EVOLUTION OF PHYTOCHEMICAL DIVERSITY IN *PILOCARPUS* (RUTACEAE) USING A COMBINED PHYLOGENETIC AND ENVIRONMENTAL ANALYSIS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Daniella Meyer Allevato

May 2018

EVOLUTION OF PHYTOCHEMICAL DIVERSITY IN *PILOCARPUS* (RUTACEAE) USING A COMBINED PHYLOGENETIC AND ENVIRONMENTAL ANALYSIS

Daniella Meyer Allevato, Ph.D.

Cornell University 2018

The diversity of plant specialized metabolites has been extensively studied to better understand plant interactions with and responses to the environment, insects/vertebrates, and other plants. *Pilocarpus*, a Neotropical genus in Rutaceae (Citrus Family) most notably known as the main source of pilocarpine for the treatment of glaucoma, is a chemically diverse genus rich in alkaloids, terpenoids and coumarins. An overharvesting of *Pilocarpus microphyllus* has led to near extinction of natural populations of this species; therefore, alternative sources of pilocarpine in other species, as well as an understanding of the important factors associated with increases in compound yields would greatly help reduce this destruction and maintain diversity.

The main aim of this research was to elucidate how phylogenetic relationships and adaptation to environmental variation shape phytochemical diversity in *Pilocarpus*. Phylogenetic regression models determined that phylogeny was better at predicting chemical traits when compared to models that only included certain bioclimatic factors. In addition, ecological niche models were used to determine the bioclimatic factors that contribute most to species distribution, including the identification of a large probability for niche overlap between species. Ancestral niche reconstruction identified geographically defined clades, as well as a potential tropical origin for the *Pilocarpus* clade. Next, an ecometabolomic analysis of *P. pennatifolius* revealed environmental

variables correlated with alkaloid and phenolic variation. Finally, a population study on *P. pennatifolius, P. spicatus,* and *P. riedelianus* discovered significant differences in the genetic diversity and structure of these wild populations. Studies utilizing both environmental and phylogenetic factors are essential to tease out the intricate processes in the evolution of chemical diversity in plants. These methods can benefit fields such as conservation management, ecology, and evolutionary biology.

BIOGRAPHICAL SKETCH

Daniella Meyer Allevato will receive a Ph.D. in Plant Biology from Cornell University in May 2018. Daniella grew up in a rural area of the San Fernando Valley, in the County of Los Angeles, California, where at an early age she developed a love for nature. This enthusiasm led her to pursue simultaneous bachelor degrees, Conservation and Resource Studies B.S. and Landscape Architecture B.A., at the University of California, Berkeley. She graduated in December 2012 with Distinction in the College of Natural Resources and High Honors in the College of Environmental Design. As an undergraduate she received a grant to study the effect of landscape heterogeneity for biocontrol of pests in Napa Valley vineyards in the Miguel Altieri Agroecology laboratory at UC Berkeley. Her senior thesis focused on designing rain gardens for environmental remediation and mosquito abatement. While a research assistant in the Elliot Meyerowitz Plant Development Laboratory at Caltech in Spring 2013, Daniella further enhanced her molecular biology skills.

As an undergraduate she had the opportunity to work in the Brazilian Savanna and she was amazed by the diversity of flora and fauna. After joining Kevin Nixon's laboratory at Cornell, she discovered an endangered medicinal plant *Pilocarpus* (Rutaceae), the only source for the drug pilocarpine. Pilocarpine has been used extensively in Western Medicine since 1876 to treat glaucoma as well as xerostomia in cancer patients. Her research on *Pilocarpus* has broad implications for environmental conservation and health, as plants in the genus *Pilocarpus* have been overharvested in the wild and compound yields in agricultural fields are low. She held a National Science Foundation Graduate Research Fellowship and received over 20 grants and awards for her Ph.D. research. She utilized her multilingual skills to successfully collect *Pilocarpus* samples from all over Brazil to evaluate the climatic and environmental influence on plant phytochemicals.

This dissertation is dedicated to my family for their love and support.

In loving memory of my grandfathers.

ACKNOWLEDGMENTS

I would like to thank many individuals and institutions for their assistance:

Kevin Nixon; Thank you for challenging me, and helping me become a better scientist. I am very grateful for all of the advice and assistance during my Ph.D. I also wanted to thank you and Alejandra for always being there, and making me feel like part of the family.| Eloy Rodriguez; I have learned so much and truly enjoyed all of our long meetings and discussions. Thank you so much for your guidance and mentorship throughout my dissertation. | Robert Raguso; thank you for challenging me and helping me become a better writer. | Brazilian Collaborators; Dr. Paulo Mazzafera, Dr. Milton Groppo, and Dr. Eduardo Kiyota I offer my sincerest gratitude for all the time spent reviewing and working on these projects together. | Field assistants in Brazil; Very thankful for the memories and adventures searching for *Pilocarpus,* I especially Lukas Daneu from CEPEC Herbarium in Bahia. | Brazilian Institutes and Professors; Dr. Jose Sena from Quercegen for providing samples to be analyzed for biochemical and sequencing analyses; Dr. Marcelo Caxambu and Dr. Pedro Dias for assisting with field collections. | Bill Crepet; I have really enjoyed my time at Cornell, thank you being a great mentor. | Herbaria and Gardens (NYBG, BH, USBG, MOBOT, KEW, SPF, SPFR, RBJ); Immensely grateful for the access to herbarium specimens and the opportunity to do DNA extractions and SEM with specimens. | Paul Cooper for taking amazing care of my *Pilocarpus* seedlings. | Adrienne Roeder for helping me with Flow Cytometry and being a great friend who understood the California to Ithaca transition. | Cornell Statistical Consulting Unit and Richard Furnas for assistance with statistical analyses. | Laura Joseph thank you for your friendship and many discussions. | All the Plant Biology office

administrators; Karin, Tara R., Tara N., and Polly for assisting me with important documents and fieldwork finances. | National Science Foundation Graduate Research Fellowship Program (DGE-1650441), Latin American Studies Program Fellowship, and the Section of Plant Biology for providing fellowships and teaching assistantships that supported my research. | All the societies and endowments (Harold Moore and Robert Clausen Endowment Fund, Einaudi Center, Anne S. Chatham Medicinal Botany GCA, Society of Systematic Biologists, Botanical Society of America, American Society of Plant Taxonomists, Provost Diversity Fellowship, Experiment.com, and Cornell Graduate School) that provided funding for my research and fieldwork in Brazil. | Amazing communities/organizations (OADI, GSMU, DI, BEST, CU-CIRTL) that I had the opportunity to work with on campus. | All the graduate students in Plant Biology and in the School of Plant Science at Cornell University for your friendship, support, and great discussions. Ian, thank you for being a wonderful friend and always supporting my research. I would especially like to thank Penelope, Camila, Nathan, Mischa, Dana, Coralie, Jose, Miles, and my lab members Rob and Thereis. | Finally, I would like to thank my family for their endless support and love. My parents, Denise and Eugene, for always believing in me and supporting my decisions; my brother Michael for always making me smile; my grandfather Camillo for being my biggest supporter and visiting me in Ithaca; my grandmother Ivone for always checking in with me and making me smile; and my grandmother Zelia, grandfather Ernani, and aunt Sheila for their love and support. | Love you all!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Pilocarpus, a Neotropical genus of Rutaceae rich in alkaloids, terpenoids and coumarins, is the main commercial source of the alkaloid pilocarpine for the treatment of glaucoma. Overharvesting of *Pilocarpus microphyllus* for pilocarpine production has led to near-extinction of natural populations of this species; therefore, alternative sources of pilocarpine in other species of *Pilocarpus*, as well as an understanding of genetic and environmental factors needed for increases in compound yields would greatly help reduce this destruction and maintain biodiversity.

Chapter 1 introduces a variety of theories associated with plant specialized metabolite (PSM) diversity in plants, including a discussion of various genetic and biochemical mechanisms that allow for the production of PSMs, genetic and environmental factors leading to plasticity, and a focus on alkaloid and coumarin diversity in plants. In addition, this chapter provides an in depth description of our study system, *Pilocarpus*, describing the species in the genus as well as the history of the use of pilocarpine by indigenous groups and in western medicine. Chemotaxonomy of the plant family Rutaceae is discussed to provide a background on the evolution of chemical diversity in the family, and to provide context for the unique set of compounds that have developed in *Pilocarpus.*

To better understand the evolution of chemical diversity in *Pilocarpus*, my study has utilized LCMS data in conjunction with phylogenetic methods, niche modeling methods (environmental variables), and genetic diversity (ddRADseq). An introduction to these topics is presented in Chapter 1; however, a more in depth discussion of the

methods / theories of each topic can be found in the introduction of each subsequent chapter.

As there is still much to be learned about the biosynthesis of imidazole alkaloids, a phylogenetic comparative analysis of imidazole alkaloids is presented in **Chapter 2.** Coumarins present in the genus were also assessed, as they could be an alternative form of plant defense in species of the genus that lack the imidazole alkaloids. An understanding of the evolution of imidazole alkaloids could aid in understanding the biosynthesis of these compounds as well as in genetic studies of this genus.

To better understand the genus *Pilocarpus* and to formulate a sampling scheme, I utilized ecological niche modeling in Maxent. **Chapter 3** describes the bioclimatic preferences of the species in the genus while also exploring the niche overlap between species. Niche overlap data provides a stepping-stone for studies on over-dispersion in communities. Ancestral niche reconstruction was also used to understand the evolution of vegetation types in *Pilocarpus.* This research led to the development of potential sites for field collections maximizing environmental variation.

Although genetics play an immense role in determining chemical diversity, another important factor is the effect of the environment or the adaptation to environmental pressures. As *Pilocarpus pennatifolius* has a large variety of imidazole alkaloids and also grows in a large latitudinal gradient, we focused on assessing chemical variation in *P. pennatifolius* in **Chapter 4.** This environmental metabolomics analysis identified both bioclimatic and soil factors that most significantly affected the variation of chemical profiles in six wild populations of *P. pennatifolius.*

Finally, as this genus is the only natural source of pilocarpine, which cannot be artificially synthesized, a better understanding of population structure is important to uncover genetic diversity and discover demographic patterns in wild populations. In **Chapter 5,** we assessed the genetic diversity of multiple wild populations of three species, *P. pennatifolius* (great number of alkaloids), *P. spicatus* (low number of alkaloids), and *P. riedelianus* (low number of alkaloids). Through our analyses we noted vastly different population structures among these three species, this included: very structured and geographically differentiated populations of *P. riedelianus,* high genetic similarity in two sympatric subspecies of *P. spicatus,* as well as extensive admixture and gene flow in populations of *P. pennatifolius*.

CHAPTER 1

Literature Review

INTRODUCTION

Plant Specialized Metabolite (PSM) Diversity

The evolution of biosynthetic pathways, in response to environmental factors, as well as in response to plant defenses from predation of insects/vertebrates, contributes to phytochemical diversity. Levels and types of plant compounds such as alkaloids and coumarins have been found to differ based on different environmental conditions, including changes in temperature, elevation, precipitation and soil type (Levin, 1976; Zobel and Brown, 1995; Martz *et al.*, 2009).

There are over 200,000 compounds produced by plants, and this great diversity of compounds has been the focus of many studies, examining their evolutionary significance and function. The first biochemical studies that examined the group of compounds not associated with primary metabolites (metabolites needed for basic growth and physiological processes) viewed this second set of compounds as plant secondary metabolites (metabolites not needed for basic growth, but that aid in survival). Many regarded these compounds as waste products from primary metabolism, accidents (i.e., with no obvious function) or aberrant metabolism (Haslam, 1986). This is in great contrast to today as plant secondary metabolites are now considered to have multiple functions, and are just as important as primary metabolism. This has led to the modification of the term "*plant secondary metabolites*" to "*plant specialized metabolites*" (PSMs) (Wink, 1999; Cipollini, 2000; Izhaki, 2002; Wink, 2004). The variability of PSMs can be seen at various scales in a plant. Compounds can vary within an individual due to ontogenetic changes, plant parts analyzed, and phenotypic plasticity (Rhoades and Cates, 1976; Avato *et al.*, 1987; Rhee *et al.*, 1998;

Moore *et al.*, 2014). Variation can also been analyzed in a single species both between individuals in a population and among populations, as well as between different species (Rønsted *et al.*, 2012; Moore *et al.*, 2014; Sampaio *et al.*, 2016). Variation in chemical diversity can be assessed qualitatively and quantitatively, and several studies have found that plants adapt and modify their expressed chemistry in response to both time and space, encompassing many ecological roles (Moore *et al.*, 2014).

There are many mechanisms in plants that allow for the diversity of PSMs. The simplest of which includes genetic point mutations, which can change the compound or parts of a compound produced. At a much grander scale, gene duplication can revolutionize PSMs present. This duplication can lead to neofunctionalization, where the original gene function is not needed in the new copy, which therefore can mutate under low selection pressure to produce new PSMs (Zhang, 2003). A gene duplication event can also lead to subfunctionalization, where the new gene copy can become more specialized for a specific part of a pathway or be important in certain temporal or spatial conditions (Lynch and Conery, 2000). In addition, if the duplication is due to allopolyploid it has the potential to generate new PSMs, as it accompanies the joining together of two different genomes. In this case, with multiple genomes, there is a greater diversity of genotypes, which could lead to a greater diversity of plant chemistry. There are still other mechanisms that play important roles in PSM evolution and function, such as small RNA's and epigenetics; however our understanding remains fragmented. One factor which has contributed to the diversity of large branching pathways is the functional "promiscuity" of enzymes in plants, when the same enzyme can catalyze many different reactions (Pandya *et al.*, 2014). This flexibility allows for a greater ability to produce vast numbers of compounds without the need for highly specific enzymes. In addition, some compounds can be biosynthesized through

more than one pathway, thus ensuring that the malfunction of one enzyme will not detrimentally effect the production of the compound (Firn and Jones, 2000, 2003; Pichersky and Lewinsohn, 2011).

The evolutionary conservation of the great array of naturally occurring compounds is also a topic that is highly debated, as the production of compounds in a plant does come with a cost (i.e. carbon). This cost is a tradeoff and can lead to decreases in other processes such as plant growth (Nitao and Zangerl, 1987). Therefore, it is easy to see how directional selection of adaptive evolution could lead to a reduction of inactive compounds (as they have a high cost for no reward) and an increase in production of beneficial compounds (as they improve fitness and survival)(Olson-Manning *et al.*, 2012). Though this theory makes logical sense, it appears that we know very little of the function of most PSMs. Whether each and every compound has a function is another question. The screening hypothesis on the other hand, supports the view that biological activity is a rare occurrence and the only way to increase the chance for activity is to have a vast array of compounds produced and maintained to ensure for the greatest probability of acquiring a biologically active compound (Romeo *et al.*, 1996). A somewhat similar view is the idea of silent metabolism, where every PSM is present universally in all plants, but only detected when under a particular stress (Lewinsohn and Gijzen, 2009). Of course, one of the greatest stresses comes from herbivory, and coevolution with herbivores has long been thought to lead to an arms-race of chemical weapons that constantly escalate in response to each other (Dawkins and Krebs, 1979; Vermeij, 1994).

There are many attributes that can affect the evolution of PSMs in plants. The response of each plant depends on its genetic variation, its environment, as well as the

interaction between genotype and the environment. Plasticity is the term used to describe one genotype which is able to generate multiple phenotypes, depending on the environment (Thoday, 1953). Local adaptation depends on both genetics and environmental factors to define phenotypes with the greatest fitness; however there is a gradient of importance for each quality (Spichtig and Kawecki, 2004). Many studies have also found that areas of constant warm stable temperatures and environmental factors are correlated with lower plasticity of plant secondary metabolites. There are many instances of a single plant species having different phytochemical profiles in different locations (Torras *et al.*, 2007). This could be due to different chemotypes (distinct genetically determined phytochemical profiles that are not evolutionary plastic), plasticity of a single genotype, or gradients of plasticity with genetic variance (Santesson, 1968). It can be difficult to determine the plasticity of a given trait without common garden or transplanting experiments.

A major compound class that has been studied extensively in regards to phylogenetic correlation of traits, biological activity, and plasticity is alkaloids. Alkaloids are chemical compounds that contain a nitrogen atom. These compounds are usually basic, and most commonly have a heterocyclic nitrogen (Scott, 1980). They are derived from amino acids and amino acid derivatives, often classified by their specific precursor or pathway. Over 3000 alkaloids have been described in over 4000 species of plants, and many continue to be discovered today (The Editors of Encyclopaedia Britannica 1998). Although there are some ubiquitous alkaloids across plant lineages, alkaloids are much more accurate for chemotaxonomic classifications when assessing relationships of smaller taxonomical units. Understanding the distribution of compounds and their biosynthesis allows for the prediction of compound presence in small clades (Rønsted *et al.*, 2012, Scott 1980). A recent study on Amaryllidaceae showed

that alkaloid diversity and biological activity, such as acetylcholine esterase inhibition, did correlate with phylogeny (Rønsted *et al.*, 2012). Since alkaloids often have extreme toxicity levels or pronounced biological effects, they are potential sources for biocontrol and medicine. Alkaloids commonly utilized as drugs include quinine, morphine, codeine, cocaine, taxol, atropine, physostigmine, and pilocarpine (Scott, 1980). In addition, some alkaloid compounds are present in our everyday lives: caffeine in coffee and chocolate, theobromine in chocolate, nicotine in cigarettes, and ephedrine in teas. The complexity of alkaloid biosynthetic pathways with stereospecific roles indicates high levels of specialization of alkaloids to their environment (Scott, 1980). Several studies have examined correlations between possible factors essential for alkaloid distribution. Levin (1976) demonstrated that alkaloids are "not distributed randomly in space or among taxonomic groups" through X^2 contingency tests, revealing how the percentage of alkaloids in tropical families is significantly higher than in temperate families. This latitudinal cline was postulated as an effect of coevolution caused by greater pest pressures in the tropics. Many studies have followed examining the close relationship between plant chemistry and herbivore interactions (Saunders *et al.*, 1992; Adler Lynn S. *et al.*, 2001).

Another group of compounds induced by biotic and abiotic stresses is coumarins (Hamerski and Matern, 1988; Katz *et al.*, 1998). Coumarins are phenylpropanoid compounds synthesized from the amino acid phenylalanine, which are further modified to cinnamic acid and p-coumaric acid. Coumarins and flavonoids are also part of the class of compounds known as benzopyrones, which contain a benzene attached to a pyrone (O'Kennedy and Thornes, 1997). Coumarins have been classified into the following groups, including linear and angular forms: simple, furanocoumarins, pyranocoumarins (Murray, 2002). Coumarins are mostly synthesized in leaves and can

act as defense compounds, including many that have photo-toxic qualities. This defense quality can be verified by the ways these compounds are used today, as fungicides, insecticides, and antibacterials (Feuer, 1974; Zobel and Brown, 1995; Razavi, 2011). Studies have shown that coumarin toxicity in plants increases with UV exposure, and different amounts of exposure can lead to varying amounts of specific coumarins expressed (Nitao and Zangerl, 1987; Ekiert and Gomółka, 1999). In addition, studies on *Heracleum* and *Ruta* have shown that variation in temperatures and seasonality can also affect the quantities of coumarins produced (Zobel and Brown, 1990; Zobel *et al.*, 1991). Coumarin diversity, expression, and compartmentalization appear to be dependent on both biotic and abiotic environmental pressures, as such more research is needed to study the patterns of expression (Dixon and Paiva, 1995; Zobel and Brown, 1995; Pérez-Rodríguez *et al.*, 2001).

Study System

Pilocarpus, a neotropical genus in Rutaceae in the sub-tribe Pilocarpinae, is known for its high bioactivity due to the presence of many alkaloids, terpenoids and coumarins (Santos and Moreno, 2004). *Pilocarpus* is also classified as being in the sub-family Rutoideae and the tribe Galipeae. The genus *Pilocarpus* comprises 16 species and 8 subspecies (Table 1.1, Figure 1.1; Skorupa, 1996, 1998, 1999; Kaastra, 1982). It has a widespread distribution, from southernmost Mexico and the Antilles, to northern Argentina (Skorupa, 1996). Though it has quite a large latitudinal distribution, *Pilocarpus* diversity is centered in Brazil where 14 species can be found, 11 of which are endemic. The two species found outside of Brazil are *P. demerarae,* which is endemic to Guyana, and *P. racemosus,* which can be found from Colombia all the way to Mexico and the Antilles.

