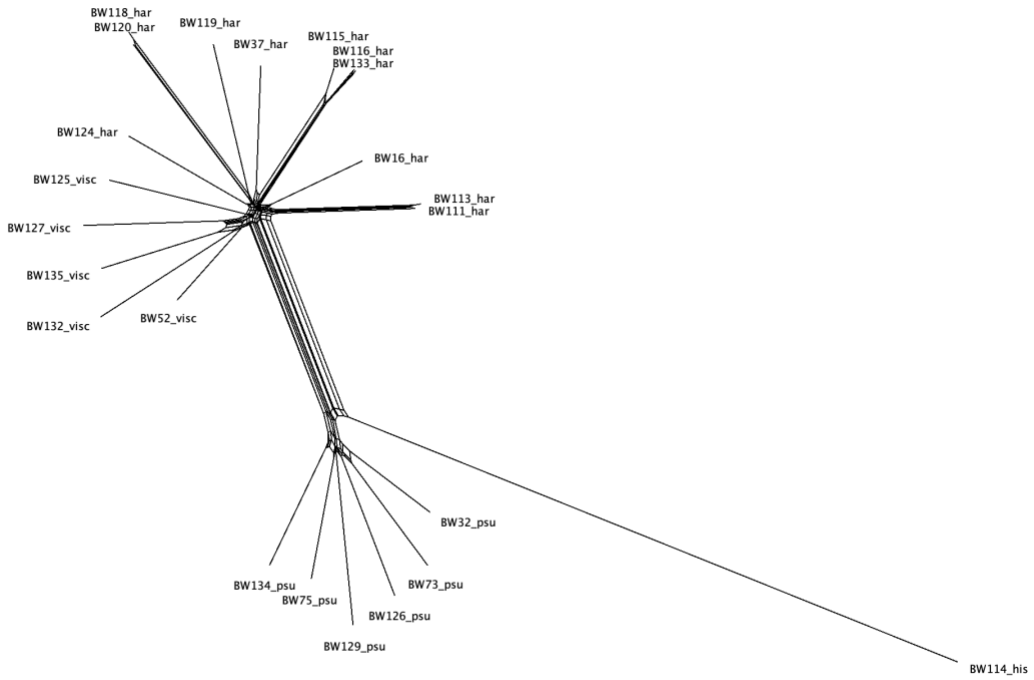

FIGURE 14. Phylogenetic Network. The network was calculated by Splitstree from the unlinked SNPs dataset (only one SNP from each RAD locus). A phylogenetic network allows for multiple paths and overlap between nodes, indicated by secondary lines. Higher numbers of secondary lines indicate a higher potential introgression between those nodes. This analysis indicates broad introgression across the entirety of taxa sampled, as there is a wealth of additional paths between taxa.

ABBA-BABA tests also appeared to show some introgression across the putative species groups, most notably *R. pseudoacacia* and *R. hispida* var. *kelseyi*, which appeared to have a higher rate of introgression than any other combination of species surveyed based on the F4 ratio of trios that included these taxa (Table 4, Figure 15). The non-zero d-statistics, combined with

high z-scores and significant p-values, indicate the strong signature of introgression between these two taxa.

TABLE 4. ABBA-BABA Statistics. ABBA-BABA scores and associated statistics calculated from the min20 dataset comparing potential rates of introgression across a priori assigned species groups. A non-zero D-statistic combined with a z-score greater than 3 is generally thought to indicate significant introgression between the P2 and P3 groups. F4 ratios also indicate high levels of introgression throughout the *Robinia* sampled.

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
hartwigii	pseudo	hispida	0.247912	27.9636	2.30E-16	0.075163	6317.92	3189.03	1921.95
hartwigii	viscosa	hispida	0.00693701	0.939994	0.347221	0.00158138	7919.84	1937.23	1910.54
hartwigii	viscosa	pseudo	0.00736215	1.1388	0.254788	0.010297	4938.12	3063.28	3018.5
viscosa	pseudo	hispida	0.24285	32.1223	2.30E-16	0.0735833	6346.89	3168.66	1930.36

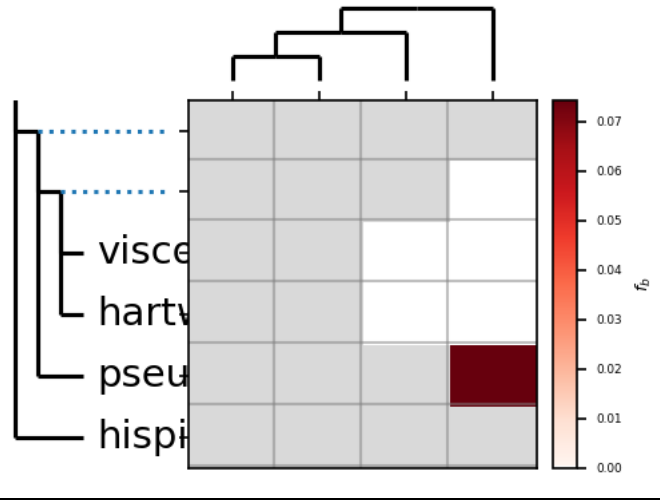

FIGURE 15. Fbranch Introgression Visualization. Visualization of f4 ratios between taxa, arranged phylogenetically. This further highlights the introgression between *Robinia hispida* var. *kelseyi* and *Robinia pseudoacacia*, as these two taxa had much higher f4 ratios than any of the other taxa sampled, despite their distant relationship. More sampling is necessary in order to examine these findings.

SPECIES GROUP VALIDATION

The results from the SNAPPER analysis are listed in Table 5. Based on Bayes Factor, splitting *Robinia hartwigii* as an independent species from *Robinia viscosa* as the treatment that best fits this dataset. This difference in Bayes Factor represents strong support for the splitting of *Robinia hartwigii* from *Robinia viscosa.* These findings are further supported by the visualization of the analysis in DensiTree, which indicates there are no samples that would fall outside of their *a priori* designated taxon (Figure 16).

TABLE 5. Bayes Factor for Delimitation. Likelihood values from SNAPPER species delimitation analysis. Lumped refers to a model that constrains all of the clammy locusts to one species, while split refers to the model where *Robinia hartwigii* is treated as a separate species. The greater Bayes Factor of the Split treatment indicates support for the treatment of *Robinia hartwigii* as a distinct taxon.