Species in this genus can vary in habit, appearing as small woody shrubs to medium trees that range from three to seven meters tall (Skorupa, 1996; Pinheiro, 1997). The leaves of all but one species, *P. sulcatus,* have the distinctive pellucid dots (essential oil glands) found in many genera of Rutaceae. In general *Pilocarpus* species have simple leaves, though there are four species with compound leaves: *P. grandiflorus, P. microphyllus, P. pennatifolius, P. trachyllophus.* All of the species have inflorescences that vary in the length of the peduncle, size of flowers, color of flowers, as well as whether flowers are sessile or stalked. The fruit type found in this group is known as a capsule, and it can be made up of 1-5 dehiscent carpels.

Table 1.1 Summary of *Pilocarpus* species

Species	Locations Collected	Subspecies Collected
Pilocarpus alatus	0	
Pilocarpus carajaensis	1	
Pilocarpus demerarae	0	
Pilocarpus jaborandi	1	
Pilocarpus giganteus	1	
Pilocarpus grandiflorus	2	
Pilocarpus manuensis	0	
Pilocarpus microphyllus	2	
Pilocarpus pauciflorus	1	
subsp. *pauciflorus*		1
subsp. *clavatus*		0
Pilocarpus pennatifolius	9	
Pilocarpus peruvianus	0	
Pilocarpus riedelianus	3	
Pilocarpus spicatus	4	
subsp. *spicatus*		1
subsp. *longeracemosus*		1
subsp. *aracatensis*		1
Pilocarpus sulcatus	1	
Pilocarpus trachyllophus	1	
Pilocarpus racemosus	1	
subsp. *racemosus*		0
subsp. *goudotianus*		0
subsp. *viridulis*		0

Figure 1.1 *Pilocarpus* species *(a) P.giganteus (b) P.pennatifolius (c) P.spicatus (d) P.trachyllophus (e) P.microphyllus (f) P.sulcatus (g) P.grandiflorus* (h) P.pauciflorus (i) P.riedelianus (j) P.peruvianus*** *provided by Alex Popovkin, who aided in collection of *P.grandiflorus*.** *P.peruvianus* courtesy of William and Lynda Steere Herbarium of NYBG(http://sweetgum.nybg.org/science/vh)

Pilocarpus is most notably known for one specific imidazole alkaloid, pilocarpine, which has been used to treat angle-closure glaucoma since 1876. Pilocarpine is the most powerful sialogogue, or salivary gland inducer, so it is used to treat xerostomia in head/neck/ mouth cancer patients and patients suffering from Sjogren's Syndrome (*Therapeutic Gazette*, 1893; Hoffmann and Kremers, 1903; Marshall, 1904). It is also the most powerful diaphoretic known, inducing up to a pint of sweat an hour. Because of this it is used to treat a variety of skin issues including chronic urticaria and dryness in addition to inducing sweat to diagnose cystic fibrosis and treat Bright's disease. When applied topically or subcutaneously it has been found to increase hair growth and prevent balding. Other uses include action as a galactagogue and treatment of lead, mercury and snake venom as the poison is eliminated in the secretions (*Therapeutic Gazette*, 1893; Hoffmann and Kremers, 1903; Marshall, 1904). The prominence of pilocarpine in the health care system has landed its position on the World Health Organizations list of most essential medicines for basic health (Pinheiro, 1997, 2002; Caldeira *et al.*, 2017). Though pilocarpine is predominantly used in the medical field, it is also used in scientific studies because it can induce epilepsy in rodents in high doses since it penetrates the blood brain barrier. As is true with most drugs, high dosages can be deadly, and in high doses pilocarpine can cause convulsions in humans (Vezzani, 2009).

Though it has a long history in Western Medicine, it has an even longer history of use by Indigenous groups in South America. The common name "Jaborandi" is derived from the Amazonian Guarani and Tupi languages for "excessive slobber/slobber of the jaguar" (Holmstedt *et al.*, 1979). This extreme salivation occurs because as a parasympathomimetic drug, pilocarpine binds to the muscarinic

acetylcholine receptor M3 found throughout the body in the smooth muscles, lungs, pancreas, brain neurons, bladder, eye and endocrine/exocrine glands. M3 receptors are important for insulin homeostasis as well as they regulate insulin secretions (Bymaster *et al.*, 2003; Unno *et al.*, 2003; Kurian *et al.*, 2009). Activated M3 receptors will bind G-protein leading to the addition of GTP, release of GDP, separation of G-protein into alpha and beta subunits, and the alpha subunit activates Phospholipase C-Phosphatidyl inositol biphosphate (PIP) complex which breaks down PIP into IP3 and diacylglycerol, and IP3 interacts with ER to release Ca2+ that forms a complex with calmodulin that binds caldesmon to allow for myosin-actin interactions. This myosin-actin interaction causes the muscle contraction, which in the case of glaucoma is the contraction of the ciliary muscle that opens the trabecular meshwork in the eye and increases outflow of aqueous humor, thereby lowering the pressure causing blindness in glaucoma (Hillman, 1974). Overall pilocarpine is known to cause the contraction of smooth muscles and to increase gland secretion by inducing sweating, salivation, urination, and increased secretions in mucous membranes of the bronchis, nose, stomach, and intestine.

The first recorded use of Jaborandi was by a European explorer, Gabriel Soares de Souza, who discovered that the Guarani Indians of southern Brazil use the leaf extract to treat mouth ulcers (Caldeira *et al.*, 2017, Holmstedt *et al.*, 1979). Tea made from the leaves of *Pilocarpus* have been used for various ailments and indigenous rituals including: detoxification (induce sweating), dry mouth disorders (induce salivation), diuretic (urination), fever reducer, anticonvulsant, diabetes, asthma, pneumonia, and psoriasis (Curry and Leard, 1948; Perry *et al.*, 1963; Minette *et al.*, 1989; Pinheiro, 1997; Reuterving C.o., 2008). Dosage is extremely important as excessive amounts above 60mg can cause vomiting, nausea, convulsions, increased heart rate and bronchial

spasms (Hirsch *et al.*, 1992; Curia *et al.*, 2008; Caldeira *et al.*, 2017). Interestingly the leaves of *Pilocarpus* also contain an alkaloid jaborine that is similar to atropine, and therefore inhibits the actions of pilocarpine. Two coumarins, osthol and imperatorin, are also anticonvulsants (Venugopala *et al.*, 2013). The presence of these compounds could be one of the reasons that there are fewer instances of toxicity in tea form. In 1873 a Brazilian doctor, Symphronio Coutinho, brought leaves of *Pilocarpus* to Paris, and their extract was readily employed across Europe to treat bronchitis, fever, stomatitis, psoriasis, and glaucoma.

Pilocarpine's growing use in Western Medicine led to the over-exploitation of many native populations of *Pilocarpus* in Brazil. Unfortunately this caused some of the species to be added to the list of Brazilian Flora threatened with extinction: Endangered (*P. microphyllus* and *P. alatus*) and Vulnerable (*P. jaborandi*, and *P. trachyllophus*) (IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis, 1992). A response to this destruction was the creation of large monoculture farms of *Pilocarpus microphyllus.* Although these farms have allowed for a reduction of collection in the wild, there has also been a loss of natural diversity in these monoculture farms. One study found that from 1975 to 1992 the leaf yield (MT/hectares) from cultivated fields has decreased by 50%, possibly due to unsustainable harvesting (Pinheiro, 2002). This decrease in leaf production has caused a 36% increase in the price of pilocarpine in the last 20 years since the demand remains the same, and exports of pilocarpine contribute $US 6.8billion a year to Brazil's economy (Caldeira *et al.*, 2017; IGBE, 2014). This issue is extremely important since pilocarpine is only extracted from the leaf tissue of *Pilocarpus* and is not synthesized in the laboratory for commercialization. In fact, studies have found that leaf tissue collected in the wild has much larger concentrations of pilocarpine when compared with leaf tissue collected in the greenhouse and farm. This

could be due to greater genetic diversity in wild populations or environmental factors that contribute to increasing compound yields. The question that arises is whether there is a way to study native wild populations to detect factors that increase compound yields, without destroying the native populations.

Chemotaxonomy of Rutaceae

Chemotaxonomy of Rutaceae was extensively studied in the 20th Century (Waterman, 1975; Gray and Waterman, 1978). Both Price (1963) and Hegnauer (1963) helped to classify the alkaloids present in various species of Rutaceae based on the biogenesis of the compounds (Swain, 1963). The precursors of the 10 main alkaloid groups found in Rutaceae include tyrosine, anthranilic acid, tryptophan, and histidine (Figure 1.2) (Waterman, 1975). *Pilocarpus* does not contain any of the typical alkaloids present in other genera of Rutaceae, but rather has a great diversity of imidazole alkaloids present. In Rutaceae, imidazole alkaloids are only found in two genera: *Pilocarpus* and *Casimoroa* (subfamily Toddaliinae from Mexico). The alkaloids in *Casimoroa* however are not similar to the ones in *Pilocarpus*, and *Pilocarpus* also has a greater diversity of imidazole alkaloids (Waterman, 1975). Of the members in the sub-tribe Pilocarpinae, four of the six genera have the typical furoquinolines found throughout Rutaceae, as well as a couple genera with acridones and indolequinazolines. Many studies of plant secondary metabolites in Rutaceae have found that there is a progression from a prominence of alkaloids (nitrogen-containing compounds) at the base of Rutaceae, leading to the prominence of coumarins (phenolics) and finally the most derived species contain a multitude of flavonoids, lignans, limonoids, quassinoids and triterpenes (Silva *et al.*, 1988). This could possibly be due to nitrogen economy or the evolution of complex

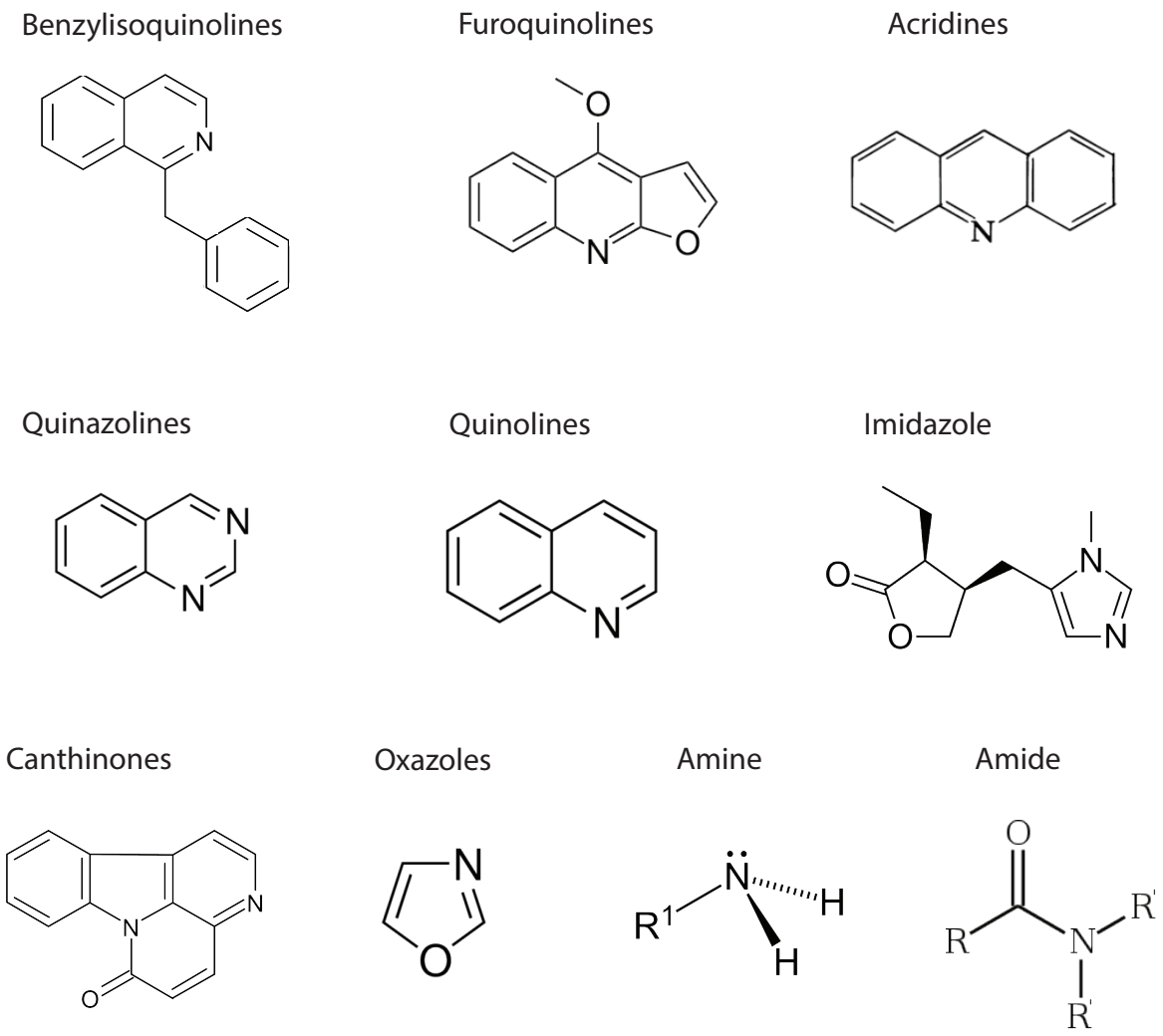


Figure 1.2 Alkaloid classes in Rutaceae

pathways involving many precursors. *Pilocarpus* has many coumarins and terpenes, none of the ubiquitous Rutaceae alkaloids, but instead has developed a diverse number of imidazole alkaloids derived from histidine.

Dissertation Objectives

The overall goal of this dissertation was to understand the evolution of chemical diversity in the Neotropical and economically important genus *Pilocarpus* (Rutaceae). This was done at various levels of hierarchy, beginning with an assessment of distributions of all the species in *Pilocarpus:*

Characterizing interspecific diversity for compound evolution

Intrageneric Phylogenetic Correlation	• Phylogeny estimation using ddRADseq • Phylogenetic signal of chemical traits • Continuous and discrete trait reconstruction • Phylogenetic regression of chemical trait-climate models
Intrageneric Environmental Variation	• Ancestral niche reconstruction • Ecological niche model for each species • Niche overlap of *Pilocarpus* species
Environmental Gradients	• Ecometabolomic study of wild populations • Alkaloid and Phenolic extraction • CCA to identify correlation of chemical variation with environment in *P. pennatifolius*
Within Species Genetic Variation	• Population genomics (ddRADseq) • 3 species: *P. pennatifolius, P. spicatus, P. riedelianus* • Genetic diversity and structure analyses

- **(AIM 1)** ***Intrageneric phylogenetic correlation*****:** I assessed the distribution of alkaloids and coumarins using phylogenetic comparative methods, to determine if variation in chemical profiles is correlated with phylogeny. Phylogeny was estimated in raxML, TNT, and BEAST using ddRADseq data (Pagel, 1992; Blomberg *et al.*, 2003; Wink, 2003; Drummond and Rambaut, 2007; Stamatakis *et al.*, 2008; Goloboff *et al.*, 2008; Rønsted *et al.*, 2012; Eaton, 2014).
- **(AIM 2)** ***Intrageneric Environmental Variation:*** Niche modeling and niche overlap analyses helped to understand the bioclimatic factors responsible for the distribution and demographic history of *Pilocarpus* species. These analyses also aided in determining sites for field collections maximizing environmental variation (Warren *et al.*, 2008; Elith *et al.*, 2011; Peterson *et al.*, 2011).
- **(AIM 3)** ***Environmental gradients*****:** This was followed by an in-depth environmental metabolomics study on one species, *P. pennatifolius,* to verify if chemoecotypes are present and determine which environmental factors (bioclimatic and soil variables) are most significant in varying alkaloids and coumarins in wild populations. Statistical unimodal ordination methods were used to analyze the LCMS data (Braak, 1986; Smith *et al.*, 2006; Pluskal *et al.*, 2010; Arbona *et al.*, 2013; Badri *et al.*, 2013; Jamil *et al.*, 2014).
- **(AIM 4)** ***Within species genetic variation*****:** Finally, as genetic variation of populations can also affect plant chemistry, we analyzed the population structure of three species: *P. pennatifolius* (largest # alkaloids), *P. riedelianus* (low # alkaloids), *P. spicatus* (low # alkaloids). Next-generation sequencing was utilized to perform population genomic analyses (Jombart, 2008; Baird *et al.*, 2008; Davey and Blaxter, 2010; Kamvar *et al.*, 2014).

Field Collections Summary

Based on the center of diversity of *Pilocarpus* species in Brazil, two field trips were taken to Brazil. My first trip was January-April 2016 and my second field trip was August 2017. In Brazil I collected *Pilocarpus* in wild populations including, 12 species and 4 subspecies, as well as 6 outgroups for phylogenetic analyses (Nixon & Carpenter 1995) (Table 1.1, Table 1.2, Appendix 1). Collections were made under my SISBIO permit (#52758-1) as well as under the collection permits of my collaborators: Dr. Milton Groppo (USP-Ribeirão Preto), Dr. Pedro Dias (USP-SP), Dr. Marcelo Caxambu (UTFPR) and Dr. Jose Sena (Quercegen). Herbarium specimens were also used for whole genome DNA sequencing to estimate phylogenetic relationships; *P. peruvianus* was added to the phylogeny and other specimens were used to confirm that each species was monophyletic. Some of this work was performed under USDA-APHIS permit number P37-15-00917. I completed the biochemical analysis (UPLC-ESI-MS) of *Pilocarpus* leaf tissue in the laboratory of Dr. Paulo Mazzafera, Professor of Plant Biology and Plant Physiology at UNICAMP (Universidade Estadual de Campinas), Brazil. Dr. Mazzafera is leading research efforts in elucidating the biosynthetic pathway of pilocarpine.

Table 1.2. *Pilocarpus* outgroups used for analyses

Subfamily	Tribe	Subtribe	Genus	Locations Collected
Rutoideae	Galipeae	Galipeinae	*Galipea*	1
			Neoraputia	
			Rauia	
		Pilocarpinae	*Esenbeckia*	1
			Metrodorea	1
			Pilocarpus	
			Raulinoa	
	Zanthoxyleae	Evoidiinae	*Zanthoxylum*	
Toddalioideae	Toddalieae	Pteleinae	*Balfourodendron*	1
			Helietta	1
			Ptelea	1

Conclusions

Plant specialized metabolites in plants are not strictly produced for growth and reproduction, but rather play a part in interactions with the environment, colonization of ecological niches, and protection against abiotic stresses. Analyzing chemical diversity using a combination of phylogenetic, niche-modeling and population genomic methods is an innovative way to investigate the evolution of chemical diversity in *Pilocarpus*. These methods are important and can be utilized to benefit a variety of fields including medicine, industry, and agriculture.

LITERATURE CITED

Adler Lynn S., Karban Richard, Strauss Sharon Y. 2001. Direct and indirect effects of alkaloids on plant fitness via herbivory and pollination. Ecology **82**, 2032–2044.

Arbona V, Manzi M, Ollas C, Gómez-Cadenas A. 2013. Metabolomics as a Tool to Investigate Abiotic Stress Tolerance in Plants. International Journal of Molecular Sciences **14**, 4885–4911.

Avato P, Bianchi G, Salamini F. 1987. Ontogenetic Variations in the Chemical Composition of Maize Surface Lipids. The Metabolism, Structure, and Function of Plant Lipids. Springer, Boston, MA, 549–551.

Badri DV, Zolla G, Bakker MG, Manter DK, Vivanco JM. 2013. Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. The New Phytologist **198**, 264–273.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers (JC Fay, Ed.). PLoS ONE **3**, e3376.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution; International Journal of Organic Evolution **57**, 717–745.

Braak CJFT. 1986. Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology **67**, 1167–1179.

Bymaster FP, Carter PA, Yamada M, Gomeza J, Wess J, Hamilton SE, Nathanson NM, McKinzie DL, Felder CC. 2003. Role of specific muscarinic receptor subtypes in cholinergic parasympathomimetic responses, in vivo phosphoinositide hydrolysis, and pilocarpine-induced seizure activity. The European Journal of Neuroscience **17**, 1403–1410.

Caldeira CF, Giannini TC, Ramos SJ, *et al.* 2017. Sustainability of Jaborandi in the eastern Brazilian Amazon. Perspectives in Ecology and Conservation **15**, 161–171.

Cipollini ML. 2000. Secondary metabolites of vertebrate-dispersed fruits: evidence for adaptive functions. Revista chilena de historia natural **73**, 421–440.

Curia G, Longo D, Biagini G, Jones RSG, Avoli M. 2008. The pilocarpine model of temporal lobe epilepsy. Journal of Neuroscience Methods **172**, 143–157.

Davey JW, Blaxter ML. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics **9**, 416–423.

Dixon R, Paiva N. 1995. Stress-Induced Phenylpropanoid Metabolism. The Plant Cell **7**, 1085–1097

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**, 214.