Treatment	Likelihood	Bayes Factor
Lumped	-370812.7988310179	-9853.5015
Split	-365886.0480793803	9853.501503

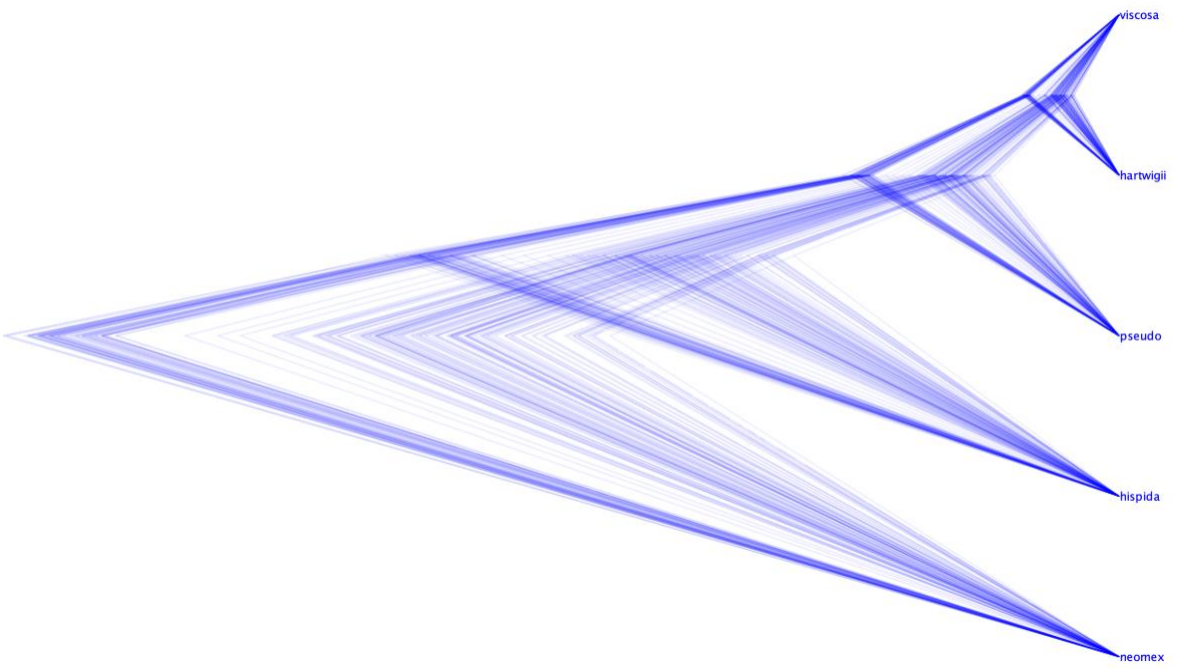

FIGURE 16. Coalescent Trees from SNAPPER Analysis. Visualized using DensiTree. In this analysis samples that did not cohere to the *a priori* species framework would be a different color, but all individuals sampled fell within their correctly assigned taxon. This lack of alternative topologies supports the continued recognition of *R. hartwigii* as a species.

MORPHOLOGICAL COMPARISONS

Utilizing quantitative methods, there appears to be little separation between *Robinia hartwigii* and *Robinia viscosa* in either quantifiable leaf or flower characters (Figure 14 and 15). Qualitatively, flowers of *Robinia hartwigii* tend to have calyxes that maintain the raised glands characteristic of the species, while *Robinia viscosa* lacks any vestiture on the calyx. Flowers of *Robinia hartwigii* also tend to be a darker pink color while *Robinia viscosa* tended to be lighter in color, but this was highly variable between populations and may not be a reliable character. In addition, *Robinia viscosa* and *Robinia hartwigii* set fruit with relatively equal likelihood and vigor across surveyed populations and I did not find fruit set to be reliable character for species

identification, as previously described. Additionally, fruit set varied wildly between 2021 and 2022 across both species, reducing the viability of this character for reliable identification of these taxa.

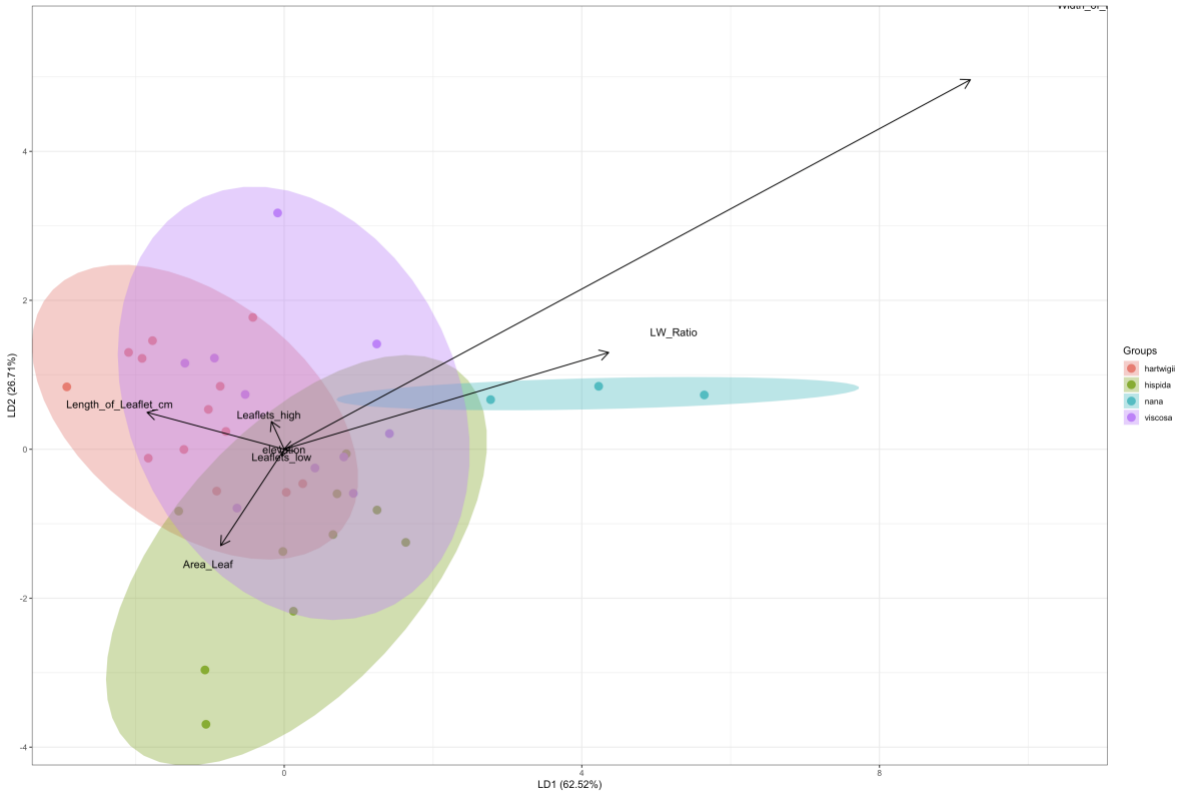

FIGURE 17. LDA of Vegetative Characters. Linear Discriminant Analysis of all *Robinia* species sampled in this work, including herbarium specimens. There appears to be few of the surveyed quantitative factors that effectively separate *R. hartwigii* and *R. viscosa* in the broader context of the genus, but there are some general trends that can be seen among taxa.

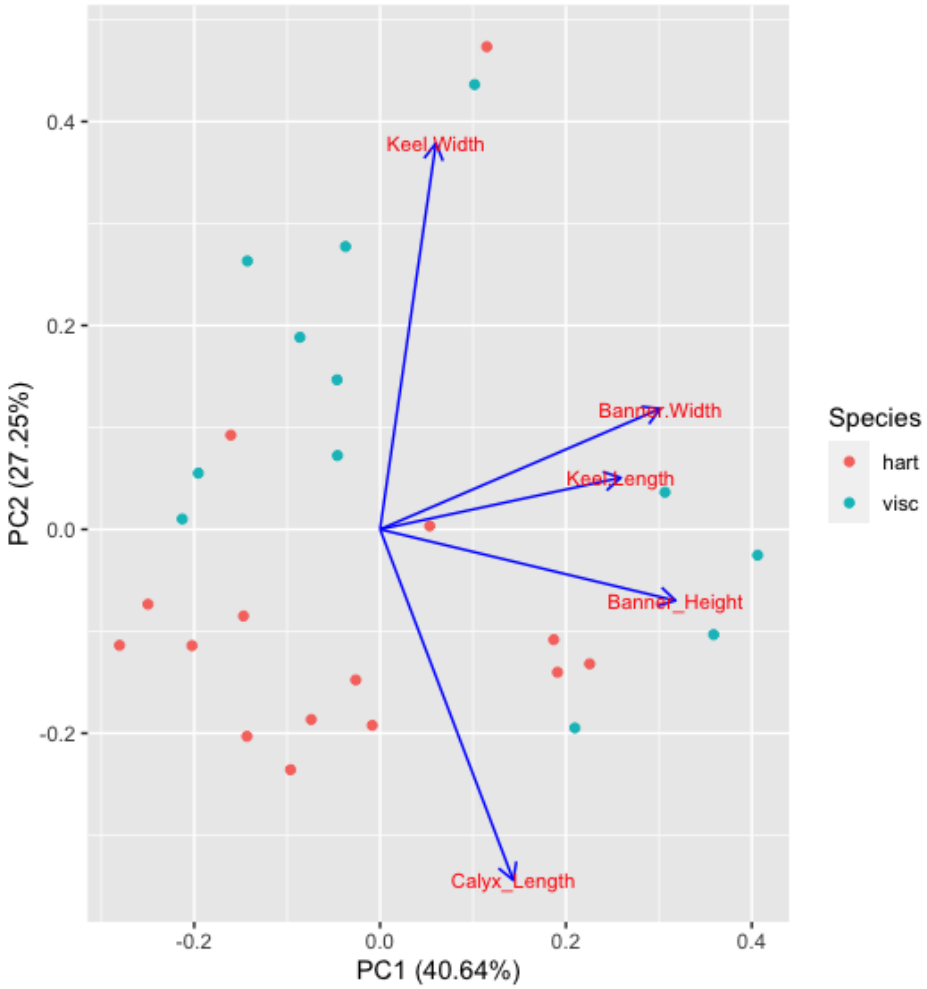

FIGURE 18. PCA of Floral Characters. Principal component analysis of floral characters in *Robinia hartwigii* and *Robinia viscosa*. there do not appear to be any floral characters measured in this work that effectively delineate between *R. hartwigii* and *R. viscosa.*

CHAPTER FOUR: DISCUSSION

In order to discern the most accurate treatment of *Robinia hartwigii* and *Robinia viscosa* it was necessary to obtain a wealth of genetic data and then analyze that dataset through a number of lenses in order to generate a number of arguments for species delimitation. Using the generalized species concept, these variable outcomes can be used as argument for and against recognition of the focal taxa as individual species. These methods, as is to be expected in almost any modern phylogenetic study, did not coalesce entirely upon the same result. Importantly, this lack of concordance among methods can be synthesized into an updated taxonomic treatment under the generalized species concept, under which I am able to parse through each method of delimitation and group assignment in order to generate an argument for either maintaining *R. hartwigii* as a distinct species or returning it to treatment as a variety of *R. viscosa.*

In terms of reciprocal monophyly, the two putative taxa are monophyletic with high values of bootstrap support for the nodes that correspond to each putative species group. These findings are strengthened by the coalescent methods and quartet inference scores, both reporting high support and few discordant patterns between *R. hartwigii* and *R. viscosa*. Strong bootstrap values at the crown nodes of each species are a strong argument for their isolation and lack of interbreeding in their native ranges and recognition of distinction. These bootstrap values are further supported by the quartet sampling analysis, which also strongly supports the crown node of each putative species group. This lack of variation between the bootstrap and quartet sampling analysis seems to indicate that the large number of RAD loci used are resistant to the effects of introgression within this lineage. While some loci may not follow a purely linear evolutionary pathway, they are heavily outweighed by the loci that do follow the expected phylogeny. Additionally, it appears incomplete lineage sorting does not significantly impact the phylogeny

of this group, as both the coalescent and gene-tree analysis result in highly supported nodes and an identical topology across the phylogeny.