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

Ekiert H, Gomółka E. 1999. Effect of light on contents of coumarin compounds in shoots of Ruta graveolens L. cultivated in vitro. Acta Societatis Botanicorum Poloniae **68**, 197–200.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**, 43–57.

Firn RD, Jones CG. 2000. The evolution of secondary metabolism – a unifying model. Molecular Microbiology **37**, 989–994.

Firn RD, Jones CG. 2003. Natural products ? a simple model to explain chemical diversity. Natural Product Reports **20**, 382.

Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. Cladistics **24**, 774–786.

Gray AI, Waterman PG. 1978. Coumarins in the Rutaceae. Phytochemistry **17**, 845–864.

Hamerski D, Matern U. 1988. Biosynthesis of psoralens. Psoralen 5-monooxygenase activity from elicitor-treated Ammi majus cells. FEBS letters **239**, 263–265.

Haslam E. 1986. Secondary Metabolism- Fact or Fiction. Natural Product Reports.

Hegnauer, R. 1963. The Taxonomic significance of alkaloids, in: *Chemical plant taxonomy* (T. Swain, ed.). London, New York: Academic Press.

Hillman JS. 1974. Management of acute glaucoma with pilocarpine-soaked hydrophilic lens. The British Journal of Ophthalmology **58**, 674–679.

Hirsch E, Baram TZ, Snead OC. 1992. Ontogenic study of lithium-pilocarpine-induced status epilepticus in rats. Brain Research **583**, 120–126.

Hoffmann F, Kremers E. 1903. *Pharmaceutical Review*. Pharmaceutical Review Publishing Company.

IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis. 1992. *Portaria No. 06N*. Brasilia.

Izhaki I. 2002. Emodin – a secondary metabolite with multiple ecological functions in higher plants. New Phytologist **155**, 205–217.

Jamil T, Kruk C, Braak CJF ter. 2014. A Unimodal Species Response Model Relating Traits to Environment with Application to Phytoplankton Communities. PLOS ONE **9**, e97583.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**, 1403–1405.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**, e281.

Katz VA, Thulke OU, Conrath U. 1998. A benzothiadiazole primes parsley cells for augmented elicitation of defense responses. Plant Physiology **117**, 1333–1339.

Kurian N, Hall CJ, Wilkinson GF, Sullivan M, Tobin AB, Willars GB. 2009. Full and Partial Agonists of Muscarinic M3 Receptors Reveal Single and Oscillatory Ca2+ Responses by 2-Adrenoceptors. The Journal of Pharmacology and Experimental Therapeutics **330**, 502–512.

Levin DA. 1976. The Chemical Defenses of Plants to Pathogens and Herbivores. Annual Review of Ecology and Systematics **7**, 121–159.

Lewinsohn E, Gijzen M. 2009. Phytochemical diversity: The sounds of silent metabolism. Plant Science **176**, 161–169.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science (New York, N.Y.) **290**, 1151–1155.

Marshall CR. 1904. On the physiological action of the alkaloids of jaborandi leaves. The Journal of Physiology **31**, 120–156.

Martz F, Peltola R, Fontanay S, Duval RE, Julkunen-Tiitto R, Stark S. 2009. Effect of Latitude and Altitude on the Terpenoid and Soluble Phenolic Composition of Juniper (Juniperus communis) Needles and Evaluation of Their Antibacterial Activity in the Boreal Zone. Journal of Agricultural and Food Chemistry **57**, 9575–9584.

Moore BD, Andrew RL, Külheim C, Foley WJ. 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytologist **201**, 733–750.

Murray RDH. 2002. The Naturally Occurring Coumarins. Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products. Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products. Springer, Vienna, 1–619.

Nitao JK, Zangerl AR. 1987. Floral Development and Chemical Defense Allocation in Wild Parsnip (Pastinaca Sativa). Ecology **68**, 521–529.

O'Kennedy R, Thornes RD. 1997. *Coumarins: biology, applications, and mode of action*. Chichester; New York: Wiley.

Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. Nature Reviews Genetics **13**, 867–877.

Pagel MD. 1992. A method for the analysis of comparative data. Journal of Theoretical Biology **156**, 431–442.

Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. 2014. Enzyme Promiscuity: Engine of Evolutionary Innovation. Journal of Biological Chemistry **289**, 30229–30236.

Pérez-Rodríguez E, Aguilera J, Gómez I, Figueroa F. 2001. Excretion of coumarins by the Mediterranean green alga Dasycladus vermicularis in response to environmental stress. Marine Biology **139**, 633–639.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. *Ecological Niches and Geographic Distributions*. Princeton, N.J: Princeton University Press.

Pichersky E, Lewinsohn E. 2011. Convergent Evolution in Plant Specialized Metabolism. Annual Review of Plant Biology **62**, 549–566.

Pinheiro CU. 1997. Jaborandi (Pilocarpus sp., rutaceae): A wild species. Economic Botany **51**, 49–58.

Pinheiro CU. 2002. Extrativismo, cultivo e privatização do jaborandi (Pilocarpus microphyllus Stapf ex Holm.; Rutaceae) no Maranhão, Brasil. Acta Botanica Brasilica **16**, 141–150.

Pluskal T, Castillo S, Villar-Briones A, Orešič M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics **11**, 395.

Price J.R. 1963. The distribution of alkaloids in Rutaceae, in: *Chemical plant taxonomy* (T. Swain, ed.). London, New York: Academic Press.

Rhee Y, Hlousek-Radojcic A, Ponsamuel J, Liu D, Post-Beittenmiller D. 1998. Epicuticular Wax Accumulation and Fatty Acid Elongation Activities Are Induced during Leaf Development of Leeks. Plant Physiology **116**, 901–911.

Rhoades DF, Cates RG. 1976. Toward a General Theory of Plant Antiherbivore Chemistry. Recent Advances in Phytochemistry. Biochemical Interaction Between Plants and Insects. Springer, Boston, MA, 168–213.

Romeo JT, Saunders JA, Barbosa P (Eds.). 1996. *Phytochemical Diversity and Redundancy in Ecological Interactions*. Boston, MA: Springer US.

Rønsted N, Symonds MRE, Birkholm T, *et al.* 2012. Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of amaryllidaceae. BMC Evolutionary Biology **12**, 182.

Sampaio BL, Edrada-Ebel R, Costa FBD. 2016. Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants. Scientific Reports **6**, 29265.

Santesson R. 1968. *Svensk Naturvetenskap*. Stockholm, Sweden: Allf.

Santos AP, Moreno PRH. 2004. Pilocarpus spp.: a survey of its chemical constituents and biological activities. Revista Brasileira de Ciências Farmacêuticas **40**, 116–137.

Saunders JA, O'neill NR, Romeo JT. 1992. Alkaloid Chemistry and Feeding Specificity of Insect Herbivores. Alkaloids: Chemical and Biological Perspectives. Alkaloids: Chemical and Biological Perspectives. Springer, New York, NY, 151–196.

Scott TA. 1980. Alkaloid biology and metabolism in plants by George R Waller and Edmund K Nowacki. pp 294. Plenum Press: New York and London. 1978. Biochemical Education **8**, 30–30.

Silva MF das GF da, Gottlieb OR, Ehrendorfer F. 1988. Chemosystematics of theRutaceae: Suggestions for a more natural taxonomy and evolutionary interpretation of the family. Plant Systematics and Evolution **161**, 97–134.

Skorupa LA. 1996. Revisao Taxonomica de Pilocarpus Vahl (Rutaceae). Ph.D. Thesis.

Skorupa LA. 1998. Three New Species of Pilocarpus Vahl (Rutaceae) from Brazil. Novon **8**, 447–454.

Skorupa LA. 1999. A New Species of Pilocarpus Vahl (Rutaceae) from Peru. Novon **9**, 257–259.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical Chemistry **78**, 779–787.

Spichtig M, Kawecki TJ. 2004. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. The American Naturalist **164**, 70–84.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Systematic Biology **57**, 758–771.

Swain T. 1963. *Chemical plant taxonomy*. London, New York: Academic Press.

The Editors of Encyclopaedia Britannica. Alkaloid | Definition, Structure, & Classification. Encyclopedia Britannica.

Therapeutic Gazette. 1893. G. S. Davis.

Thoday JM. 1953. Components of fitness. Symposium of the Society for Experimental Biology.

Torras J, Grau MD, López JF, Heras FXC de las. 2007. Analysis of essential oils from chemotypes of Thymus vulgaris in Catalonia. Journal of the Science of Food and Agriculture **87**, 2327–2333.

Unno T, Kwon S-C, Okamoto H, Irie Y, Kato Y, Matsuyama H, Komori S. 2003. Receptor signaling mechanisms underlying muscarinic agonist-evoked contraction in guinea-pig ileal longitudinal smooth muscle. British Journal of Pharmacology **139**, 337–350.

Venugopala KN, Rashmi V, Odhav B. 2013. Review on Natural Coumarin Lead Compounds for Their Pharmacological Activity. BioMed Research International.

Warren DL, Glor RE, Turelli M. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution **62**, 2868–2883.

Waterman PG. 1975. Alkaloids of the rutaceae: their distribution and systematic significance. Biochemical Systematics and Ecology **3**, 149–180.

Wink M (Ed.). 1999. *Biochemistry of Plant Secondary Metabolism*. Sheffield: WileyBlackwell.

Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64**, 3–19.

Wink M. 2004. Phytochemical diversity of secondary metabolites. Encycl Plant Crop Sci, 915–9.

Zhang J. 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution **18**, 292–298.

Zobel AM, Brown SA. 1990. Seasonal changes of furanocoumarin concentrations in leaves of *Heracleum lanatum*. Journal of Chemical Ecology **16**, 1623–1634.

Zobel AM, Brown SA. 1995. Coumarins in the interactions between the plant and its environment. Allelopathy Journal.

Zobel AM, Wang J, March RE, Brown SA. 1991. Identification of eight coumarins occurring with psoralen, xanthotoxin, and bergapten on leaf surfaces. Journal of Chemical Ecology **17**, 1859–1870.

CHAPTER 2

Phylogenetic Regression and Character Reconstruction of Phytochemical Diversity in *Pilocarpus*

Daniella M. Allevato[1,*], Milton Groppo[2], Eduardo Kiyota[3],

Paulo Mazzafera[3], Kevin C. Nixon[1]

ABSTRACT

The evolution of phytochemical diversity and biosynthetic pathways in plants can be evaluated from a phylogenetic perspective relative to environmental factors and/or in the context of chemical defenses in response to predation by insects and vertebrates. *Pilocarpus,* an economically important medicinal plant in the family Rutaceae, has a great diversity of imidazole alkaloids and coumarins. In this study, we utilize phylogenetic comparative methods to determine whether there is a phylogenetic signal for chemical traits across the genus *Pilocarpus,* including ancestral reconstructions of continuous and discrete chemical traits. Bioclimatic variables found to be associated with the distribution of this genus were used to perform *GLS* regressions between chemical traits and bioclimatic variables. Next, these regression models were compared with *pGLS* regressions, models including phylogenetic relationships, to test whether the inclusion of phylogenetic relationships led to model improvement. We found that phylogeny was a better overall predictor of chemical traits when compared to bioclimatic factors such as Annual Mean Temperature, Precipitation in Driest Quarter

---

[1]L.H. Bailey Hortorium, Section of Plant Biology, School of Plant Sciences, Cornell University, Ithaca, NY, USA
[2]Departamento de Biologia, Universidade de Sao Paulo, Ribeirão Preto, Brazil
[3]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Brazil

and Precipitation in Wettest Quarter; however, there were two bioclimatic factors, Annual Precipitation and Precipitation in Coldest Quarter, which were better predictors of chemical traits in comparison to phylogenetic relationships. Studies utilizing both environmental and phylogenetic factors are essential to tease out the intricate processes in the evolution of chemical diversity in plants. These methods can benefit fields such as conservation management, ecology, and evolutionary biology.

KEYWORDS: Chemosystematics/chemotaxonomy, phylogenetic signal, evolution, chemical pathways, coumarin, imidazole alkaloid, *Pilocarpus*

INTRODUCTION

Plants produce a vast assortment of chemical compounds ranging from primary metabolites, necessary for plant growth and survival, to plant specialized metabolites (PSM) important in growth and plant defense. Many studies have found that closely related species have highly conserved and similar metabolites (Bisby *et al.*, 1980; Harborne, 1984; Hegnauer, 1994; Thaler *et al.*, 2012; Rønsted *et al.*, 2012). The similarity of PSM chemical structures in related species is what led to the development of a classification system known as chemotaxonomy or chemosystematics (Helen C. De S. Abbott, 1886; Penfold, 1927; McNair, 1935; Hegnauer, 1994). Through time, various studies have shown that closely related species can also have divergent traits (Figure 2.1). This could be due to disruptive selection in new environments (Latta and Linhart, 1997; Geber and Griffen, 2003; Rueffler *et al.*, 2006), competition in the same environment (Mraja 2011), overdispersion of species in the same community (Becerra 2007) or convergent evolution (Fraenkel, 1959; Pichersky and Lewinsohn, 2011).

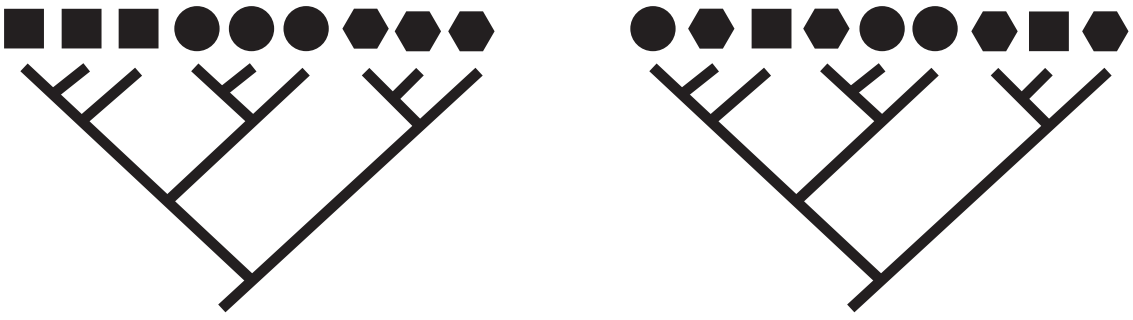


Figure 2.1. Closely related species can have conserved or divergent chemical traits. Conserved traits in (a) shows species in the same clade with the same chemical trait. In (b) we see that species in the same clade can have different chemical traits.

There are many genetic mechanisms that might increase the diversity of plant specialized metabolites. The simplest mechanism includes point mutations that can

alter enzyme activities or substrate specificity, providing for greater diversity of metabolites (Dixon, 2001). At a genomic scale, gene duplication can revolutionize PSMs present through neofunctionalization producing new PSMs (Zhang, 2003), subfunctionalization providing for specialization of a pathway (Lynch and Conery, 2000), and in the case of allopolyploidy the combination of two different genomes providing a diverse array of PSMs. Recent studies depict the role of small RNA's and epigenetics in chemical diversity; however, our understanding remains fragmented (Boyko and Kovalchuk, 2008, 2011; Urano *et al.*, 2010; Parent *et al.*, 2012). One factor, which has helped the diversity of large branching pathways, is the promiscuity of enzymes in plants (Pandya *et al.*, 2014). This flexibility facilitates the production of vast amounts of compounds without the need for highly specific enzymes (Firn and Jones, 2000, 2003; Pichersky and Lewinsohn, 2011).

Chemotaxonomists have extensively studied patterns of plant specialized metabolites to infer relationships among taxa in Rutaceae, a family with 1000's of terpenes, coumarins, and alkaloids (Waterman, 1975, 1990; Gray and Waterman, 1978; Mabry and Ulubelen, 1980; Waterman and Hussain, 1983; Silva *et al.*, 1988). In fact, Rutaceae is the plant family with the richest diversity of alkaloid classes including: acridines, furoquinolines, quinolines, quinazolines, indoloquinazolines, canthinones, benzylisoquinolines, and aromatic amides and amines (Waterman, 1975). The precursors for these alkaloid classes include histidine, anthranilic acid, tryptophan and tyrosine (McNair, 1935; Waterman, 1975; Price *et al.*, 1989). Of these classes, four are essentially limited to Rutaceae: imidazoles, furoquinolines, acridines, and indoloquinazolines. Rutaceae is also rich in coumarins, phenylpropanoids synthesized from the amino acid phenylalanine that are further modified to cinnamic acid and p-coumaric acid. Though these compounds are more widely distributed among

Angiosperm lineages, Rutaceae nevertheless has a great diversity of which major chemical markers have been studied due to the proliferation of furanocoumarins and pyranocoumarins (Gray and Waterman, 1978; Harborne, 1984, 1993; Murray, 2002).

Pilocarpus, a genus of Rutaceae in the sub-tribe Pilocarpinae, is known for its high bioactivity due to the presence of many alkaloids, terpenoids and coumarins (Santos and Moreno, 2004). *Pilocarpus* comprises 16 species and has a widespread distribution from southern Mexico and the Antilles, to northern Argentina (Skorupa, 1996). *Pilocarpus* does not contain any of the typical alkaloids present in Rutaceae, but rather has a unique set of diverse imidazole alkaloids derived from histidine. Two different imidazole alkaloids have been found in *Casimoroa,* another genus in Rutaceae (subfamily Toddaliinae), distributed in Mexico (Waterman, 1975). Of the members in the sub-tribe Pilocarpinae, four of the six genera have the typical furoquinolines found throughout Rutaceae, as well as a few with acridines and indolequinazolines. Many studies that have looked at the metabolites in the Rutaceae plant family, have found a progression (from basal species to derived species) of the prominence of certain compounds. An original prominence of alkaloids and later decline of alkaloids was correlated with a rise in coumarins (oxygen derivatives) and subsequently a rise in flavonoids, lignans, limonoids, quassinoids and triterpenes (Silva *et al.*, 1988). This could possibly be due to nitrogen economy or the evolution of complex pathways involving many precursors. *Pilocarpus,* as mentioned previously, has many coumarins and terpenes, none of the ubiquitous alkaloids found in the other Rutaceae, but instead has developed a diverse number of imidazole alkaloids from histidine. It is possible that abiotic/biotic pressures led to the development of this unique set of compounds, necessary after a loss of the traditional Rutaceae alkaloids in the genus. Though there

have been many studies on these imidazole alkaloids, their biosynthesis is not fully understood.

A phylogenetic perspective was used to gain a better understanding of the evolution of imidazole alkaloids and coumarins in *Pilocarpus*. As species are non-independent, due to phylogenetic relationships, it is important to account for the dependence of chemical traits on the relationship between species when using comparative methods. There are various statistical approaches to evaluating correlated traits in phylogenetics. Our analysis has included the quantification of phylogenetic signal using three metrics: Pagel's λ to determine the diversity in the genus, the D-metric to determine the signal of individual compounds, and Blomberg's K statistic to assess quantitative variation in compounds (Pagel, 1992; Blomberg *et al.*, 2003). Next, ancestral state reconstruction was performed to better characterize evolutionary patterns. Finally, to assess the relationship of metabolites in the genus to extrinsic bioclimatic factors, phylogenetic and non-phylogenetic models were compared.

MATERIALS AND METHODS

Field collections of plant tissue

Field collected plants that were sampled include a total of 137 individuals. Each individual plant underwent two types of extractions: coumarin and alkaloid. For each extraction there were three different leaf replicates, giving a total of 411 coumarin extraction and 411 alkaloid extractions. *Pilocarpus* species were collected from various sites, obtaining as many individuals as possible from each site (Figure 2.2, APPENDIX 1).