In contrast to the phylogenetic analyses, species group discovery methods were somewhat mixed when applied to the question of lumping or splitting the clammy locusts. Principal Component Analysis, as well as Discriminant Analysis of Principal Components, failed to represent *Robinia hartwigii* as distinct from *Robinia viscosa*. The results of the PCA were somewhat surprising, as *Robinia hartwigii* was very broadly clustered compared to the other groups in the analysis. When limiting the analyses to only *Robinia viscosa* and *Robinia hartwigii,* the resulting figure was even more scattered, as *Robinia hartwigii* varied from the sister taxon along both axes, highly varying from the tightly clustered *Robinia viscosa.* This lack of coherence could be due to genetic drift related to small sample sizes or introgression from other members of the *Robinia* genus. This variation was reduced by the DAPC, which reduced *Robinia hartwigii* to tightly overlap *Robinia viscosa*- an argument for the inclusion of *Robinia hartwigii* within *R. viscosa.* STRUCTURE results were also conflicting, only splitting *Robinia hartwigii* and *Robinia viscosa* when they were the only two groups in the analysis. When populations of *Robinia pseudoacacia* and *Robinia hispida* were added to the analysis, the algorithm no longer split *R. hartwigii* and *viscosa,* instead lumping them into one historical cluster. This could potentially be due to a high variance between the clammy locusts (*R. hartwigii* and *R. viscosa*) and the rest of the genus and the relatively low variance between the two clammy locusts. This could indicate only weak reproductive isolation between the two focal species compared to the rest of the species within the genus, which could be further isolated than the two focal species but further study, with methods that model demography, is necessary in order to confirm these initial findings.

Introgression is necessary to consider in studies of species delimitation, particularly when delimiting species of plants, as it is estimated that 25% of all plant species are able to generate viable hybrid offspring (Mallet 2005). Previous study has shown that RAD-Seq methodology is somewhat resistant to phylogenetic clouding by introgression as the high number of putative loci tends to overpower the potential influence of introgression (Yang et al., 2022). Introgression was successfully detected within *Robinia*. *Robinia hispida* var. *kelseyi* and *Robinia pseudoacacia* seem to be introgressed, or even of hybrid origin, sharing far more genetic material with each other than with *Robinia hartwigii* or *Robinia viscosa*. While this study casts doubt on the long-hypothesized hybrid origins of *Robinia hartwigii,* it does appear to indicate *Robinia hispida* var. *kelseyi* may be of hybrid origin between *Robinia pseudoacacia* and a variety of *Robinia hispida,* though more work is necessary in order to confirm these findings. After centuries of speculation relating to species of hybrid origin within this genus, this is the first molecular examination of the previously hypothesized hybrid origins of *Robinia hartwigii* and *Robinia hispida* var. *kelseyi*. Further examination of the molecular dataset collected in this study is necessary in order to confirm these hybrid origins, but the findings of the ABBA-BABA tests and phylogenetic networks indicate introgression is likely common within the *Robinia* genus. In order to fully discern the origins of these species, it will be necessary to incorporate a reference genome into the RAD Library generation or to utilize a pseudo-reference approach in order to reduce the number of orthologs when investigating the phylogeny and utilize a sophisticated phylogenetic network analysis that allows for species of hybrid origin and calculates the likelihood of this possibility (Blanco-Pastor et al., 2019; Wang et al., 2021).

This work failed to identify quantifiable differences in both the vegetative and floral characters surveyed that can be used to easily split *Robinia hartwigii* and *Robinia viscosa.*

Thankfully, many categorical characters are presented for ease of identification in both flowering and vegetative states. These characters include arguably the most charismatic difference between the two species: the position of their viscose glands. Additionally, the calyx of *Robinia hartwigii* tends to have small (~0.1 mm in length) raised glands on their surface, while the calyx of *Robinia viscosa* lacks any glands. *R. hartwigii* also tended to have deeper calyx lobes than those of *R. viscosa*. The flowers of *Robinia viscosa* also tended to be paler in color, while *Robinia viscosa* tended towards a darker pink/magenta. For this work I visited every known, supposed natural, population of *Robinia hartwigii* and *Robinia viscosa* in the Southeastern United States as well as extensive study of herbarium material in order to identify these qualitative characters and attempt to identify delineating quantitative characters.

From this collection of results, it appears that *Robinia hartwigii*, if we are to treat it as such, has potentially diverged recently from *Robinia viscosa*, leading to the discordance among methods due to potentially porous barriers to reproduction between the species surveyed. When examining these taxa, we are likely seeing two taxa in the process of creating two distinct lineages, leading to the lack of coherence among methods in this study. Although not all methods support this distinction, it is clear that the overall trend of the results support the *Robinia hartwigii* a distinct lineage worthy of recognition at the species level.

In this work, I have utilized molecular and morphological methods in order to understand delimitation of the clammy locusts of the Southern Appalachians. This work has provided ample support to recognize *Robinia hartwigii* and *Robinia viscosa* as distinct species. These methods have described *Robinia hartwigii* as monophyletic and highlighted the genetic and morphological distinction of the taxon from *Robinia viscosa*. Accurate circumscription of both of these rare taxa allows for their continued protection on the landscape. As both *Robinia hartwigii*

and *Robinia viscosa* are rare in their native range, an accurate treatment allows for improved record keeping and an updated understanding of the ecological and environmental requirements to maintain populations of these species in their native range.

REFERENCES

Ahles HE, Radford AE, Bell C. Ritchie. 1968. Manual of the Vascular Flora of the Carolinas. :605–607.

Ashe WW. 1922. The Eastern Shrubby Species of *Robinia*. Journal of the Elisha Mitchell Scientific Society. 37(3):175–177. [accessed 2022 Dec 28]. https://www.jstor.org/stable/24331491#metadata_info_tab_contents.

Blanco-Pastor JL, Bertrand YJK, Liberal IM, Wei Y, Brummer EC, Pfeil BE. 2019. Evolutionary Networks from RADseq Loci Point to Hybrid Origins of *Medicago carstiensis* and *Medicago cretacea.* Am J Bot. 106(9):1219–1228. doi:10.1002/AJB2.1352.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 10(4). doi:10.1371/JOURNAL.PCBI.1003537.

Bouckaert RR. 2010. DensiTree: Making Sense of Sets of Phylogenetic Trees. Bioinformatics. 26(10):1372–1373. doi:10.1093/BIOINFORMATICS/BTQ110. [accessed 2023 Mar 16]. https://academic.oup.com/bioinformatics/article/26/10/1372/192963.

Brown J, Pirrung M, Mccue LA. 2017. FQC Dashboard: Integrates FastQC Results into a Web-Based, Interactive, and Extensible FASTQ Quality Control Tool. Bioinformatics. 33(19):3137–3139. doi:10.1093/BIOINFORMATICS/BTX373.