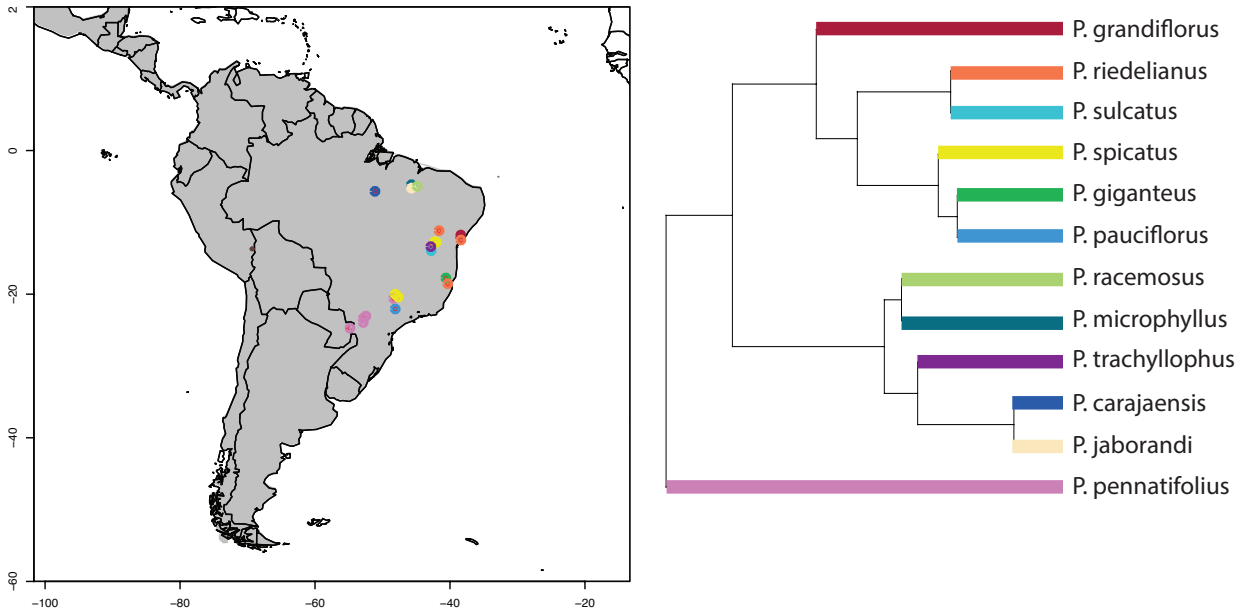


Figure 2.2. *Pilocarpus* species collection localities. Dots on the map represent collection sites. The colors of the dots correspond to the species on the phylogeny to the right, with some species collected at multiple localities. *This phylogeny was estimated using Bayesian methods described in this paper.

Biochemical extraction methods

To validate the biochemical extractions from field collected leaf tissue, a degradation study on two live specimens of *Pilocarpus pennatifolius* from the New York Botanic Garden was performed to examine degradation of alkaloids and coumarins in the leaves. Silica-dried leaf tissue on ice showed the least degradation and was not significantly more degraded relative to immediately frozen tissue. This method of preservation of leaf tissue for biochemical extractions has been used in many studies that require plant collections in remote areas, as the removal of water reduces enzymatic activity in the leaf tissue and the cold temperature helps to prevent enzymatic reactions due to handling (Kim and Verpoorte, 2010; Fine *et al.*, 2013).

Extraction protocol for alkaloids and coumarins

Biochemical extractions to elucidate alkaloids and coumarins present in *Pilocarpus* leaf tissues were performed on silica-dried leaf tissue, at UNICAMP, Campinas, Brazil. Alkaloid extraction protocols are described in Allevato Chapter 4, and these modifications on the method developed by Avancini *et al.*, (2003) were made by Dr. Sawaya (Dr. Mazzafera Lab). The coumarin extraction protocol followed the methods developed in Vialart *et al.*, (2012) and Durand-Hulak *et al.*, (2015).

UPLC-ESI-MS methods

These alkaloid and coumarin extracts were run using UPLC-ESI-MS.

Alkaloid

Five microliters of each sample were injected into an Acquity UPLC coupled with a TQD triple-quadrupole mass spectrometer (Micromass-Waters, Manchester, England). Mass spectrometer conditions were: capillary 3.0 kV, cone 30 V, extractor 1V, ion source temperature 150° C, desolvation temperature 300° C and column temperature 30° C, in electrospray ionization in positive mode. The chromatographic column used was a Polaris 3 C18-A 100 x 2.0mm column (Varian) and the elution was carried out using an ammonium acetate buffer, 10mmol / L, pH 3.0 (solvent A) and acetonitrile (solvent B) in a gradient ranging from 5 to 25% of solvent B in 8 min. The retention time and m/z were used to identify alkaloid compounds in the samples.

Coumarin

Sample extracts were diluted (1:4) in 80% ethanol (ethanol and water). Four microliters of each sample were injected into an Acquity UPLC coupled with a TQD triple-quadrupole mass spectrometer (Micromass-Waters, Manchester, England). Mass spectrometer conditions were: capillary 3.0 kV, cone 35 V, extractor 1V, ion source temperature 120° C, desolvation temperature 350° C and column temperature 30° C, in

electrospray ionization in positive mode. The chromatographic column used was a Waters Acquity C18-BEH (2.1 × 100 mm,1.7 μm) column and the elution was carried out using Milli-Q water with 0.1% formic acid (solvent A) and acetonitrile (solvent B) in a gradient ranging from 10 to 100% of solvent B in 8 min. The retention time and m/z were used to identify alkaloid compounds in the samples using a SIM (single ion monitoring) mode.

UPLC-MS data processing, compound identification, and quantification

A subset of 22 alkaloids was identified and quantified based on a standard curve calibration with a pilocarpine standard using a method developed in Sawaya *et al.*, (2008), for all 375 samples. For the coumarin extracts (375 total samples), a subset of 8 coumarins was quantified using chemical standards to compare retention times and m/z fragments. Quantification of coumarins was through the normalization of peak areas of the UPLC chromatograms (Masson *et al.*, 2016)

UPLC-MS data scoring

Chemical data on alkaloids and coumarins were scored for presence/absence, as well as concentration (Figure 2.3). One alkaloid, pilocarpine, was quantified with a reference standard and eight coumarins were quantified from reference standards. Each species had multiple individual replicates, as well as three biological replicates (APPENDIX 2). For the 30 compounds we evaluated in this study, there were only four instances where there was one individual with a compound absent (*P. pennatifolius:* coumarin and citropten; *P. riedelianus: 13-nor-7(11)dehydropilocarpine; P. spicatus:* citropten). For the analyses we considered a compound to be present for the species when the majority had the compound present. The greatest variation detected between sites and individuals was the concentration of the nine quantified compounds.

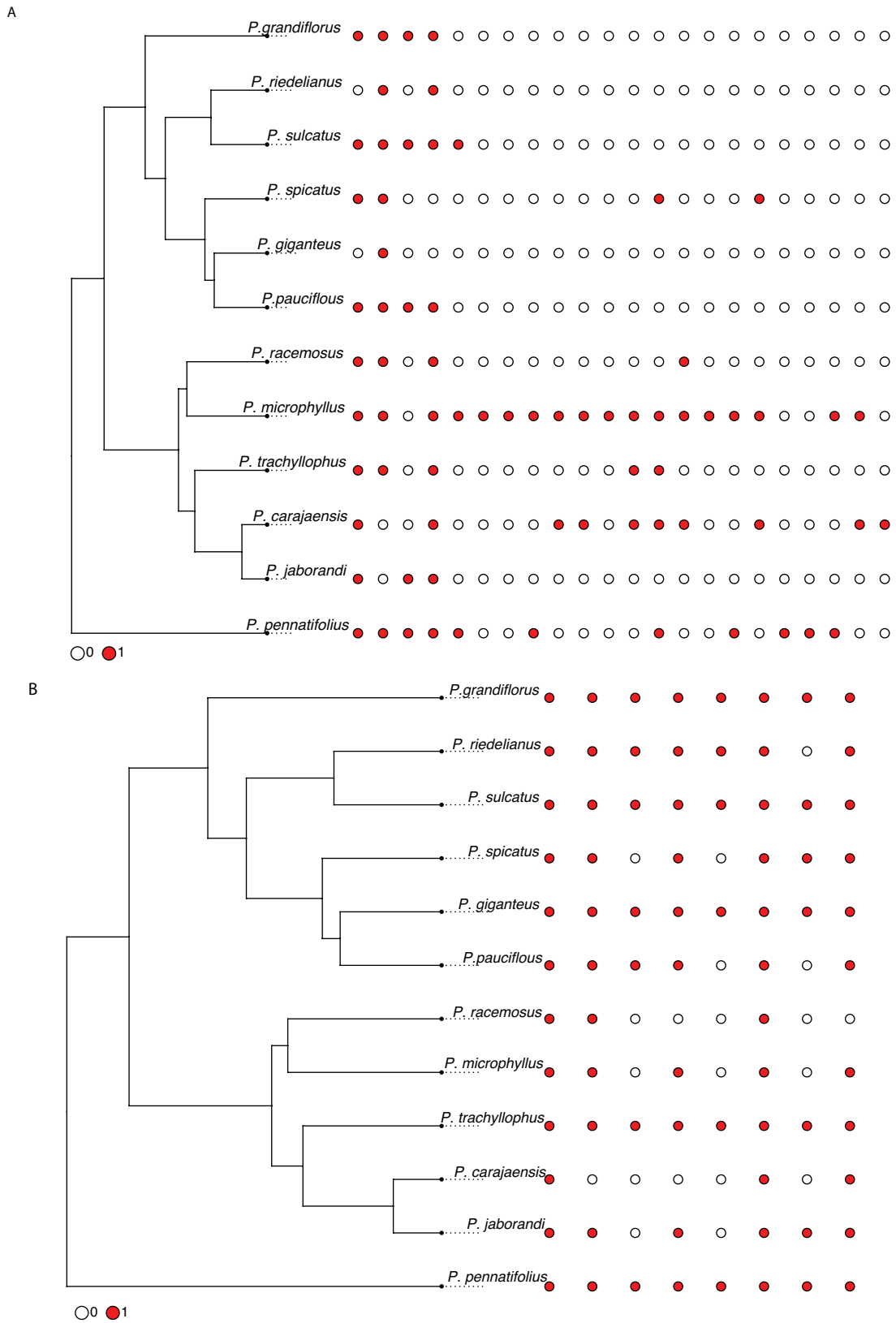


Figure 2.3. Presence/absence of compounds in *Pilocarpus*. Each column represents a compound, red represents compound presence and white represents compound absence. (a) alkaloids (b) coumarins *This phylogeny was estimated using Bayesian methods

DNA extraction and sequencing

Whole genome DNA extractions were done with a few modifications of Barry *et al.*, (2005). Modifications include the addition of RNase A before incubation in the water bath, the use of ice-cold ethanol and ice-cold isopropanol. Genome size (2C=5.05pg) for optimization of sequencing was determined using Flow Cytometry at Cornell (APPENDIX 10). DNA extractions were sent to University of Minnesota Genomics Center, Minneapolis, Minnesota, USA for enzyme optimization, library preparation, and RADseq / SBG Sequencing-based Genotyping. Enzymes used for digestion were PstI-HF and BtgI, and the samples were run on a Nextseq 500 High Output 150-bp SR run. Mean quality scores for all libraries were greater than or equal to Q30, indicating an accurate base call probability of 99.9%, and all barcodes were detected.

Phylogenetic methods

The sample SR raw RADseq fastq files were demultiplexed and assembled de novo using *ipyrad* (Eaton, 2014; Eaton *et al.*, 2017). This included quality filtering, clustering within samples, calculation of joint estimation of error and heterozygosity, consensus base calling, clustering across samples, and finally alignment. This concatenated SNP dataset was exported (tnt, phylip, nexus formats) for further analysis.

Phylogenies were estimated using Parsimony in *tnt,* Bayesian in BEAST 2.4.8, and Maximum Likelihood in RAxML v. 8.0. (Nixon and Carpenter, 1993; Nixon, 1999; Drummond and Rambaut, 2007; Stamatakis *et al.*, 2008; Goloboff *et al.*, 2008). The outgroups used in the analyses are in Table 2.1, and all phylogenetic trees were rooted with *Ptelea* (Nixon and Carpenter, 1993). Extractions from herbarium specimens were done for *P. alatus, P. demerarae*, and *P. manuensis*; however, RADseq library preparations were not successful so they were not used in this analysis. These species distributions include: *P. alatus* in the Northern Amazonian region, *P. demerarae* in Guyana, and finally

P. manuensis in Peru and in the western Amazonian Brazilian state of Acre. The distribution of *P. manuensis* is similar to *P. peruvianus,* so it is possible that its inclusion in the phylogeny could assist with a better estimate of the *Pilocarpus* phylogeny.

Parsimony analyses were run in TNT using the parsimony ratchet, drift, and tree-bisection-reconnection (TBR) branch swapping methods to estimate a consensus tree (Figure 2.5a). RAxML v. 8.0 was used to estimate a maximum consensus tree using 10,000 bootstrap iterations (Figure 2.5b) (Stamatakis *et al.*, 2008). BEAST 2.4.8 analyses were run using the GTR + Γ +I model, implementing an uncorrelated lognormal relaxed molecular clock as well as a Yule and birth-death process of speciation (Gernhard, 2008). Two separate runs of 30 million MCMC generations (logged every 100) converged and the effective sample sizes (ESS) were assessed positively in Tracer.

Table 2.1. *Pilocarpus* outgroups used for analyses

Subfamily	Tribe	Subtribe	Genus	Locations Collected
Rutoideae	Galipeae	Galipeinae	*Galipea*	Foz de Iguaçu, SP,BR
			Neoraputia	
			Rauia	
		Pilocarpinae	***Esenbeckia***	Piracicaba, SP, BR
			Metrodorea	Piracicaba, SP, BR
			Pilocarpus	*Many sites in BR*
			Raulinoa	
Toddalioideae	Toddalieae	Pteleinae	***Balfourodendron***	Piracicaba, SP, BR
			Helietta	Piracicaba, SP, BR
			Ptelea	Ithaca, NY, USA

*Bolded genera were collected and used as outgroups

The independent runs were combined using LogCombiner with a 20% burn in, and then were used to estimate the maximum clade credibility tree in Tree Annotator (Drummond and Rambaut, 2007) (Figure 2.5c).

The BEAST tree was chosen for further phylogenetic comparative analyses since we could calibrate the tree using two *Ptelea* fossils. *Ptelea paliuruoides,* the oldest *Ptelea* fossil, is from the mid-Eocene and was collected in the "Middle Eocene Parachute Creek

Member of the Green River Formation of Colorado and Utah"(Manchester and O'Leary, 2010). *Ptelea enervosa,* the second *Ptelea* fossil, is from the mid-Miocene and it was collected from localities on the Oregon-Idaho border (Call and Dilcher, 1995). A mid-Eocene fossil calibration was implemented as a minimum age for the *Ptelea trifoliata* node, (indicated by a red star on the phylogeny) using a log-normal distribution (Figure 2.5c). The outgroups were trimmed for subsequent phylogenetic comparative analyses using the *ape* package in R, and the branch lengths estimated with the full data set were saved for the subset of 12 taxa with phytochemical data.

Phylogenetic signal methods

Diversity signal

There are various statistical approaches to evaluating correlations of traits in phylogenetics. Phylogenetic signal is the extent that related species resemble each other (Blomberg 2003). It is important to first examine overall diversity of compounds present in each species. The phylogenetic signal for the diversity of compounds in each species was determined using Pagel's λ, calculated using the 'phylosig' function in the *phytools* R Package (Pagel, 1992; Revell, 2012). Pagel's λ ranges from 0 to 1, where λ is the transformation of the phylogeny that best predicts the variation in traits across the tree. λ =0 means that the tree topology does not affect the variation in the trait i.e. no phylogenetic signal. When λ=1 there is a perfect correlation between the trait and tree topology following a Brownian motion model, i.e. phylogenetic signal. Likelihood ratio tests are employed to test whether λ is significantly greater than 0.

Individual compounds presence/absence signal

Next, the phylogenetic signal of individual compounds was examined by using a discrete binary trait: Compound present=1, Compound absent= 0. The phylogenetic

signal of these discrete binary traits was calculated using the D metric through the R packages *caper* and *ape* (Fritz and Purvis, 2010). In the case of the D metric, D=1 indicates that a trait has evolved randomly across the tree (no signal), whereas D=0 indicates that a trait has evolved in a way correlated with phylogeny by Brownian motion. In order to measure significance, 1000 permutations are run to obtain a p-value.

Quantitative signal

Finally, the phylogenetic signal of the quantitative variation of each compound present was calculated using Blomberg's K statistic, utilizing the 'phylosig' function in the *phytools* R package (Blomberg *et al.*, 2003; Revell, 2012). K>1 indicates that traits have more phylogenetic signal than expected under Brownian motion when K=1. When K<1 related species are less similar. To test whether K is significant, 1000 randomizations can be run to generate a null distribution and provide a p-value in *phytools* (Revell, 2012). In addition, Pagel's was also calculated to assess the phylogenetic signal of quantitative variation (Pagel, 1992). An assessment of evolutionary models was done through examination of AIC values for the following methods: OU= Ornstein-Uhlenbeck; BM= Brownian motion, EB= Early Burst model.

Bioclimatic Variables

Bioclimatic variables (BIO1-BIO19) at 30-arcseconds were extracted from the WorldClim website (www.worldclim.org) using the longitude and latitude coordinates of each collection locality through the R package *raster* (Etten and Hijmans, 2010). The five most significant bioclimatic variables that differentiate niches in the genus were identified in an ecological niche modeling study of the genus (Allevato Chapter 3).

Comparative Analyses

Biosynthetic Diversity

To assess similarities of coumarin compounds due to non-independence of pathways and shared enzymes, we first created a master matrix of enzyme presence-absence data to portray the enzymes used in the biosynthesis of each coumarin compound (Figure 2.4 coumarin biosynthetic pathway and enzymes). This master matrix was next used to create a new separate enzyme presence-absence matrix for each species that only included the coumarin compounds present for that species. Each species matrix was then used to calculate the Sørensen distance for each pairwise compound comparison. A small Sørensen distance represents a very similar pathway and many shared enzymes; whereas, a large Sørensen distance represents a different pathway with few or no shared enzymes. The mean of the Sørensen distances for each species was identified as the Biosynthetic Diversity of each species (Sørensen, 1948; Junker et al., 2017). We next reconstructed the Biosynthetic Diversity for each node using Maximum Likelihood through the 'fastAnc' function in *phytools* in R (Revell, 2012). And we assessed the phylogenetic signal for the Biosynthetic Diversity of compounds in each species was determined using Pagel's λ, calculated using the 'phylosig' function in the *phytools* R Package (Pagel, 1992; Revell, 2012).

Ancestral trait reconstruction

Character reconstruction was assessed for both continuous and discrete chemical traits. Continuous traits were reconstructed for each node using Maximum Likelihood through the 'fastAnc' function in *phytools* in R (Revell, 2012). Discrete chemical traits were modeled using two methods, the continuous-time Markov chain model (Mk) (Yang *et al.*, 1995) and Stochastic character mapping (SCM) (Nielsen, 2002). Mk model

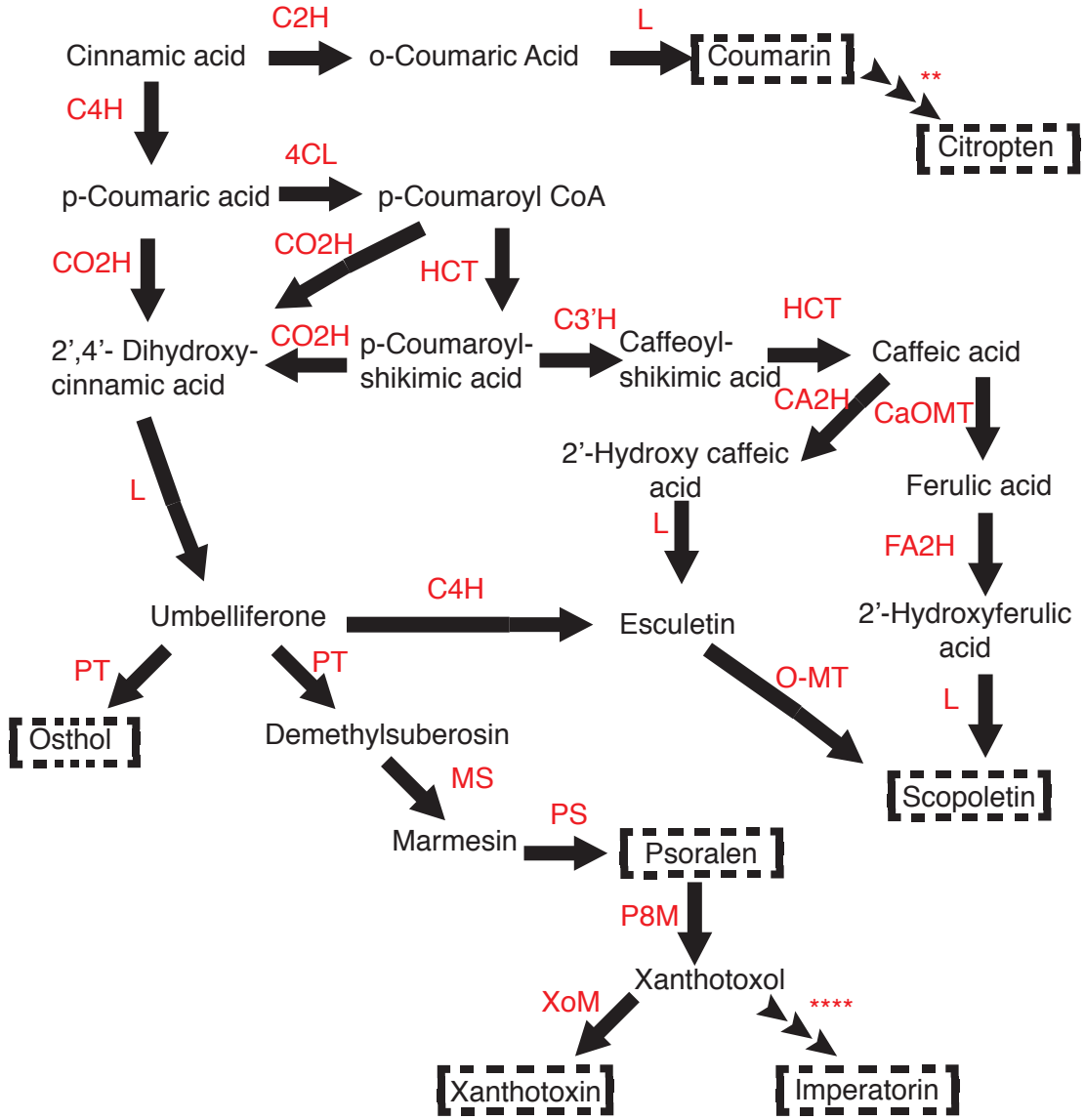


Figure 2.4. Biosynthesis of coumarins in Rutaceae. Coumarins quantified and identified in *Pilocarpus* extractions are highlighted with a dashed box. Enzymes and compound modifications are in red (Bourgaud *et al* 2006). *C2H*= Cinnamic acid 2-hydroxylase; *C4H*= cinnamic acid 4-hydroxylase; *4CL*= 4-coumarate; **=esterification; *CO2H*= 4-coumaric acid 2 hydroxylase; *HCT*= Hydroxycinnamoyl-transferase; *C3'H*= cinnamoyl ester 3'-hydroxylase; *CA2H*= caffeic acid 2-hydroxylase; *CaOMT*= caffeic acid *O*-methyltransferase; *FA2H*= Ferulic acid 2-hydroxylase; *O-MT*= *O*-methyltransferase; *PT*= Prenyltransferase; *MS*= Marmesin synthase; *PS*= Psoralen synthase; *P8M*= Psoralen 8-monooxygenase; *XoM*= Xanthotoxol *O*-methyltransferase; *L*= Lactonization; ***= O-alkylation.

was performed using 'equal rates of changes' (ER) in the R function 'rerootingMethod' of phytools (Yang *et al.*, 1995; Revell, 2012). SCM was performed in *phytools* using the 'make.simmap' function (Revell, 2012), summarizing of a set of 100 stochastic maps.