Bryant D, Moulton V. 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Mol Biol Evol. 21(2):255–265. doi:10.1093/MOLBEV/MSH018.

Burda RI, Koniakin SN. 2019. The Non-Native Woody species of the flora of Ukraine: Introduction, naturalization, and invasion. Biosyst Divers. 27(3):276–290. doi:10.15421/011937.

Burge DO, Parker VT, Mulligan M, Valderamma CG. 2018. Conservation Genetics of the Endangered Del Mar Manzanita (*Arctostaphylos glandulosa* subsp. *crassifolia*) Based on Rad Sequencing Data. https://doi-org.proxy195.nclive.org/103120/0024-9637-653117. 65(3):117–130. doi:10.3120/0024-9637-65.3.117. [accessed 2021 Oct 17]. https://bioone-org.proxy195.nclive.org/journals/madroño/volume-65/issue-3/0024-9637-65.3.117/Conservation-Genetics-of-the-Endangered-Del-Mar-Manzanita-Arctostaphylos-glandulosa/10.3120/0024-9637-65.3.117.full.

Candy J, Campbell N, Grinnell M, Beacham T, Narum S. 2015. Population Differentiation Determined from Putative Neutral and Divergent Adaptive Genetic Markers In Eulachon (*Thaleichthys pacificus*, Osmeridae), an Anadromous Pacific Smelt. Mol Ecol Resour. 15(6):1421–1434. doi:10.1111/1755-0998.12400. [accessed 2021 Oct 30]. https://pubmed.ncbi.nlm.nih.gov/25737187/.

Chandler GT, Crisp MD. 1998. Plant Systematics and Evolution Morphometric and Phylogenetic Analysis of the *Daviesia ulicifolia* Complex (Fabaceae, Mirbelieae). P1 Syst Evol. 209:93–122.

Chattopadhyay B, Garg KM, Ramakrishnan U. 2014. Effect Of Diversity and Missing Data On Genetic Assignment With RAD-Seq Markers. BMC Res Notes. 7(1):1–3. doi:10.1186/1756-0500-7-841/FIGURES/1. [accessed 2023 Mar 2]. https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-7-841.

Chifman J, Kubatko L. 2014a. Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics. 30(23):3317–3324. doi:10.1093/BIOINFORMATICS/BTU530. [accessed 2022 Nov 30]. https://academic.oup.com/bioinformatics/article/30/23/3317/206559.

Cohen DH, Schenk JJ. 2022. Investigating Phylogenetic Placement and Species-Level Relationships in a Recent Radiation of *Mentzelia* section Bartonia (Loasaceae) from the Mojave Desert. doi:10.1600/036364422X16442668423455.

Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol. 37(1):291–294. doi:10.1093/MOLBEV/MSZ189. [accessed 2022 Oct 5]. https://academic.oup.com/mbe/article/37/1/291/5552155.

Davey JW, Blaxter ML. 2010. RADSeq: Next-Generation Population Genetics. Brief Funct Genomics. 9(5–6):416–423. doi:10.1093/BFGP/ELQ031. [accessed 2021 Oct 30]. https://academic.oup.com/bfg/article/9/5-6/416/182576.

Dixon P. 2003. VEGAN, A Package of R Functions for Community Ecology. Journal of Vegetation Science. 14(6):927–930. doi:10.1111/J.1654-1103.2003.TB02228.X. [accessed 2023 Feb 26]. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1654-1103.2003.tb02228.x.

Donoghue MJ. 1985. A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative. Bryologist. 88(3):172. doi:10.2307/3243026.

Doyle J. 1991. DNA Protocols for Plants. Molecular Techniques in Taxonomy.:283–293. doi:10.1007/978-3-642-83962-7_18.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for Ancient Admixture between Closely Related Populations. Mol Biol Evol. 28(8):2239–2252.

doi:10.1093/MOLBEV/MSR048. [accessed 2022 Oct 9]. https://academic.oup.com/mbe/article/28/8/2239/1052492.

Eaton DAR, Overcast I. 2020. ipyrad: Interactive Assembly and Analysis of RADseq Datasets. Bioinformatics. 36(8):2592–2594. doi:10.1093/BIOINFORMATICS/BTZ966.

Eaton DAR, Ree RH. 2013. Inferring Phylogeny And Introgression Using RADseq data: An Example From Flowering Plants (Pedicularis: Orobanchaceae). Syst Biol. 62(5):689–706. doi:10.1093/sysbio/syt032.

Eaton DAR, Spriggs EL, Park B, Donoghue MJ. 2017. Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. Syst Biol. 66(3):399–412. doi:10.1093/SYSBIO/SYW092.

Edler D, Klein J, Antonelli A, Silvestro D. 2021. raxmlGUI 2.0: A Graphical Interface and Toolkit For Phylogenetic Analyses Using RAxML. Methods Ecol Evol. 12(2):373–377. doi:10.1111/2041-210X.13512.

Evanno G, Regnaut S, Goudet J. 2005. Detecting The Number Of Clusters Of Individuals Using The Software Structure: A Simulation Study. Mol Ecol. 14(8):2611–2620. doi:10.1111/J.1365-294X.2005.02553.X.

Evans BJ, Zeng K, Esselstyn JA, Charlesworth B, Melnick DJ. 2014. Reduced Representation Genome Sequencing Suggests Low Diversity on the Sex Chromosomes of Tonkean Macaque Monkeys. Mol Biol Evol. 31(9):2425–2440. doi:10.1093/MOLBEV/MSU197.

Fellers JP. 2008. Genome Filtering Using Methylation-Sensitive Restriction Enzymes with Six Base Pair Recognition Sites. Plant Genome. 1(2). doi:10.3835/PLANTGENOME2008.05.0245.

FigTree. [accessed 2023 Mar 16]. http://tree.bio.ed.ac.uk/software/figtree/.

Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork VL. 2017. Phylogenomic Inferences from Reference-Mapped And De Novo Assembled Short-Read Sequence Data Using RADseq Sequencing Of California White Oaks (*Quercus* section Quercus)1. https://doi.org/101139/gen-2016-0202. 60(9):743–755. doi:10.1139/GEN-2016-0202.

Francis RM. 2017. pophelper: An R Package and Web App To Analyse And Visualize Population Structure. Mol Ecol Resour. 17(1):27–32. doi:10.1111/1755-0998.12509.