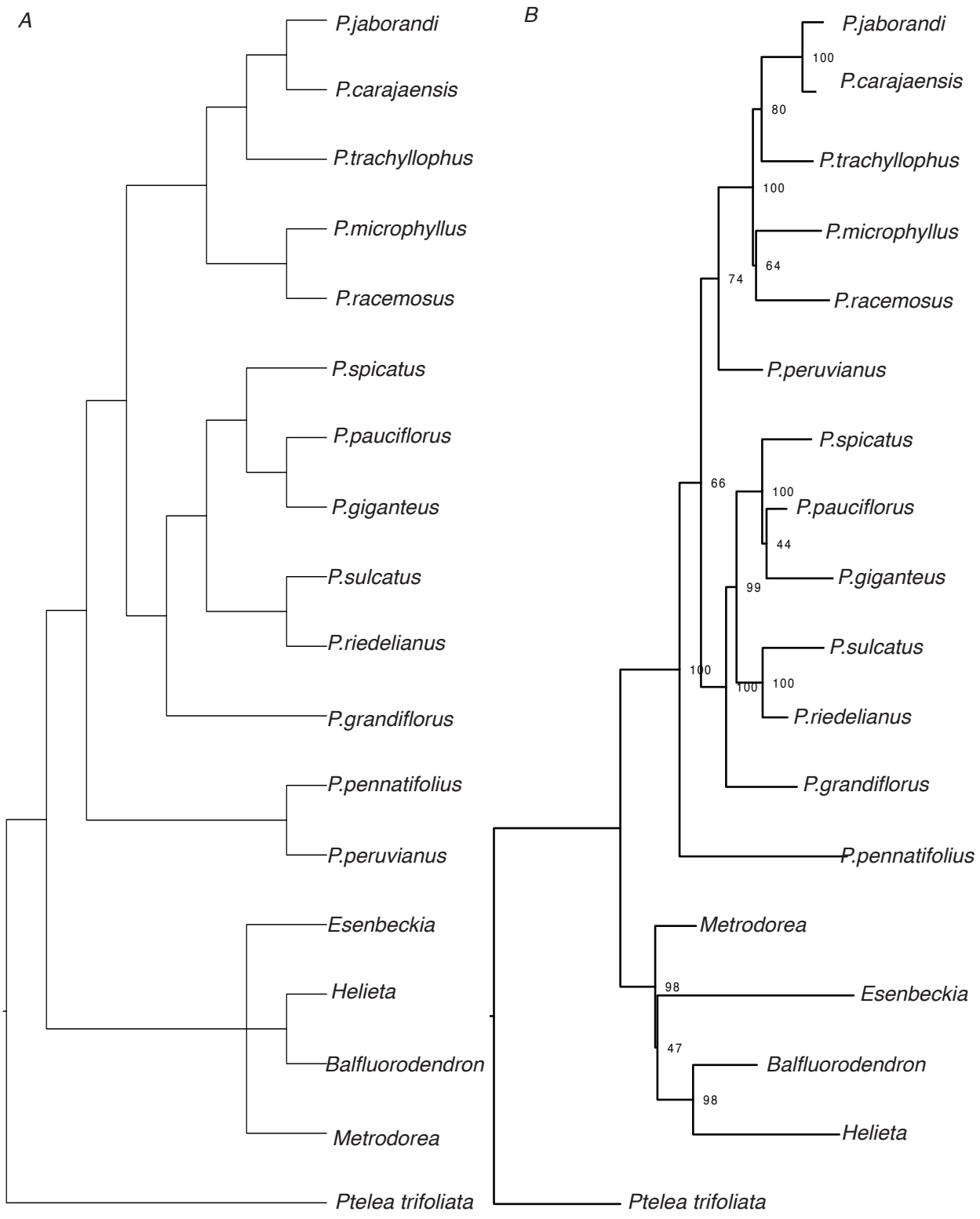
Non-phylogenetic generalized least squares (GLS) and phylogenetic GLS (pGLS)

Regressions of chemical and climatic traits were calculated in R using *ape* and *nlme* packages (Paradis *et al.*, 2004; Pinheiro *et al.*, 2017). Phylogenetic GLS was calculated in R using the *geiger* package, with branch lengths scaled using Pagel's λ (Pagel, 1992; Harmon *et al.*, 2008; Revell, 2012). Both GLS and *p*GLS models for each chemical trait were evaluated using the Akaike Information Criterion (AIC) to establish whether phylogeny improved the model's fit. A lower AIC score denotes an improved model fit (Butler and King, 2004).

RESULTS AND DISCUSSION

Phylogeny of *Pilocarpus*

All three methods (Parsimony, Bayesian, Maximum Likelihood) confirmed that *Pilocarpus* species are monophyletic, and the genus is made up of two major clades (Figure 2.5). These two major clades are geographically defined: Clade 1 (*P. microphyllus, P. trachyllophus, P. carajaensis, P. jaborandi, P. racemosus, P. peruvianus)* present in the tropical northern region of Brazil as well as Central America, and Clade 2 (*P. grandiflorus, P. riedelianus, P. sulcatus, P. spicatus, P. giganteus, and P. pauciflorus)* present in mid-coastal to southern regions of Brazil. In addition, these two clades are subtended by *P. pennatifolius*, which is determined to be the first branch of the *Pilocarpus* clade according to both Bayesian and Maximum Likelihood estimates. However, the Parsimony analysis resulted in the placement of both *P. pennatifolius* and *P. peruvianus*


A
P.jaborandi
P.carajaensis
P.trachyllophus
P.microphyllus
P.racemosus
P.spicatus
P.pauciflorus
P.giganteus
P.sulcatus
P.riedelianus
P.grandiflorus
P.pennatifolius
P.peruvianus
Esenbeckia
Helieta
Balfluorodendron
Metrodorea
Ptelea trifoliata
B
P.jaborandi
100
P.carajaensis
80
P.trachyllophus
100
P.microphyllus
64
74
P.racemosus
P.peruvianus
P.spicatus
66
100
P.pauciflorus
44
P.giganteus
99
P.sulcatus
100
100
100
P.riedelianus
P.grandiflorus
P.pennatifolius
Metrodorea
98
Esenbeckia
47
Balfluorodendron
98
Helieta
Ptelea trifoliata

C

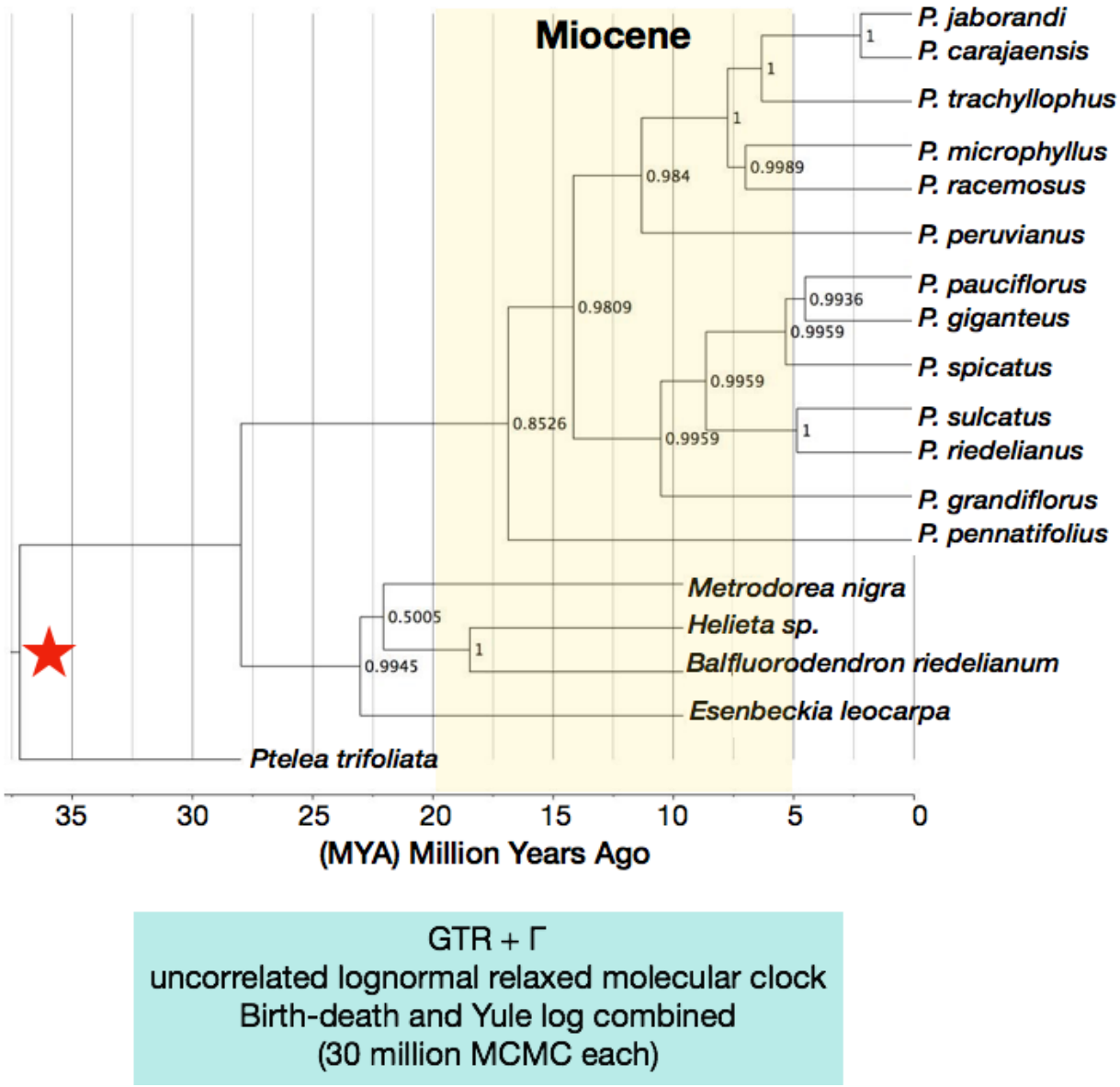


Figure 2.5. Phylogeny of *Pilocarpus* estimated using three methods. (a) Parsimony in tnt, (b) Maximum Likelihood in RaxML, and (c) Bayesian in BEAST. The red star in (c) refers to the node where two fossils of *Ptelea,* one from the mid-Eocene and one from mid-Miocene, were used to date the Bayesian tree in BEAST. The calibrated phylogeny suggests a Miocene diversification of *Pilocarpus* (in yellow).

as the first branch. This uncertainty can be confirmed in both the Bayesian and Maximum Likelihood analyses, which portray less certainty (via posterior probabilities and bootstrap support) when placing *P. peruvianus* at the base of Clade 1 (Clade with *P. microphyllus)* (Figure 2.5). There has only been one other study on the phylogeny of *Pilocarpus,* and it was estimated using Bayesian methods and with the following genes: *trn*G-S, 5.8S, ITS 1 and 2 (Oliveira, 2008). In this study, *P. peruvianus* was placed at the base, and *P. pennatifolius* was nested within the monophyletic clade of *Pilocarpus,* as a sister clade of *P. spicatus* (Oliveira, 2008).

Since *P. peruvianus* was not collected in the field and was an herbarium specimen extraction, we did not have any chemical data for this species. Therefore, we pruned it from the tree (leaving *P. pennatifolius* as the first branch), and *P. peruvianus* was thus not included in the rest of the analyses in this paper. For our comparative analyses we chose to use the branch lengths from our Bayesian phylogenetic tree, as we were able to calibrate the tree using *Ptelea* fossils (Figure 2.5c).

Diversity of coumarins and alkaloids across *Pilocarpus* is reversed

To examine the diversity of compounds among the species in this clade, the total number of compounds present in each species was scored, and the chemical diversity trait for both alkaloids and coumarins was reconstructed on the phylogeny using Maximum Likelihood (Figure 2.6a total alkaloids, Figure 2.6b total coumarins). The color-coded branches in Figure 2.6 indicate low diversity (red) to high diversity (blue) of total compounds. In Figure 2.6a, two major clades are observed: one with a higher diversity of alkaloids (Clade 1= *P. microphyllus, P. trachyllophus, P. carajaensis, P. jaborandi, P. racemosus)* and the other with a lower diversity of alkaloids (Clade 2= *P. grandiflorus, P. riedelianus, P. sulcatus, P. spicatus, P. giganteus, and P. pauciflorus).*

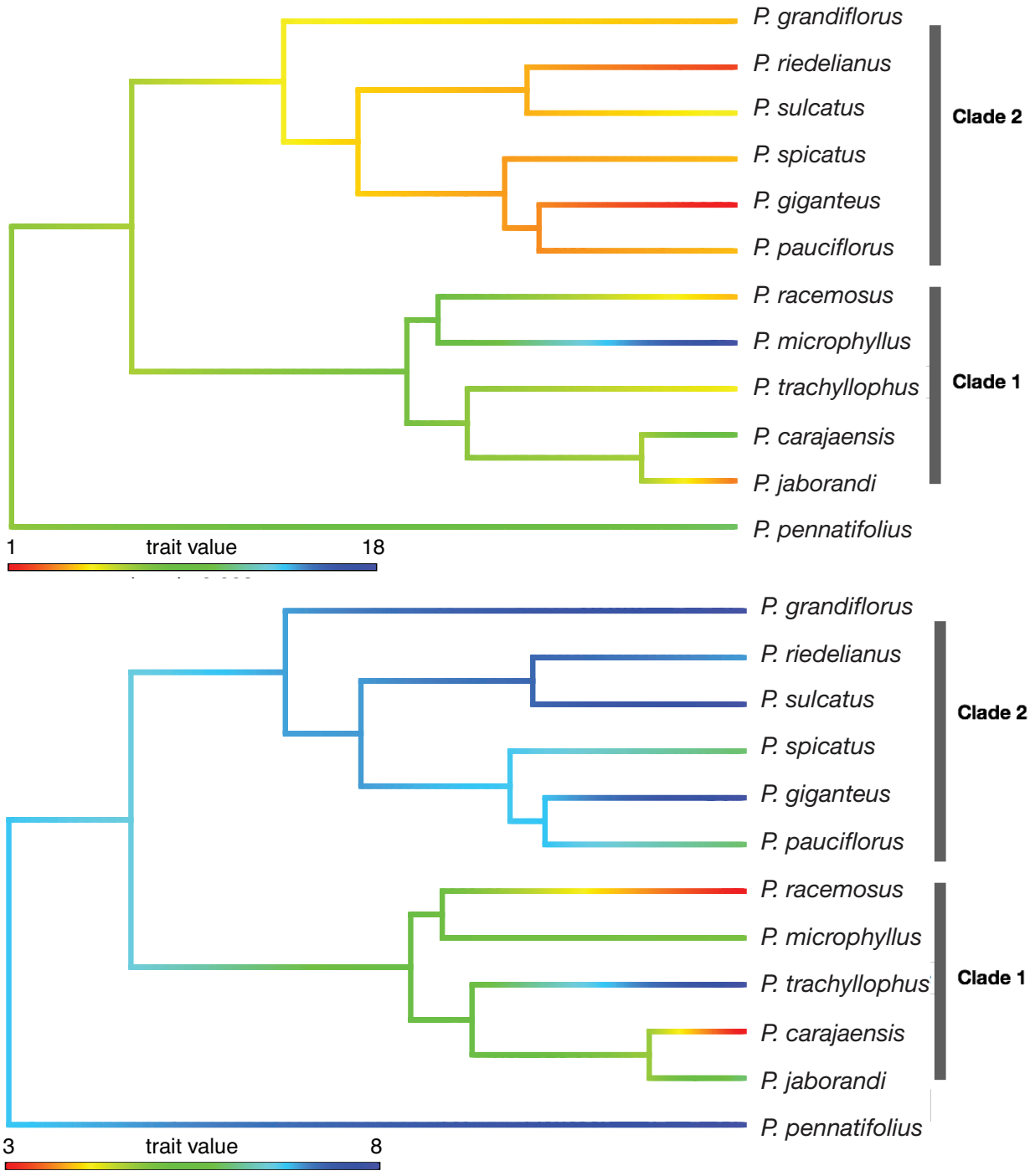


Figure 2.6. Chemical diversity (total number of compounds) reconstructed across the phylogeny of *Pilocarpus* is reversed when comparing (a) alkaloid diversity with (b) coumarin diversity. The trait value gradient bar depicts the number of compounds present for that taxon ranging from a low (red) to high (blue) number of compounds.

Interestingly, the coumarin diversity results are reversed on the phylogeny (Figure 2.6b). There are still two major clades: one with high coumarin diversity (Clade 2= *P. grandiflorus, P. riedelianus, P. sulcatus, P. spicatus, P. giganteus, and P. pauciflorus)* and the other with lower coumarin diversity (Clade 1= *P. microphyllus, P. trachyllophus, P. carajaensis, P. jaborandi, P. racemosus).* However, we now see that Clade 1 (the clade with the greatest alkaloid diversity) has the lowest coumarin diversity, and Clade 2 (the clade with the lowest alkaloid diversity) has the greatest coumarin diversity. In addition, *P. pennatifolius* at the base contains very high numbers of coumarins and next there is the opposite of escalation, a decline in numbers of coumarins, especially for Clade 1.

The phylogenetic signal calculated with Blomberg's K, found that the phylogenetic signal was not significant for alkaloid diversity or coumarin diversity (p-value > 0.05). This implies that evolutionary history was not the main contributor for the diversity of compounds, and instead there are other factors affecting the diversity of compounds. One thing to keep in mind is that these two clades are regionally distinct; Clade 1 is mostly found in northern Brazil, the Amazon, and Central America, whereas Clade 2 is mostly found in the southern and eastern regions of Brazil. This distinction is important because as these species dispersed and developed into these different regions they were exposed to a variety of environmental pressures. One interesting example of possible convergence is the similarity of coumarin and alkaloid diversity in *P. sulcatus* and *P. trachyllophus,* two sympatric species growing in the Caatinga of Brazil. It is also important to note that *P. pennatifolius,* at the base of the *Pilocarpus* clade, has a higher diversity for both alkaloids and coumarins.