Grummer JA, Bryson RW, Reeder TW. 2014. Species Delimitation Using Bayes Factors: Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata: Phrynosomatidae). Syst Biol. 63(2):119–133. doi:10.1093/SYSBIO/SYT069.

Isely D, Peabody FJ. 1984. *Robinia* (Leguminosae: Papilionoidea) . Castanea. 49(4):187–202. [accessed 2021 Nov 20]. https://www-jstor-

org.proxy195.nclive.org/stable/4033299?Search=yes&resultItemClick=true&searchText=duane+isely+robinia&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dduane%2Bisely%2Brobinia%26so%3Drel&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3Aa27e9a65e89f918175a2aab9efc3c5c9&seq=1#metadata_info_tab_contents.

Jombart T, Bateman A. 2008. adegenet: A R Package for The Multivariate Analysis of Genetic Markers. Bioinformatics. 24(11):1403–1405. doi:10.1093/BIOINFORMATICS/BTN129.

Jombart T, Devillard S, Balloux F. 2010. Discriminant Analysis of Principal Components: A New Method For The Analysis Of Genetically Structured Populations. BMC Genet. 11(1):1–15. doi:10.1186/1471-2156-11-94/FIGURES/9.

Karbstein K, Tomasello S, Hodač L, Wagner N, Marinček P, Barke BH, Pätzold C, Hörandl E. 2021 Aug 31. Unraveling Phylogenetic Relationships, Reticulate Evolution, and Genome Composition of Polyploid Plant Complexes by Rad-Seq and Hyb-Seq. bioRxiv.:2021.08.30.458250. doi:10.1101/2021.08.30.458250. [accessed 2022 Nov 21]. https://www.biorxiv.org/content/10.1101/2021.08.30.458250v1.

Koehne VE. 1913. Eine neue Robinie. Mitteilungen der Deutschen Dendrologischen. 22.

Konarska A, Łotocka B. 2020. Glandular Trichomes of *Robinia viscosa* Vent. var. *hartwigii* (Koehne) Ashe (Faboideae, Fabaceae)—Morphology, Histochemistry And Ultrastructure. Planta. 252(6):1–19. doi:10.1007/S00425-020-03513-Z/FIGURES/8.

Lavin M, Wojciechowski MF, Gasson P, Hughes C, Wheeler E. 2003. Phylogeny of Robinioid Legumes (Fabaceae) Revisited: *Coursetia* and *Gliricidia* Recircumscribed, and a Biogeographical Appraisal of the Caribbean Endemics. Syst Bot. 28(2):387–409.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR. 2014. Species Delimitation using Genome-Wide SNP Data. Syst Biol. 63(4):534–542. doi:10.1093/SYSBIO/SYU018.

Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's Phylogenetic Bootstrap in The Era of Big Data. Nature 2018 556:7702. 556(7702):452–456. doi:10.1038/s41586-018-0043-0.

Maddison W, Knowles L. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. Syst Biol. 55(1):21–30. doi:10.1080/10635150500354928.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite - Fast D-statistics and Related Admixture Evidence from VCF Files. Mol Ecol Resour. 21(2):584–595. doi:10.1111/1755-0998.13265.

Mallet J. 2005. Hybridization As an Invasion of The Genome. Trends Ecol Evol. 20(5):229–237. doi:10.1016/J.TREE.2005.02.010.

Massatti R, Reznicek AA, Knowles LL. 2016. Utilizing RADseq Data for Phylogenetic Analysis Of Challenging Taxonomic Groups: A Case Study in *Carex* sect. Racemosae. Am J Bot. 103(2):337–347. doi:10.3732/ajb.1500315.

Mayr E. 1942. Systematics And the Origin of Species. New York: Columbia University Press.

Michener CD, Sokal RR. 1956. A Quantitative Approach to A Problem in Classification 1.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in Ecological and Conservation Genomics. Mol Ecol. 22(11):2841–2847. doi:10.1111/mec.12350.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. Genetics. 192(3):1065–1093. doi:10.1534/GENETICS.112.145037/-/DC1.

Peabody FJ. 1984. Revision of the Genus *Robinia* (Leguminosae: Papilionoideae). Iowa State University.

Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. 2018. Quartet Sampling Distinguishes Lack of Support from Conflicting Support in The Green Plant Tree of Life. Am J Bot. 105(3):385–403. doi:10.1002/AJB2.1016.

Perkins MT, Zhebentyayeva T, Sisco PH, Craddock JH. 2021. Genome-Wide Sequence-Based Genotyping Supports a Nonhybrid Origin of *Castanea alabamensis*. https://doi.org/101600/036364421X16370109698524. 46(4):973–984. doi:10.1600/036364421X16370109698524.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 155(2):945–959. doi:10.1093/GENETICS/155.2.945.

Revell LJ. 2012. phytools: An R Package for Phylogenetic Comparative Biology (and other things). Methods Ecol Evol. 3(2):217–223. doi:10.1111/J.2041-210X.2011.00169.X.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 61(3):539–542. doi:10.1093/SYSBIO/SYS029.

Spriggs EL, Eaton DAR, Sweeney PW, Schlutius C, Edwards EJ, Donoghue MJ. 2019. Restriction-Site-Associated DNA Sequencing Reveals a Cryptic *Viburnum* Species on the North American Coastal Plain. Syst Biol. 68(2):187–203. doi:10.1093/sysbio/syy084. [accessed 2022 Oct 5]. https://academic.oup.com/sysbio/article/68/2/187/5232147.

Stamatakis A. 2014. RAxML version 8: A Tool for Phylogenetic Analysis and Post-Analysis Of Large Phylogenies. Bioinformatics. 30(9):1312–1313. doi:10.1093/BIOINFORMATICS/BTU033.

Stoltz M, Baeumer B, Bouckaert R, Fox C, Hiscott G, Bryant D. 2021. Bayesian Inference of Species Trees using Diffusion Models. Syst Biol. 70(1):145–161. doi:10.1093/SYSBIO/SYAA051.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 38(7):3022–3027. doi:10.1093/MOLBEV/MSAB120.

Team RCP. 2020. R: A Language and Environment for Statistical Computing.

Venables WN, Ripley BD. 1997. Statistics Complements to Modern Applied Statistics with S-Plus Second edition. [accessed 2023 Apr 2]. http://www.stats.ox.ac.uk/pub/MASS2/.

Ventenat EP. 1799. Description Des Plantes Nouvelles Et Peu Connues. In: Description Des Plantes Nouvelles Et Peu Connues : Cultivées Dans Le Jardin de J.M. Cels : avec figures /. Paris, France: De l'imprimerie de Crapelet,. p. 15–20.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. 2013. Genome-wide RAD Sequence Data Provide Unprecedented Resolution of Species Boundaries and Relationships In The Lake Victoria Cichlid Adaptive Radiation. Mol Ecol. 22(3):787–798. doi:10.1111/MEC.12023.

Wang N, Kelly LJ, McAllister HA, Zohren J, Buggs RJA. 2021. Resolving Phylogeny and Polyploid Parentage Using Genus-Wide Genome-Wide Sequence Data from Birch Trees. Mol Phylogenet Evol. 160. doi:10.1016/J.YMPEV.2021.107126.

Weakley AS. 2020. Flora of the Southeastern United States. North Carolina Botanical Garden.

Weisrock DW, Smith SD, Chan LM, Biebouw K, Kappeler PM, Yoder AD. 2012. Concatenation and Concordance in the Reconstruction of Mouse Lemur Phylogeny: An Empirical Demonstration of the Effect of Allele Sampling in Phylogenetics. Mol Biol Evol. 29(6):1615–1630. doi:10.1093/MOLBEV/MSS008.

Whitakker TW. 1934. The Shrubby Robinias. Bulletin of Popular Information . 2(6):25–28.

Whittaker TW. 1934. A Karyo-Systematic Study of *Robinia*. Journal of the Arnold Arboretum. 15(4):353–357. [accessed 2021 Jul 26]. https://www.jstor.org/stable/43782263?seq=1#metadata_info_tab_contents.

Wichmann B. 2021. Natural Heritage Program List of Rare Plant Species of North Carolina. Raleigh, NC.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis.

Wilgenbusch JC, Swofford D. 2003. Inferring Evolutionary Trees with PAUP*. Curr Protoc Bioinformatics. 00(1):6.4.1-6.4.28. doi:10.1002/0471250953.BI0604S00.

Yang YT, Yang XC, Wang MC, Zhong LL, Ma R, Ma T, Liu JQ, Davis CC, Xi ZX. 2022. Species Delimitation of North American *Nyssa* Species. J Syst Evol. 60(4):747–758. doi:10.1111/JSE.12725/SUPPINFO.

APPENDICES

APPENDIX A. HERBARIUM SPECIMENS USED IN MORPHOLOGICAL ANALYSIS IN ADDITION TO AUTHOR'S COLLECTIONS

Acc#	Length_of_Leaflet_cm	Width_of_Leaflet_cm	LW_Ratio	Area_Leaf	Leaflets_low	Leaflets_high	species
NCU00007724	3.941	2.034	1.93756146	6.03	7	9	hispida
NCU00007727	3.216	1.821	1.7660626	4.425	11	15	nana
NCU00007728	1.92	1.068	1.79775281	1.668	11	15	hispida
NCU00007734	2.427	1.386	1.75108225	2.785	11	17	hartwigii
NCU00007735	3.243	1.977	1.64036419	5.141	13	19	hartwigii
NCU00007736	2.264	1.764	1.28344671	1.129	13	21	hartwigii
NCU00007737	3.367	1.544	2.18069948	3.971	19	25	viscosa
NCU00007737	3.861	1.649	2.34141904	4.71	19	21	viscosa
NCU00007739	3.079	1.505	2.04584718	3.512	11	19	viscosa
NCU00012584	2.957	1.603	1.84466625	3.689	17	21	hartwigii
NCU00020454	3.593	1.536	2.33919271	4.378	19	25	viscosa
NCU00031712	2.064	1.175	1.75659575	1.828	17	23	viscosa
NCU00059938	2.366	1.197	1.97660819	2.217	17	21	hartwigii

NCU00059939	2.655	1.24	2.14112903	2.563	17	21	hartwigii
NCU00059940	2.646	1.584	1.67045455	3.194	9	17	hartwigii
NCU00059941	2.452	1.382	1.77424023	2.513	15	17	hartwigii
NCU00059944	3.044	1.302	2.33794163	3.224	11	17	hartwigii
NCU00059945	4.018	2.261	1.77708978	6.557	19	21	hartwigii
NCU00059947	2.576	1.108	2.32490975	2.353	13	21	hartwigii
NCU00059950	3.879	1.758	2.20648464	5.017	13	17	hartwigii
NCU00059951	2.462	1.344	1.83184524	2.462	13	15	hartwigii
NCU00059952	3.435	1.765	1.94617564	4.716	15	19	hartwigii
NCU00059960	2.408	1.189	2.02523129	2.273	11	15	hartwigii
NCU00059964	3.155	1.573	2.00572155	3.854	14	21	hartwigii
NCU00059965	2.781	1.33	2.09097744	3.015	NA	NA	viscosa
NCU00059968	2.789	1.384	2.01517341	2.85	11	13	viscosa
NCU00059968	2.628	1.504	1.74734043	3.534	13	19	viscosa
NCU00059969	3.866	2.049	1.88677404	6.301	15	17	hartwigii
NCU00059969	3.462	2.093	1.65408505	5.705	13	17	hartwigii
NCU00060175	3.276	1.09	3.00550459	2.838	15	19	viscosa
NCU00060175	3.651	1.351	2.70244264	3.742	13	19	viscosa

NCU00060180	5.141	2.116	2.42958412	7.322	13	19	viscosa
NCU00060184	1.707	0.702	2.43162393	1.358	NA	NA	viscosa
NCU00060185	2.95	1.638	1.8009768	4.208	9	15	viscosa
NCU00060189	3.778	1.563	2.41714651	4.55	11	15	viscosa
NCU00060190	3.341	1.285	2.6	3.631	9	18	viscosa
NCU00060193	3.532	2.192	1.61131387	6.243	8	19	viscosa
NCU00060195	2.277	1.388	1.64048991	2.447	15	19	viscosa
NCU00063124	2	0.9	2.22222222	NA	11	17	viscosa
NCU00063125	2.106	1.092	1.92857143	1.882	15	21	hartwigii
NCU00090364	1.897	0.98	1.93571429	1.488	11	15	viscosa
NCU00104231	3.118	2.076	1.50192678	5.134	NA	NA	viscosa
NCU00104232	3.258	1.496	2.17780749	3.732	NA	NA	viscosa
NCU00104233	2.447	1.102	2.22050817	2.142	11	17	pseudoacacia
NCU00110096	3.644	1.314	2.77321157	3.947	9	15	viscosa
NCU00306690	3.048	1.986	1.5347432	4.031	11	13	nana
NCU00306732	4.107	2.106	1.95014245	5.908	13	19	hartwigii
NCU00306769	3.198	1.775	1.80169014	4.582	11	15	hartwigii
NCU00306803	2.642	2.377	1.11148507	5.129	9	13	nana
NCU00306806	5.667	3.607	1.57111173	10.415	5	9	nana