Reconstruction of biosynthetic diversity across *Pilocarpus*

Specialized metabolites can also be considered non-independent traits, since they can be part of the same biosynthetic pathway or network. Although the pathway of imidazole alkaloids is unknown, the relationship between coumarin compounds and pathway enzymes are mostly known or approximated. Therefore, to determine whether there was a large correlation between coumarins, the Sørensen distance between coumarin compounds was assessed for each species. Figure 2.7. depicts the reconstruction of biosynthetic diversity using the mean of Sørensen distances for each pairwise compound comparison. Clade 2 is mostly blue, representing large Sørensen distances or few shared enzymes in the pathways for compounds present in Clade 2 species. The majority of species in Clade 1 have small Sørensen distances; therefore they have more

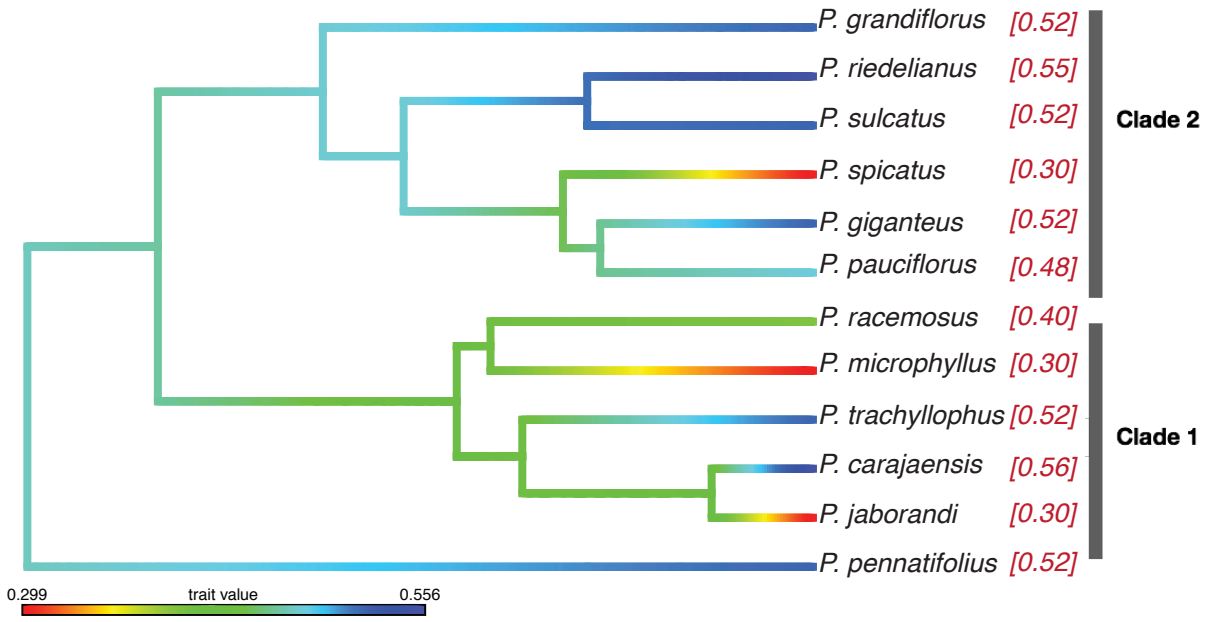


Figure 2.7. Reconstruction of the biosynthetic diversity or similarity of coumarin biosynthetic enzymes in *Pilocarpus*. Trait value gradient represents the mean of the Sørensen distances for pairwise compounds present in each species, and this exact value is labeled to the right of each taxon name. A small Sørensen distance (red) signifies small differences in enzymes, therefore a greater amount of shared enzymes in biosynthesis. A greater Sørensen distance (blue) signifies few or no shared enzymes in the biosynthesis.

shared enzymes and possibly a greater correlation between compounds present. Overall the majority of species in the *Pilocarpus* clade have larger mean Sørensen distance values, representing fewer shared biosynthesis enzymes and a lower correlation of coumarin chemical traits (Sørensen, 1948). Mean values for these species are approximately in the same range as the values found in previous studies for the correlation of chemical traits in plant species (Sørensen, 1948; Junker *et al.*, 2017). In addition, the phylogenetic signal of the mean Sørensen distance was not significant, λ=7.58^-0.05 and p-value > 0.05, thus there is a more random distribution of biosynthetic diversity across the genus.

Reconstruction of pilocarpine reveals new species with greater concentration

The discrete (stochastic character mapping of presence/absence) and continuous (Maximum Likelihood analysis of quantitative variation) reconstructions of pilocarpine are visualized in (Figure 2.8). Figure 2.8a depicts the presence/absence of pilocarpine, establishing its presence in all but two species: *P. spicatus* and *P. giganteus.* The pilocarpine trait is lost at the orange "/" on the phylogeny and later regained at the black "/", this is further confirmed by the marginal probabilities at each node. Next, in our continuous trait reconstructions we determined that *P. sulcatus* had the greatest concentration of pilocarpine, compared to the other species in the genus (* Star by species in Figure 2.8b). This discovery is intriguing, as the other species in the same clade as *P. sulcatus* (Clade 2) have some of the lowest values for concentrations of pilocarpine, including the absence of pilocarpine itself. Therefore, pilocarpine concentration is not conserved on the phylogeny and appears to be a divergent trait. In this case, there must be other factors affecting pilocarpine expression. A comparison of the distributions of species in Clade 2 confirmed that the distribution of *P. sulcatus* is

A. *B.*

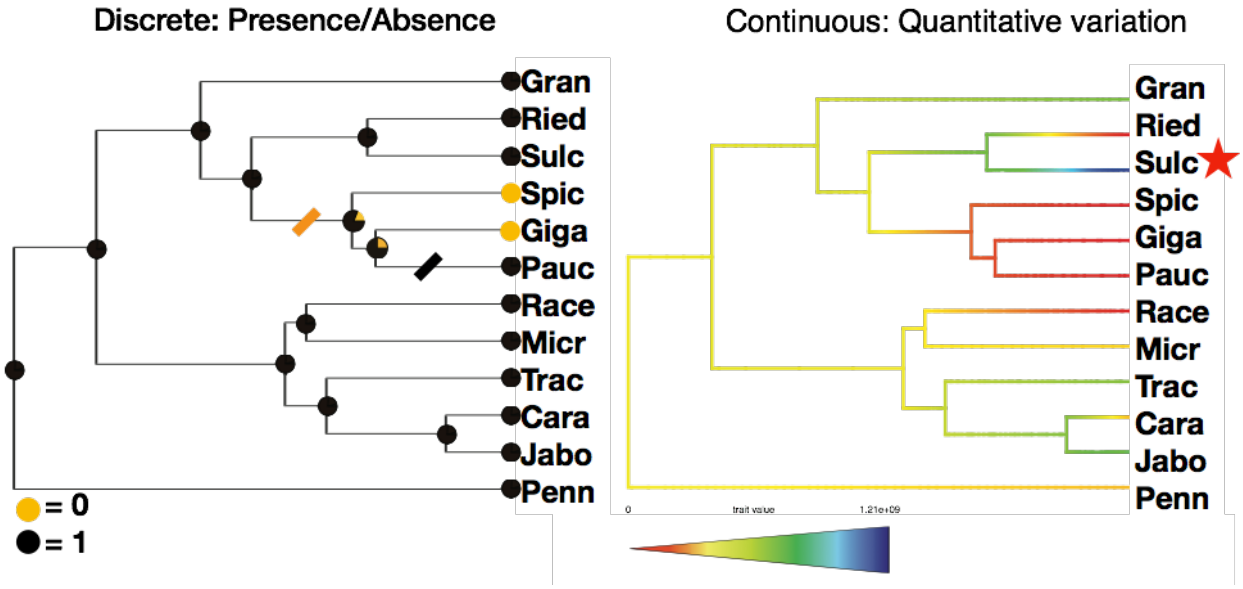


Figure 2.8. Pilocarpine trait reconstructions. (a) Stochastic character mapping (SCM) of presence (black)/absence (orange) for pilocarpine. Orange "/" marks loss of the trait and black "/" the trait is gained. (b) Maximum Likelihood analysis of quantitative variation in pilocarpine, with "star" referring to species with greatest pilocarpine concentration.

distinct. *Pilocarpus sulcatus* is the only species in Clade 2 that has a limited distribution in the Caatinga vegetative areas of southern Bahia and northern Minas Gerais. Although it is one of the smallest shrubs in the genus (height=1-2 m), its growth habit is unique with a low density of branches, and simple leaf clusters present at long internodes. In addition, *P. sulcatus* differs from other species in *Pilocarpus* as it has very conspicuous veins, unique perforated pollen, reddish leaf trichomes, and few pellucid dots.

Reconstruction of chemical traits across *Pilocarpus*

In Figure 2.9, the discrete character reconstructions for three specific compounds are visualized. In (a) C3: coumarin was detected in all of Clade 2 with one subsequent loss on the *P. spicatus* branch. This is in contrast to Clade 1, which appears to lose this trait at the base of Clade 1, though it is regained by *P. trachyllophus*. As coumarin is the

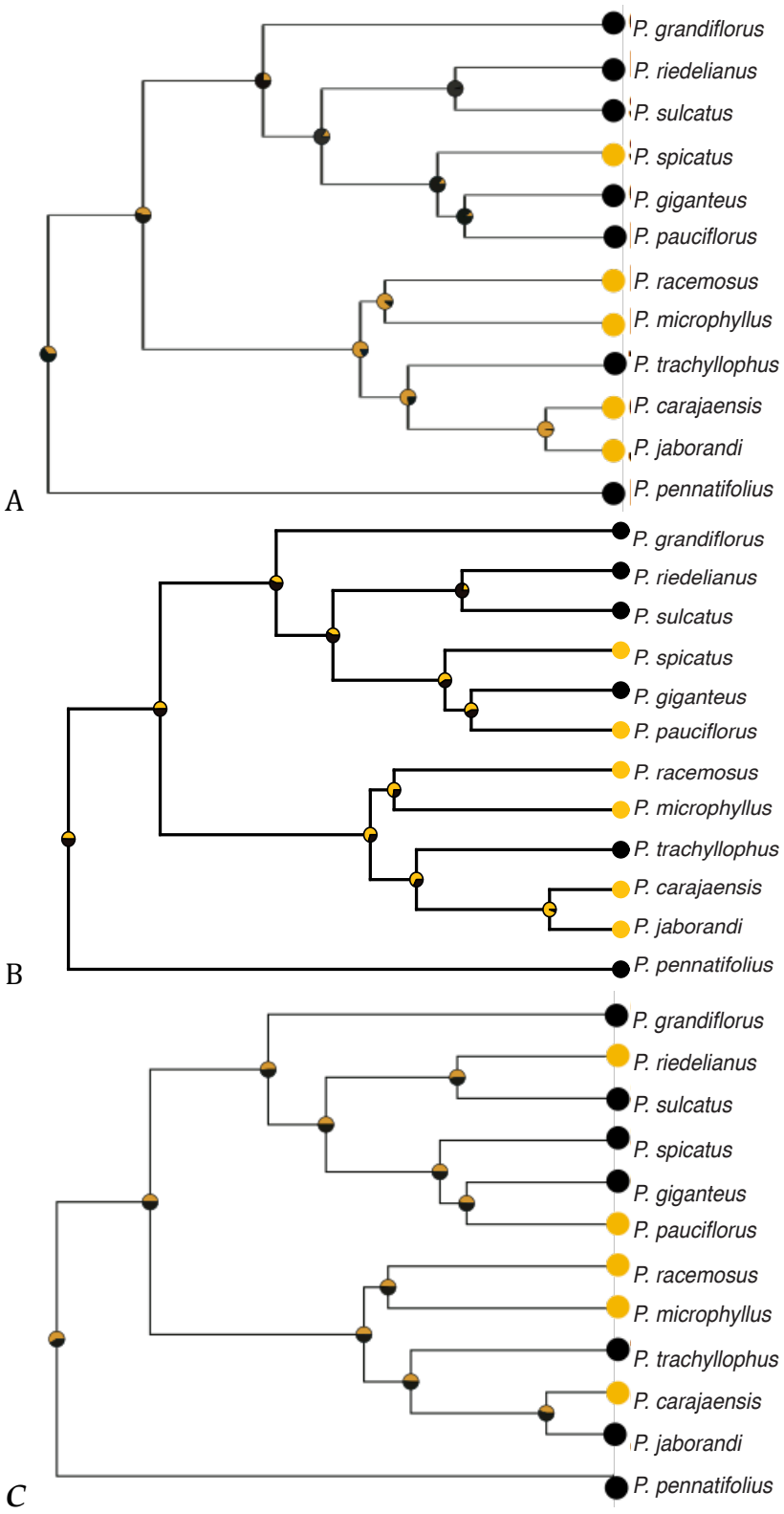


Figure 2.9. Ancestral chemical trait reconstructions. Stochastic character mapping (SCM) of discrete presence (black)/absence (orange) for each trait; C3 (one gain and one loss), C5 (two gains and two losses), C7 multiple gains and losses which appears random
(a) C3=Coumarin (b) C5= Citropten (c) C7= Osthol.

precursor to a variety of coumarins and furanocoumarins in the genus, it is possible that its disappearance in Clade 1 could be due to its role in forming other specialized coumarins further down the pathway. Since these two major clades are also geographically restricted, it is not possible to determine if these differences are due to random mutation, environmental specialization, or phylogenetic relationships. Next in (b) for C5= Citropten, there appears to be two losses and two gains in each clade. Finally, in Figure 2.9(c) the reconstruction of the coumarin osthol appears to have a somewhat random distribution across the genus, with multiple gains and losses. Although some compounds were not identified in certain species, it is possible that these species could still produce them, but their concentrations could have been too low to be detected. All of the continuous chemical traits were reconstructed on the phylogeny and are visualized in APPENDIX 3.

Interestingly it appears that C3=coumarin, C5= citropten, and C7= osthol have a similar distribution in Clade 1, being absent in all but one species, *P. trachyllophus*. One possibility is that *P. trachyllophus* needed to adapt to the precipitation and temperature extremes, through the expression of more coumarins for defense (Zobel and Brown, 1995). This could be tested with a common garden experiment, verifying if *P. trachyllophus* continued to produce these compounds in a tropical humid environment. The other species in Clade 1 could also be tested to evaluate whether translocation to a drier environment could lead to expression of C3, C5, and C7. On the other hand, the absence of C3, C5, and C7 in Clade 1 species could be a trade-off. All the other species in Clade 1 can be found in tropical humid regions, facing a greater diversity of pathogens and herbivores. These biological pressures could lead to an increase of other beneficial

defense compounds, which could be accompanied by a decrease of the C3, C5, C7 coumarins.

Phylogenetic signal of the presence/absence of individual compounds in the genus

Results from the D-metric used to assess the phylogenetic signal for the "presence/absence" of individual chemical compounds vary greatly among the compounds (Table 2.2). Two compounds, "C1" and "C6" were not analyzed, as they are present in all species of our study. Compounds with significant phylogenetic signal that are more phylogenetically conserved than one would expect under Brownian motion include: "A2", "A8", "A16", "A18", "A19", "A20". A few compounds, "A3, A17, C7", had values for the D-metric that were significantly greater than 1, implying that they are dispersed more randomly than expected across the phylogeny. Calculations of the D-metric for all other compounds were not significantly different for the probability that D=0 or the probability that D=1; therefore, no other assumptions can be made for the distribution of these compounds. It is likely that the rest of the coumarin compounds would benefit from a larger scale phylogenetic study examining the phytochemical signal in the sub-tribe, tribe, or even sub-family of Rutaceae. This expansion is not possible for the imidazole alkaloids as they are restricted to the *Pilocarpus* clade and thus can only be assessed in this manner.

Phylogenetic signal of quantitative variation in compounds and bioclimatic variables

Blomberg's K and Pagel's lambda were utilized to assess the phylogenetic signal of quantitative variables (Table 2.3). This included the concentration of referenced compounds and the average values of the five most important bioclimatic variables for each species in the genus. There are two compound concentration values that have phylogenetic signal. In terms of Blomberg's K, Psoralen has significant phylogenetic signal (K=2.43, p=0.005), whereas Imperatorin has a moderate significant phylogenetic

Table 2.2. Phylogenetic signal of presence/absence of chemical traits using the D-metric

ID	Compounds	D	P (D=1)	P (D=0)
Imidazole Alkaloids				
a1	a1	2.225038	0.704	0.108
a2	***a2***	***-2.074712***	***0.04***	0.833
a3	a3	1.856038	0.854	***0.017***
a4	a4	0.5762607	0.384	0.375
a5	a5	-0.1749972	0.18	0.585
a6	a6	-1.366992	0.151	0.609
a7	a7	-2.097808	0.172	0.632
a8	***a8***	***-2.464568***	***0.025***	0.899
a9	a9	1.206666	0.495	0.247
a10	a10	1.29005	0.514	0.265
a11	a11	-2.545128	0.169	0.641
a12	a12	-0.0485798	0.202	0.516
a13	a13	1.07127	0.48	0.196
a14	a14	0.2765816	0.167	0.532
a15	a15	-2.028621	0.148	0.639
a16	***a16***	***-2.659028***	***0.027***	0.891
a17	a17	1.997286	0.8	***0.054***
a18	***a18***	***-4.957525***	***<0.001***	0.802
a19	***a19***	***-6.42181***	***<0.001***	0.818
a20	***a20***	***-2.339238***	***0.04***	0.892
a21	a21	1.162835	0.487	0.257
a22	a22	2.94775	0.677	0.224
Coumarins				
c1	Scopoletin	*	*	*
c2	Psoralen	3.481022	0.667	0.129
c3	Coumarin	-0.0189366	0.132	0.546
c4	Xanthotoxin	1.264286	0.575	0.21
c5	Citropten	0.8296914	0.36	0.259
c6	Imperatorin B	*	*	*
c7	Osthol	2.240458	0.957	**0.007**
c8	Imperatorin	0.07397704	0.254	0.549

P= probability that D-metric is equal to 0 or 1. If P is greater than 0.05 then it is not significantly different from that value of D. When D=1 there is no phylogenetic signal, and there is a random distribution. When D=0 there is phylogenetic signal and the trait follows Brownian motion.

signal (K=1.94, p=0.049). Under Pagel's Lambda both Psoralen (λ = 1.131, p < 0.001) and Imperatorin (λ = 1.150, p < 0.001) have significant phylogenetic signal. When running all the variables under three models (Ornstein-Uhlenbeck, Brownian motion, and Early Burst), and comparing the AIC values, the Brownian motion model performs best since the AIC is lowest for the compounds with significant phylogenetic signal. This is further confirmation of our phylogenetic signal values.

Table 2.3. Phylogenetic signal of coumarin and alkaloid concentration in the genus

Variable	K	p-value	λ	p-value	OU Model AIC	BM Model AIC	EB Model AIC
Compound Concentration							
Pilocarpine	0.366	0.79	7.17E-05	1	515.399	520.1997	511.7323
Scopoletin	0.637	0.302	7.17E-05	1	49.81229	48.07356	46.14562
Psoralen	2.437	**0.005**	1.131533	**0.0007**	64.32504	60.65837	71.34135
Coumarin	0.663	0.193	7.17E-05	1	24.07424	22.07728	20.40758
Xanthotoxin	0.769	0.169	6.68E-05	1	112.1079	110.8701	108.4413
Citropten	0.493	0.544	7.17E-05	1	31.66613	32.85086	27.99947
Imperatorin B	0.676	0.199	4.95E-05	1	111.8662	110.8349	108.1996
Osthol	0.485	0.586	7.17E-05	1	33.86382	35.44863	30.19715
Imperatorin	1.942	**0.049**	1.150471	**0.0003**	123.0903	119.4236	127.4142
Bioclimatic Variables							
AMT	1.209	**0.006**	0.9491407	**0.04948**	51.4243	47.91605	51.61654
AP	0.260	0.957	7.17E-05	1	183.8934	192.998	180.2267
PCQ	0.382	0.768	7.17E-05	1	169.1206	173.3624	165.4539
PWQ	1.032	**0.013**	0.9243483	**0.04916**	156.5369	153.1888	156.7398
PDQ	1.014	**0.022**	0.7885843	0.26539	148.9799	145.9502	146.5536

OU= Ornstein-Uhlenbeck; BM= Brownian motion; EB= Early Burst model; AMT= Annual Meant Temperature, AP= Annual Precipitation, PCQ= Precipitation Coldest Quarter, PDQ= Precipitation Driest Quarter, PWQ= Precipitation Wettest Quarter

Assessing Blomberg's K for the bioclimatic variables, there is a significant phylogenetic signal for Annual Mean Temperature (AMT) [K=1.210, p= 0.006], Precipitation in the Wettest Quarter (PWQ) [K=1.032, p=0.013], and Precipitation in the Driest Quarter (PDQ) [K=1.015, p=0.022] (Table 2.3). When considering Pagel's λ there

is a significant phylogenetic signal only for AMT (λ=0.949, p=0.0495) and PWQ (λ=0.923, P= 0.0492), but no significant signal for PDQ (λ=0.789, p=0.265). This variation in PDQ significance could be because Blomberg's K has been found to have greater sensitivity in detecting smaller changes in phylogenetic signal when compared to other metrics such as Pagel's λ, and Abouheif's C_{mean} (Münkemüller Tamara *et al.*, 2012).

Regression models: the relationship between chemical traits, climate, and phylogeny

Generalized Least Squares (GLS) regressions were run in the R package *ape* comparing each bioclimatic trait to the concentration of quantified compounds. Next, phylogenetic Generalized Least squares (*p*GLS) regressions of each relationship were also run in R to determine if the models were or were not improved with the inclusion of phylogeny in the analysis. In Table 2.4, the chemical trait-climate regression model results are displayed, and the bolded AIC values are identified as the models with lower AIC scores, thus recognized as an improved/best model. Both Annual Precipitation (AP) and Precipitation in Coldest Quarter (PCQ) have AIC scores lower in GLS regressions, thereby AP and PCQ present a better fit of the data compared to a model including phylogeny. When comparing regression analyses of the remaining bioclimatic variables (AMT, PDQ, PWQ), lower AIC scores were detected when the phylogeny was taken into consideration (*p*GLS), though a few of the compounds in PDQ and PWQ had lower AIC values for the GLS regressions. For these three bioclimatic variables it is also important to note that they also had phylogenetic signal, this could therefore be an additional reason as to why the *p*GLS regressions appear more accurate. In this case there could be obfuscation caused by the phylogenetic signal for the bioclimatic traits in conjunction. Whenever AIC is lowest, the standard error for both slope and intercept is also lower. The greatest slopes were observed in C3, C5, and C7. From these analyses,

we see that the relationships between phylogenetic relationships and climate vary greatly when comparing different compounds and different bioclimatic factors. It is important to incorporate these different factors into our analyses to better understand the evolution of chemical traits.

Table 2.4. Chemical trait-Climate GLS and *p*GLS models

	GLS					pGLS				
	Y-int	**SE**	**Slope**	**SE**	**AIC**	**Y-int**	**SE**	**Slope**	**SE**	**AIC**
			AMT					**AMT**		
209	23.14	0.69	0.00	0.00	52.26	22.72	1.16	0.00	0.00	**47.12**
cc1	23.15	0.73	0.04	0.40	52.27	22.73	1.26	-0.20	0.31	**48.11**
cc2	23.65	0.45	-0.27	0.11	46.45	24.03	1.29	-0.32	0.16	**44.49**
cc3	24.41	0.61	-2.35	0.89	45.95	23.35	1.14	-1.51	0.80	**44.88**
cc4	23.91	0.50	-0.06	0.02	46.30	23.39	1.04	-0.04	0.02	**43.16**
cc5	23.52	0.63	-0.70	0.82	51.44	22.69	1.19	-0.52	0.57	**47.63**
cc6	24.73	0.46	-0.08	0.02	39.23	23.95	0.85	-0.06	0.01	**37.31**
cc7	23.42	0.62	-0.48	0.76	51.80	22.77	1.16	-0.59	0.50	**46.98**
cc8	23.74	0.43	-0.03	0.01	44.80	24.04	1.16	-0.03	0.01	**42.93**
adv	22.57	0.82	0.11	0.11	51.18	21.61	1.24	0.11	0.07	**46.23**
cdv	26.45	1.61	-0.51	0.24	47.89	24.36	1.58	-28.00	0.17	**45.70**
	Yint	**SE**	**Slope**	**SE**	**AIC**	**Yint**	**SE**	**Slope**	**SE**	**AIC**
			AP					**AP**		
209	1632.64	114.02	0.00	0.00	**174.95**	1752.01	326.99	0.00	0.00	182.49
cc1	1422.07	153.86	19.74	84.16	**180.80**	1541.58	541.90	-31.70	132.00	193.59
cc2	1435.01	121.71	7.20	29.44	**180.82**	1635.38	639.94	-27.65	77.90	193.51
cc3	1574.00	160.22	-247.44	233.87	**179.62**	1723.32	533.91	-380.74	373.70	192.47
cc4	1572.32	115.86	-10.32	5.38	**177.14**	1961.58	389.70	-21.66	6.84	185.34
cc5	1595.50	114.45	-322.90	147.92	**176.21**	1663.50	464.47	-363.87	221.95	190.81
cc6	1706.77	127.85	-13.10	4.82	**174.26**	2154.07	334.16	-24.87	5.67	180.79
cc7	1580.04	109.09	-299.10	133.90	**176.00**	1707.60	435.64	-394.86	186.40	189.21
cc8	1465.88	123.19	-1.01	2.83	**180.74**	1814.85	590.51	-6.21	6.49	192.61
adv	1335.06	176.82	18.91	23.61	**180.15**	1093.57	506.73	52.15	30.41	190.57
cdv	2055.58	358.01	-96.10	54.39	**177.63**	2782.64	507.69	-188.74	54.69	184.25
	Yint	**SE**	**Slope**	**SE**	**AIC**	**Yint**	**SE**	**Slope**	**SE**	**AIC**
			PCQ					**PCQ**		
209	345.23	76.10	0.00	0.00	**165.25**	396.53	183.27	0.00	0.00	168.59
cc1	330.44	82.39	-22.05	45.06	**165.84**	392.65	227.13	-59.32	55.41	172.73

cc2	305.00	65.94	-1.41	15.95	**166.11**	383.73	281.50	-14.88	34.31	173.81
cc3	447.09	67.98	-281.24	99.24	**159.05**	478.52	211.31	-284.74	147.51	170.23
cc4	362.56	65.07	-4.93	3.02	**163.29**	487.80	191.61	-8.32	3.37	168.31
cc5	395.60	56.61	-201.99	73.16	**159.32**	399.62	190.33	-197.40	90.96	169.40
cc6	455.29	64.29	-7.70	2.42	**157.77**	588.73	157.19	-10.58	2.67	162.70
cc7	369.46	60.94	-150.09	74.84	**162.06**	420.66	172.40	-208.80	73.77	166.97
cc8	320.48	65.71	-0.95	1.51	**165.66**	456.91	259.27	-2.89	2.85	172.86
adv	165.04	81.74	23.26	10.92	**161.63**	72.12	196.48	30.66	11.80	167.83
cdv	744.71	167.33	-69.80	25.42	**159.38**	901.85	211.51	-86.99	22.79	163.24
	Yint	**SE**	**Slope**	**SE**	**AIC**	**Yint**	**SE**	**Slope**	**SE**	**AIC**
			PDQ					**PDQ**		
209	209.37	52.92	0.00	0.00	**138.78**	125.56	71.77	10.14	9.48	144.85
cc1	137.03	37.77	-6.09	20.66	147.12	184.20	76.28	-4.82	18.61	**146.54**
cc2	108.30	24.99	12.71	6.05	**142.83**	125.56	71.77	10.14	9.48	144.85
cc3	106.07	40.36	45.40	58.91	146.53	190.74	78.35	-22.40	54.69	**146.42**
cc4	130.82	33.30	-0.12	1.55	147.21	235.27	60.57	-2.70	1.06	**140.67**
cc5	145.68	33.08	-35.45	42.74	146.42	202.17	64.40	-54.07	30.78	**143.39**
cc6	133.20	41.40	-0.19	1.56	147.20	253.22	60.11	-2.87	1.02	**139.63**
cc7	142.49	31.94	-29.48	39.22	146.56	195.77	69.32	-34.39	2.66	**145.10**
cc8	113.58	28.20	0.84	0.65	**145.35**	176.13	73.39	-0.13	0.88	146.13
adv	133.73	44.81	-0.74	5.98	147.20	162.39	80.52	1.94	4.83	**146.42**
cdv	37.99	96.25	14.43	14.62	146.11	244.75	101.76	-9.88	10.96	**145.68**
	Yint	**SE**	**Slope**	**SE**	**AIC**	**Yint**	**SE**	**Slope**	**SE**	**AIC**
			PWQ					**PWQ**		
209	405.21	50.19	0.00	0.00	155.25	420.99	82.23	0.00	0.00	**149.36**
cc1	372.74	57.69	-10.14	31.55	157.28	380.77	103.34	3.80	25.21	**153.83**
cc2	344.17	44.15	9.54	10.68	156.47	352.60	106.48	6.82	14.03	**154.22**
cc3	270.00	50.90	175.21	74.30	152.10	319.26	88.35	105.89	68.01	**151.90**
cc4	354.74	50.83	0.43	2.36	157.37	426.33	88.66	-2.26	1.78	**152.78**
cc5	351.31	52.07	18.83	67.29	157.31	385.40	90.79	-5.88	50.86	**154.49**
cc6	320.41	61.10	1.99	2.31	156.54	426.02	96.20	-1.74	1.88	**153.52**
cc7	371.62	49.73	-26.18	61.08	157.19	395.34	99.00	-17.26	42.36	**153.66**
cc8	344.79	45.25	0.81	1.04	156.70	385.25	117.47	0.02	1.30	**153.86**
adv	437.43	61.15	-13.09	8.17	154.66	413.47	98.91	-4.20	6.80	**154.04**
cdv	160.23	139.45	31.54	21.18	155.00	441.65	141.05	-8.15	15.20	**153.52**

AP= Annual Precipitation; AMT= Annual Meant Temperature; PCQ= Precipitation Coldest Quarter; PDQ= Precipitation Driest Quarter; PWQ= Precipitation Wettest Quarter; 209= Pilocarpine concentration; adv= alkaloid diversity; cdv= coumarin diversity

***Bolded AIC values indicate the model that best predicts chemical variation. A lower AIC value represents a better model score.*

Conclusions

Although most agree that related species share similar traits, it is also very possible that related species have divergent chemistry. This difference in chemistry could be driven by many factors including speciation due to expansion into different environments or differentiation into different niche/ adaptive spaces in the same location (Mraja *et al.*, 2011; Becerra, 2015). On the other hand, a lack of correlation of specialized chemistry with phylogeny could be due to convergent evolution, induction by environmental pressures, variation in development, or silent metabolism (Wink, 2003; Carrari *et al.*, 2006; Pichersky and Lewinsohn, 2011). Within a species, compounds such as alkaloids and coumarins have been shown to differ across broad groups in type and quantity based on different environmental conditions, such as changes in temperature, elevation, precipitation and soil type (Levin, 1976; Dixon and Paiva, 1995; Zobel and Brown, 1995; Martz *et al.*, 2009). The advancement of phylogenetic comparative methods enables the discovery of relationships between the environment and phytochemical diversity.

Together these complementary approaches have aided in a better understanding of the coumarins and alkaloids present in the genus *Pilocarpus,* as well as provided a background for studies looking at the biosynthesis of these compounds. Future work identifying biosynthetic enzyme presence and absence in the different species would further aid in understanding the evolution of these biosynthetic pathways. Although expanding the sample of genera would potentially provide greater significance for phylogenetic signal, this was not possible for the imidazole alkaloids as they are only present in the small scale of the genus *Pilocarpus*. On the other hand, more research can be done to expand the detection of coumarins, more broadly within Rutaceae, at the level of tribe or even at the sub-family level within Rutoideae, as this would further

assist in identifying the evolutionary patterns of coumarins. Rutaceae is a phylogenetically complex family with a rich diversity of chemistry. The combination of phylogenetic comparative methods, environmental factors, and biochemical data is beneficial for studies on the evolution of chemical diversity, conservation of bio-diverse areas, as well as ecological studies.

ACKNOWLEDGMENTS

The author(s) thank various individuals and funding institutes for their assistance throughout this project. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1650441). In addition, this work was supported by Harold E. Moore Jr. and Robert T. Clausen Endowment Fund, BSA, SSB, ASPT, Garden Club of America, Einaudi Center for International Studies, Latin American Studies Program of Cornell, and the Cornell Graduate School. We would like to thank Professor Milton Groppo (USP-RP), Professor Pedro Dias (USP-SP), and Professor Marcelo Caxambu (Universidade Tecnologica Federal do Parana), Dr. Rubens Coelho (USP-SP), Alex Popovkin (Bahia, Brazil), Lukas Daneu (CEPEC Herbaria) for helping with the collection of *Pilocarpus* specimens.; Dr. Jose Sena (Quercegen) for assistance in collection of *Pilocarpus* specimens in northern Brazil; Iguaçu Itaipu and Parque Estadual Lago Azul, PR for allowing the collection of specimens from their property for helping with the collection of *Pilocarpus* specimens. Dr. Adrienne Roeder for assistance with Flow Cytometry. Dr. Alexandra Christine Helena Frankland Sawaya for providing the updated extraction protocol for imidazole alkaloids; Instituto Agronomico in Campinas, SP, Brazil for analyzing the soil samples and the statistical consulting group of Cornell University for confirming the statistical

analyses. We would like to acknowledge Dr. Eloy Rodriguez and Dr. Robert Raguso for thoughtful discussions throughout this project. PM thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brasil) for a research fellowship and Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp - grant 2008/58035-6) for research support.

Plants collected in public locations were collected under ICMBio SISBIO permit #52758-1, allowing for the collection of botanical, fungal and microbiological specimens. In addition, plants were collected in collaboration with and under the permits of Professor Milton Groppo, Professor Pedro Dias, and Professor Marcelo Caxambu.

LITERATURE CITED

Avancini G, Abreu IN, Saldaña MDA, Mohamed RS, Mazzafera P. 2003. Induction of pilocarpine formation in jaborandi leaves by salicylic acid and methyljasmonate. Phytochemistry **63**, 171–175.

Barry CS, McQuinn RP, Thompson AJ, Seymour GB, Grierson D, Giovannoni JJ. 2005. Ethylene insensitivity conferred by the Green-ripe and Never-ripe 2 ripening mutants of tomato. Plant Physiology **138**, 267–275.

Becerra JX. 2015. Macroevolutionary and geographical intensification of chemical defense in plants driven by insect herbivore selection pressure. Current Opinion in Insect Science **8**, 15–21.

Bisby FA, Vaughan JG, Wright CA. 1980. *Chemosystematics: principles and practice*. Published for the Systematics Association by Academic Press.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution; International Journal of Organic Evolution **57**, 717–745.

Bourgaud F, Hehn A, Larbat R, Doerper S, Gontier E, Kellner S, Matern U. 2006. Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochemistry Reviews **5**, 293–308

Boyko A, Kovalchuk I. 2008. Epigenetic control of plant stress response. Environmental and Molecular Mutagenesis **49**, 61–72.

Boyko A, Kovalchuk I. 2011. Genome instability and epigenetic modification—heritable responses to environmental stress? Current Opinion in Plant Biology **14**, 260–266.

Butler MA, King AA. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist **164**, 683–695.

Call VB, Dilcher DL. 1995. Fossil Ptelea Samaras (Rutaceae) in North America. American Journal of Botany **82**, 1069–1073.

Carrari F, Baxter C, Usadel B, *et al.* 2006. Integrated Analysis of Metabolite and Transcript Levels Reveals the Metabolic Shifts That Underlie Tomato Fruit Development and Highlight Regulatory Aspects of Metabolic Network Behavior. PLANT PHYSIOLOGY **142**, 1380–1396.

Dixon R, Paiva N. 1995. Stress-Induced Phenylpropanoid Metabolism. The Plant Cell **7**, 1085–1097.

Dixon RA. 2001. Natural products and plant disease resistance. Nature.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**, 214.

Durand-Hulak M, Dugrand A, Duval T, Bidel LPR, Jay-Allemand C, Froelicher Y, Bourgaud F, Fanciullino A-L. 2015. Mapping the genetic and tissular diversity of 64 phenolic compounds in Citrus species using a UPLC-MS approach. Annals of Botany **115**, 861–877.

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

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. Systematic Biology **66**, 399–412.

Etten J van, Hijmans RJ. 2010. A Geospatial Modelling Approach Integrating Archaeobotany and Genetics to Trace the Origin and Dispersal of Domesticated Plants. PLOS ONE **5**, e12060.

Fine PVA, Metz MR, Lokvam J, Mesones I, Zuñiga JMA, Lamarre GPA, Pilco MV, Baraloto C. 2013. Insect herbivores, chemical innovation, and the evolution of habitat specialization in Amazonian trees. Ecology **94**, 1764–1775.

Firn RD, Jones CG. 2000. The evolution of secondary metabolism – a unifying model. Molecular Microbiology **37**, 989–994.

Firn RD, Jones CG. 2003. Natural products ? a simple model to explain chemical diversity. Natural Product Reports **20**, 382.

Fraenkel GS. 1959. The Raison d'Être of Secondary Plant Substances. Science **129**, 1466–1470.

Fritz SA, Purvis A. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conservation Biology **24**, 1042–1051.

Geber MA, Griffen LR. 2003. Inheritance and Natural Selection on Functional Traits. International Journal of Plant Sciences **164**, S21–S42.

Gernhard T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology **253**, 769–778.

Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. Cladistics **24**, 774–786.

Gray AI, Waterman PG. 1978. Coumarins in the Rutaceae. Phytochemistry **17**, 845–864.

Harborne JB 1984. *Plant chemosystematics*. London ; Orlando, Fla: Academic Press.

Harborne JB 1993. *Introduction to ecological biochemistry*. London ; San Diego: Academic Press.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics **24**, 129–131.

Hegnauer R. 1994. *Chemotaxonomie der Pflanzen*. Birkhäuser Basel.

Helen C. De S. Abbott. 1886. *Certain Chemical Constituents of Plants Considered in Relation to Their Morphology and Evolution*. Botanical Gazette.

Junker RR, Kuppler J, Amo L, *et al.* 2017. Covariation and phenotypic integration in chemical communication displays: biosynthetic constraints and eco-evolutionary implications. The New Phytologist.

Kim HK, Verpoorte R. 2010. Sample preparation for plant metabolomics. Phytochemical Analysis **21**, 4–13.

Latta RG, Linhart YB. 1997. Path Analysis of Natural Selection on Plant Chemistry: The Xylem Resin of Ponderosa Pine. Oecologia **109**, 251–258.

Levin DA. 1976. The Chemical Defenses of Plants to Pathogens and Herbivores. Annual Review of Ecology and Systematics **7**, 121–159.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science (New York, N.Y.) **290**, 1151–1155.

Mabry TJ, Ulubelen A. 1980. Chemistry and utilization of phenylpropanoids including flavonoids, coumarins, and lignans. Journal of agricultural and food chemistry **28**, 188–196.

Manchester SR, O'Leary EL. 2010. Phylogenetic Distribution and Identification of Fin-winged Fruits. The Botanical Review **76**, 1–82.

Martz F, Peltola R, Fontanay S, Duval RE, Julkunen-Tiitto R, Stark S. 2009. Effect of Latitude and Altitude on the Terpenoid and Soluble Phenolic Composition of Juniper (Juniperus communis) Needles and Evaluation of Their Antibacterial Activity in the Boreal Zone. Journal of Agricultural and Food Chemistry **57**, 9575–9584.

Masson J, Liberto E, Beolor J-C, Brevard H, Bicchi C, Rubiolo P. 2016. Oxygenated heterocyclic compounds to differentiate Citrus spp. essential oils through metabolomic strategies. Food Chemistry **206**, 223–233.

McNair JB. 1935. Angiosperm phylogeny on a Chemical Basis. Bulletin of the Torrey Botanical Club **62**, 515–532.

Mraja A, Unsicker SB, Reichelt M, Gershenzon J, Roscher C. 2011. Plant Community Diversity Influences Allocation to Direct Chemical Defence in Plantago lanceolata. PLoS ONE **6**, e28055.

Münkemüller Tamara, Lavergne Sébastien, Bzeznik Bruno, Dray Stéphane, Jombart Thibaut, Schiffers Katja, Thuiller Wilfried. 2012. How to measure and test phylogenetic signal. Methods in Ecology and Evolution **3**, 743–756.

Murray RDH. 2002. The Naturally Occurring Coumarins. Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products. Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products. Springer, Vienna, 1–619.

Nielsen R. 2002. Mapping mutations on phylogenies. Systematic Biology **51**, 729–739.

Nixon KC. 1999. The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. Cladistics **15**, 407–414.

Nixon KC, Carpenter JM. 1993. On Outgroups. Cladistics **9**, 413–426.

Oliveira PD de. 2008. Filogenética de Pilocarpinae (Rutaceae).

Pagel MD. 1992. A method for the analysis of comparative data. Journal of Theoretical Biology **156**, 431–442.

Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. 2014. Enzyme Promiscuity: Engine of Evolutionary Innovation. Journal of Biological Chemistry **289**, 30229–30236.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20**, 289–290.