NCU00306827	3.918	2.336	1.67722603	7.685	9	13	nana
NCU00311224	3.708	2.632	1.40881459	7.55	7	9	hispida
NCU00311225	2.66	1.25	2.128	2.544	13	15	hispida
NCU00311226	4.385	2.36	1.85805085	8.588	11	15	hispida
NCU00311228	2.833	1.882	1.5053135	3.972	9	13	hispida
NCU00311230	1.407	0.57	2.46842105	0.714	9	11	hispida
NCU00311231	5.759	2.371	2.42893294	10.635	7	13	hispida
NCU00311232	3.55	2.157	1.64580436	6.083	9	13	hispida
NCU00311246	4.057	1.789	2.26774735	5.139	9	12	hispida
NCU00311247	2.944	1.841	1.59913091	4.423	9	13	hispida
NCU00311248	4.37	3.067	1.42484513	8.987	7	11	hispida
NCU00311249	4.079	2.599	1.56944979	8.11	NA	NA	hispida
NCU00311250	2.89	1.662	1.73886883	3.866	7	11	hispida
NCU00311251	5.394	3.067	1.75872188	13.07	13	13	hispida
NCU00311252	3.395	1.899	1.78778304	5.122	NA	NA	hispida
NCU00311252	3.025	1.639	1.84563758	4.132	NA	NA	hispida
NCU00311254	3.913	1.658	2.36007238	5.152	9	13	hispida
NCU00311255	3.534	2.117	1.66934341	5.966	11	13	hispida
NCU00311257	4.354	3.06	1.42287582	10.615	9	11	hispida

NCU00311259	3.595	2	1.9021164	5.328	7	7	hispida
NCU00311260	3.594	1.566	2.29501916	4.506	9	15	hispida
NCU00311262	4.831	3.644	1.32574094	14.453	9	11	hispida
NCU00311263	4.955	2.184	2.26877289	8.048	11	11	hispida
NCU00311264	2.973	1.422	2.0907173	3.323	15	17	hispida
NCU00311265	3.069	1.574	1.9498094	3.642	9	15	hispida
NCU00311266	3.065	1.396	2.19555874	3.5	13	17	hispida
NCU00311267	4.113	2.78	1.4794964	9.16	9	11	hispida
NCU00311268	2.664	1.423	1.8721012	2.977	NA	NA	hispida
NCU00427548	2.283	1.436	1.58983287	NA	9	11	viscosa
WCUH0012403	NA	NA	NA	NA	9	19	hartwigii
WCUH0012427	NA	NA	NA	NA	13	19	viscosa
WCUH0012428	NA	NA	NA	NA	13	17	hartwigii
WCUH0012429	NA	NA	NA	NA	11	15	viscosa
WCUH0012431	4.46	1.803	2.47365502	6.068	11	13	pseudoacacia
WCUH0012432	3.816	1.707	2.23550088	5.114	11	17	pseudoacacia
WCUH0012433	3.558	1.706	2.08558031	4.84	11	15	pseudoacacia
WCUH0012434	3.451	2.482	1.39041096	6.762	11	13	hartwigii
WCUH0012436	4.343	2.255	1.92594235	7.16	7	15	pseudoacacia
WCUH0012440	4.432	2.513	1.76362913	8.503	9	19	pseudoacacia

WCUH0012441	NA	NA	NA	NA	15	23	pseudoacacia
WCUH0012446	4.538	2.634	1.72285497	9.208	7	11	pseudoacacia
WCUH0012448	3.13	1.392	2.24856322	3.96	13	15	pseudoacacia
WCUH0012449	2.302	1.191	1.93282956	1.999	11	13	pseudoacacia
WCUH0012457	NA	NA	NA	NA	11	13	pseudoacacia
WCUH0029161	3.264	1.494	2.18473896	3.884	11	15	pseudoacacia

APPENDIX B. MEASUREMENTS OF FRESH COLLECTED FLOWERS

Location	Species	Calyx Length	Vest	Banner Height	Banner Width	Keel Length	Keel Width	Color
CUCR	visc	7	Hair	11	16	13	9	White
SATU	hart	5	Glands and Hair	13	14	15	8	Pink
WTSD	hart	8	Glands and Hair	10	14	15	5	White/Pink
WTSD	hart	9	Glands and Hair	11	12	11	4	White/Pink
WTSD	hart	9	Glands and Hair	12	9	13	5	White/Pink
WTSD	hart	6	Glands and Hair	10	10	14	5	White/Pink
WTSD	hart	9	Glands and Hair	12	9	11	5	White/Pink
WTSD	hart	9	Glands and Hair	11	9	10	6	White/Pink
WTSD	hart	7	Glands and Hai	10	9	11	4	White/Pink

			r					
WTSD	hart	8	Glands and Hair	9	11	10	5	White/Pink
WTSD	hart	7	Glands and Hair	11	11	12	4	White/Pink
ROCR	visc	6	Hair	14	12	11	6	White
ROCR	visc	6	Hair	10	12	11	4	White
ROCR	visc	7	Hair	12	13	11	8	White
ROCR	visc	7	Hair	10	15	12	5	White
ROCR	visc	6	Hair	11	12	11	7	White
ROCR	visc	6	Hair	12	11	10	5	White
ROCR	visc	7	Hair	12	14	9	7	White
CUCR	visc	9	Hair	14	15	14	4	White
CUCR	visc	9	Hair	16	16	15	5	White
CUCR	visc	7	Hair	17	16	14	5	White
CUCR	visc	9	Hair	17	16	15	6	White
SATU	hart	9	Glands and Hair	14	11	11	5	Pink
SATU	hart	9	Glands and Hair	15	14	14	5	Pink
SATU	hart	10	Glands and Hai	14	16	11	6	Pink

			r					
SATU	hart	8	Glands and Hair	13	15	9	4	Pink
SATU	hart	9	Glands and Hair	15	14	13	5	Pink