Parent J-S, Martínez de Alba AE, Vaucheret H. 2012. The origin and effect of small RNA signaling in plants. Frontiers in Plant Science **3**.

Penfold AR. 1927. *The Essential Oil of Eucalyptus Micrantha (De Candolle)*.

Pichersky E, Lewinsohn E. 2011. Convergent Evolution in Plant Specialized Metabolism. Annual Review of Plant Biology **62**, 549–566.

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2017. nlme: Linear and nonlinear mixed effects models.

Price PW, Waring GL, Julkunen-Tiitto R, Tahvanainen J, Mooney HA, Craig TP. 1989. Carbon-nutrient balance hypothesis in within-species phytochemical variation ofSalix lasiolepis. Journal of Chemical Ecology **15**, 1117–1131.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution **3**, 217–223.

Rønsted N, Symonds MRE, Birkholm T, *et al.* 2012. Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of amaryllidaceae. BMC Evolutionary Biology **12**, 182.

Rueffler C, Dooren TJMV, Leimar O, Abrams PA. 2006. Disruptive selection and then what? Trends in Ecology & Evolution **21**, 238–245.

Santos AP, Moreno PRH. 2004. Pilocarpus spp.: a survey of its chemical constituents and biological activities. Revista Brasileira de Ciências Farmacêuticas **40**, 116–137.

Sawaya A, Abreu IN, Andreazza NL, Eberlin MN, Mazzafera P. 2008. HPLC-ESI-MS/MS of Imidazole Alkaloids in Pilocarpus microphyllus. Molecules **13**, 1518–1529.

Silva MF das GF da, Gottlieb OR, Ehrendorfer F. 1988. Chemosystematics of theRutaceae: Suggestions for a more natural taxonomy and evolutionary interpretation of the family. Plant Systematics and Evolution **161**, 97–134.

Skorupa LA. 1996. Revisao Taxonomica de Pilocarpus Vahl (Rutaceae). Ph.D. Thesis.

Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol. Skr. **5**, 1–34.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Systematic Biology **57**, 758–771.

Thaler JS, Humphrey PT, Whiteman NK. 2012. Evolution of jasmonate and salicylate signal crosstalk. Trends in Plant Science **17**, 260–270.

Boyko A, Kovalchuk I. 2008. Epigenetic control of plant stress response. Environmental and Molecular Mutagenesis **49**, 61–72.

Boyko A, Kovalchuk I. 2011. Genome instability and epigenetic modification—heritable responses to environmental stress? Current Opinion in Plant Biology **14**, 260–266.

Call VB, Dilcher DL. 1995. Fossil Ptelea Samaras (Rutaceae) in North America. American Journal of Botany **82**, 1069–1073.

Dixon RA. 2001. Natural products and plant disease resistance. Nature.

Geber MA, Griffen LR. 2003. Inheritance and Natural Selection on Functional Traits. International Journal of Plant Sciences **164**, S21–S42.

Junker RR, Kuppler J, Amo L, *et al.* 2017. Covariation and phenotypic integration in chemical communication displays: biosynthetic constraints and eco-evolutionary implications. The New Phytologist.

Latta RG, Linhart YB. 1997. Path Analysis of Natural Selection on Plant Chemistry: The Xylem Resin of Ponderosa Pine. Oecologia **109**, 251–258.

Manchester SR, O'Leary EL. 2010. Phylogenetic Distribution and Identification of Fin-winged Fruits. The Botanical Review **76**, 1–82.

Masson J, Liberto E, Beolor J-C, Brevard H, Bicchi C, Rubiolo P. 2016. Oxygenated heterocyclic compounds to differentiate Citrus spp. essential oils through metabolomic strategies. Food Chemistry **206**, 223–233.

Parent J-S, Martínez de Alba AE, Vaucheret H. 2012. The origin and effect of small RNA signaling in plants. Frontiers in Plant Science **3**.

Rueffler C, Dooren TJMV, Leimar O, Abrams PA. 2006. Disruptive selection and then what? Trends in Ecology & Evolution **21**, 238–245.

Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol. Skr. **5**, 1–34.

Urano K, Kurihara Y, Seki M, Shinozaki K. 2010. 'Omics' analyses of regulatory networks in plant abiotic stress responses. Current Opinion in Plant Biology **13**, 132–138.

Vialart G, Hehn A, Olry A, *et al.* 2012. A 2-oxoglutarate-dependent dioxygenase from Ruta graveolens L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. The Plant Journal: For Cell and Molecular Biology **70**, 460–470.

Waterman PG. 1975. Alkaloids of the rutaceae: their distribution and systematic significance. Biochemical Systematics and Ecology **3**, 149–180.

Waterman PG. 1990. Chemosystematics of the Rutaceae: Comments on the interpretation ofda Silva & al. Plant Systematics and Evolution **173**, 39–48.

Waterman PG, Hussain RA. 1983. Chemosystematics in the Rutaceae: A chemotaxonomic survey of flavonoids and monoterpenes in leaves of Acmadenia species. Botanical Journal of the Linnean Society **86**, 227–235.

Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64**, 3–19.

Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141**, 1641–1650.

Zhang J. 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution **18**, 292–298.

Zobel AM, Brown SA. 1995. Coumarins in the interactions between the plant and its environment. Allelopathy Journal.

CHAPTER 3

Niche overlap and phylogeography in the genus *Pilocarpus* (Rutaceae)

Daniella M. Allevato [1], Kevin C. Nixon [1]

ABSTRACT

- *Background and aims*: *Pilocarpus* is a Neotropical genus in the Rutaceae family known as the sole source of the alkaloid pilocarpine, a drug used in the treatment of glaucoma. *Pilocarpus* is an extremely bioactive genus rich in alkaloids, terpenoids and coumarins. Overharvesting of *Pilocarpus microphyllus* has led to near extinction of natural populations of this species; therefore, alternative sources of pilocarpine in other species of *Pilocarpus,* as well as an understanding of the important factors needed for increasing pilocarpine yields would greatly help reduce this destruction and conserve diversity in natural populations. Several studies have found that different environmental conditions such as temperature, elevation, precipitation and soil type can alter plant phytochemicals, including alkaloids, terpenoids and coumarins. The main aim of this research was to understand the bioclimatic factors affecting the distribution of species in this genus to guide conservation management and field collections of *Pilocarpus.*
- *Methods:* Locality data from 2655 geo-referenced herbarium specimens was used to estimate an ecological niche model for each species in Maxent using the 19 bioclimatic variables from WorldClim. These niche models identified the

1. Section of Plant Biology, School of Plant Sciences, Cornell University, Ithaca, NY, USA

environmental variables most significant for the distribution of species and were used to assess niche overlap between *Pilocarpus* species.

- *Key results:* Analysis of niche overlap found that *P. pauciflorus* (mean *D*=0.56, mean *I*=0.64), *P. riedelianus* (mean *D*=0.4, mean *I*=0.65), and *P. spicatus* (mean *D*=0.39, mean *I*=0.67) have the greatest niche overlap, whereas *P. pennatifolius* has the least similarity or overlap of niches in the genus (mean *D*=0.19, mean *I*=0.41)
- *Conclusions:* Ecological niche modeling in conjunction with niche overlap provides insight into the distribution of this genus, which can guide management strategies and field collections. In addition, ancestral niche reconstructions can aid in understanding the evolution and speciation of species in this genus.

KEYWORDS: Ecological niche modeling, *Pilocarpus,* Rutaceae, niche overlap, Brazil, climate

INTRODUCTION

Each species has a preferred set of biotic and abiotic environmental conditions that allow for its survival, growth, and the success of its future progeny. This possible range of environmental conditions intrinsic to a species is known as the fundamental niche (Hutchinson, 1957). This fundamental niche is constrained by a variety of factors including competition, biotic interactions, and human interference; therefore, the subset of locations where one can actually find a species is known as its realized niche (Hutchinson, 1957). For a given environmental variable, a species will be found along an interval of that environmental gradient; nevertheless, there is usually an optimum where one can find the greatest abundance of a particular species. As such, the response of a species to a particular environmental factor has been found to mostly follow a unimodal response, with the greatest abundance of a species occurring at an optimum for each environmental variable (Braak, 1986; Peterson *et al.*, 2011; Jamil *et al.*, 2014).

Species distribution models (SDM) or ecological niche models (ENM) can be used to estimate the environmental requirements or fundamental niches of each species, to determine the potential ecological niche of a species (Elith and Leathwick, 2009). This is done by correlating environmental factors with geo-referenced occurrence data for a particular species (Buckley *et al.*, 2010). These models have been used widely in ecology and conservation biology as well as in studies examining speciation, niche conservatism, and climate change (Franklin, 2010; Peterson, 2011; Franklin, 2013; Porfirio *et al.*, 2014). There are several issues with ENM's, including bias towards areas that are more easily accessible (and more heavily collected) as well as bias in areas where species could occur but do not, due to a variety of reasons including changes of land use through time, human intervention/cultivation, dispersal barriers etc. (Godsoe, 2010). Nevertheless, these models are believed to be effective for the prediction of

potential species distributions. In particular, a machine learning method known as Maxent is widely used to determine species habitat associations (Phillips *et al.*, 2006). Maxent uses presence only data in conjunction with background locations, comparing their probability densities. Probability densities represent the probability or suitability of finding a species in a particular location. This is estimated by investigating the unique values for a set of bioclimatic variables (e.g. temperature, precipitation, seasonality) at each geo-referenced point. The probability densities of the presence-only data is compared with that of the background locations, minimizing the relative entropy between probability densities, to predict distribution of a species within each cell (Phillips *et al.*, 2006; Elith *et al.*, 2011). The background locationsare used to assist in alleviating the error due to only having presence data points in the analysis.

In this paper we are exploring the distribution pattern of the neotropical genus *Pilocarpus*. This genus comprises 16 species, with a distribution from the southern border of Mexico to the northern provinces of Argentina (Skorupa, 1996, 1998, 1999). The center of species diversity of *Pilocarpus* is Brazil, where 14 species can be found, 11 of which are endemic. Overharvesting of wild *Pilocarpus* populations has occurred since it is the only source of the imidazole alkaloid pilocarpine to treat glaucoma. This has led to the near extinction of species in this genus including: *P. microphyllus, P. jaborandi, P. trachyllophus,* and *P. alatus* (IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis), 1992). Although many species are collected in the wild, studies have shown that pilocarpine production is quite variable among species in the genus (Sawaya *et al.*, 2011; Allevato Chapter 2). This could be due to a variety of reasons including environmental adaptation to pressures/stressors in the environment. In the case of areas with little seasonal variation you would expect species to have a lower capacity or less environmental plasticity (Molina-Montenegro and Naya, 2012). There is

still much to be understood at multiple levels of hierarchy including phylogenetic diversity between *Pilocarpus* species, as well as diversity within species due to environmental/horticultural factors that could lead to greater pilocarpine for each species.

The objectives of this study were to (1) Determine the climatic preferences of all species of *Pilocarpus* (2) Explore the niche overlap of species in the genus to assess the possibility for species coexistence (3) Determine the evolution of climatic suitability in the genus with ancestral reconstruction. Ecological niche modeling in Maxent and ancestral reconstructions were completed using locality data from 2655 herbarium specimens, to evaluate areas of diversity and identify environmental variables contributing to the distribution patterns.

MATERIALS AND METHODS

Herbarium specimens used in analysis

Locality data / geographic coordinate data were compiled based on 2655 herbarium specimens from two sources: Global Biodiversity Information Facility (GBIF) (www.gbif.org), and INCT Brazilian Herbaria (http://inct.florabrasil.net) (APPENDIX 4) ('GBIF Occurrence Download', 2016; 'INCT', 2018). After assessing and removing duplicates and inappropriate coordinates, specimens remaining included 65 *P. microphyllus, 431 P. pennatifolius, 320 P. spicatus, 26 P. giganteus, 44 P. grandiflorus, 140 P. pauciflorus, 19 P. jaborandi, 5 P. alatus, 3 P. demerareae, 26 P. carajaensis, 42 P. sulcatus, 99 P. trachyllophus, 152 P. riedelianus, 87 P. racemosus, 6 P. manuensis and 8 P. peruvianus*

Bioclimatic Variables

Bioclimatic variables (BIO1-BIO19) at 30-arcseconds (identified in Table 3.1) were extracted from the WorldClim website (www.worldclim.org) using the longitude and

latitude coordinates of each collection locality through the R package *raster* (Etten and Hijmans, 2010). Some of these bioclimatic variables vary more than others, and this is demonstrated by their coefficient of variation, which ranges from 5.2% to 71%(APPENDIX 5).

Table 3.1 Bioclimatic variables from WorldClim

BIO1 = Annual Mean Temperature **BIO2** = Mean Diurnal Range (Mean of monthly (max temp - min temp)) **BIO3** = Isothermality (BIO2/BIO7) (* 100) **BIO4** = Temperature Seasonality (standard deviation *100) **BIO5** = Max Temperature of Warmest Month **BIO6** = Min Temperature of Coldest Month **BIO7** = Temperature Annual Range (BIO5-BIO6) **BIO8** = Mean Temperature of Wettest Quarter **BIO9** = Mean Temperature of Driest Quarter **BIO10** = Mean Temperature of Warmest Quarter	**BIO11** = Mean Temperature of Coldest Quarter **BIO12** = Annual Precipitation **BIO13** = Precipitation of Wettest Month **BIO14** = Precipitation of Driest Month **BIO15** = Precipitation Seasonality (Coefficient of Variation) **BIO16** = Precipitation of Wettest Quarter **BIO17** = Precipitation of Driest Quarter **BIO18** = Precipitation of Warmest Quarter **BIO19** = Precipitation of Coldest Quarter

Ecological Niche-Modeling / Species Distribution Modeling

Ecological niche modeling (ENM) was performed using Maxent in R version 3.4.3 (http://www.cs.princeton.edu/~schapire/maxent) with presence-only data from herbarium specimens of each species (Elith *et al.*, 2011). Maxent uses maximum entropy to predict a potential distribution of species using presence-only geo-location data, generated background data, and bioclimatic environmental variables. Models were run using default parameters and pseudo-absences were randomly sampled from a restricted background (80km radius) to prevent model over-fitting. We sampled 10,000 random points across each background point for every predictor (Phillips and Dudík, 2008). To test the performance of the model we assessed the AUC (Area under the receiver operating characteristic Curve). This area represents the ability to discriminate

and correctly classify a randomly chosen point in space as being a suitable or unsuitable niche for a species. AUC area values of 0.5 and below are considered to be random or poor models, >0.7 are considered good, and >0.9 are excellent models.

Niche-overlap

Niche overlap was calculated using the nicheOverlap function in the R package *dismo*, based on Hellinger's I and Schoener's D as described in Warren *et al.*, 2008 (Hijmans *et al.*, 2017*b*). Both statistics use a probability distribution, calculated in Maxent, to make a pairwise comparison of niche overlap. Schoener's D measures niche similarity and I statistic measures niche equivalency. Schoener's D is a traditional statistic used in ecological studies to describe niche overlap; however, it differs from the *I* statistic because the suitability scores from Maxent are considered proportional to species abundance (Warren *et al.*, 2008). The *I* similarity statistic only considers the ENM as probability distributions. Both metrics range from 0 (no overlap) to one (total overlap), and had 100 replications (Schoener, 1970). The heatmap was created in Excel to visualize the greatest/least niche overlap among species in the genus.

DNA extraction and sequencing

Whole genome DNA extractions were done with a few modifications of Barry *et al.*, (2005). Modifications include the addition of RNase A before incubation in the water bath, the use of ice-cold ethanol and ice-cold isopropanol. DNA extractions were sent to University of Minnesota Genomics Center, Minneapolis, Minnesota, USA for enzyme optimization, library preparation, and RADseq / SBG Sequencing-based Genotyping. Enzymes used for digestion were PstI-HF and BtgI, and the samples were run on a Nextseq 500 High Output 150-bp SR run. Mean quality scores for all libraries was greater than or equal to Q30, indicating an accurate base call probability of 99.9%, and all barcodes were detected.

Phylogeny estimation

The sample SR raw RADseq fastq files were demultiplexed and assembled de novo using *ipyrad* (Eaton, 2014; Eaton *et al.*, 2017). This included quality filtering, clustering within samples, calculation of joint estimation of error and heterozygosity, consensus base calling, clustering across samples, and finally alignment. This concatenated SNP dataset was exported as tnt, phylip, and nexus files for further analysis.

Phylogeny was estimated using Bayesian in BEAST 2.4.8(Drummond and Rambaut, 2007). The outgroups used in the analyses are in Table 2.1, and all phylogenetic methods were rooted with *Ptelea* (Nixon and Carpenter, 1993). Extractions from herbarium specimens were done for *P. alatus*, *P. demerarae*, and *P. manuensis*; however, RADseq library preparations were not successful so they were not used in this analysis. These species distributions include: *P. alatus* in the Northern Amazonian region, *P. demerarae* in Guyana, and finally *P. manuensis* in Peru and in the western Amazonian Brazilian state of Acre. The distribution of *P. manuensis* is similar to *P. peruvianus,* so it is possible that its inclusion in the phylogeny could assist with a better estimate of the *Pilocarpus* phylogeny.

BEAST 2.4.8 analyses were run using the GTR + Γ +I model, implementing an uncorrelated lognormal relaxed molecular clock as well as a Yule and birth-death process of speciation (Gernhard, 2008). Two separate runs of 30 million MCMC generations (logged every 100) converged and the effective sample sizes (ESS) were assessed positively in Tracer. The independent runs were combined using LogCombiner with a 20% burn in, and then were used to estimate the maximum clade credibility tree in Tree Annotator (Drummond and Rambaut, 2007) (Figure 2.5c).

The BEAST tree was chosen for further phylogenetic comparative analyses since we could calibrate the tree using two *Ptelea* fossils. *Ptelea paliuruoides,* the oldest *Ptelea* fossil, is from the mid-Eocene and was collected in the "Middle Eocene Parachute Creek Member of the Green River Formation of Colorado and Utah"(Manchester and O'Leary, 2010). *Ptelea enervosa,* the second *Ptelea* fossil, is from the mid-Miocene and it was collected from localities on the Oregon-Idaho border (Call and Dilcher, 1995). A mid-Eocene fossil calibration was implemented as a minimum age for the *Ptelea trifoliata* node, (indicated by a STAR on the phylogeny) using a log-normal distribution (Figure 2.5c). The outgroups were trimmed for subsequent phylogenetic comparative analyses using the *ape* package in R, and the branch lengths estimated with the full data set were saved for the analyses.

Ancestral climatic niche reconstruction

Vegetation types were reconstructed across the phylogeny as a categorical variable. This analysis was done using a continuous-time Markov chain model with the 'rerootingmethod' function in *phytools* in R (Revell, 2012).

RESULTS AND DISCUSSION

Bioclimatic Variation in the Genus

Using the R package *raster*, 19 bioclimatic variables from WorldClim were extracted for each herbarium specimen locality (Hijmans *et al.*, 2017*a*) . The climatic preference for each species was determined by calculating the mean of all collection localities per species (APPENDIX 5). The six most variable bioclimatic factors in Table 3.2 include: Precipitation in Driest Quarter (CV%= 66.07), Precipitation in the Coldest Quarter (CV%=60.93), Precipitation in the Warmest Quarter (CV%=46.93), Annual Precipitation (CV%=24.03), Mean Diurnal Range (CV%=12.37), and Annual Mean Temperature

Table 3.2 Bioclimatic preferences of each species in *Pilocarpus*

Species	Annual Mean Temp. (C)	Mean Diurnal Range	Annual Precip. (mm)	Precip. Driest Quarter (mm)	Precip. Warmest Quarter (mm)	Precip. Coldest Quarter (mm)
P. alatus	27.14	10.44	1517.60	40.60	73.00	410.60
P. carajaensis	25.11	11.34	1925.62	85.73	259.23	713.65
P. demerarae	26.87	9.40	1854.67	152.67	157.33	465.00
P. giganteus	22.32	8.56	1683.12	199.19	599.96	211.15
P. grandiflorus	24.04	7.68	1586.45	265.52	408.50	354.90
P. jaborandi	24.07	11.34	1084.67	27.44	165.67	380.06
P. manuensis	25.65	10.83	1844.83	151.50	584.50	191.67
P. microphyllus	26.00	10.05	1663.66	89.12	166.32	573.12
P. pauciflorus	20.91	10.61	1448.19	163.95	541.74	208.71
P. pennatifolius	19.83	11.63	1578.14	283.33	477.31	307.77
P. peruvianus	25.50	12.14	2045.13	88.50	529.50	150.50
P. racemosus	24.35	10.34	1823.10	110.76	332.13	331.70
P. riedelianus	23.63	9.69	1735.46	184.06	442.19	335.34
P. spicatus	22.50	10.13	1059.50	106.41	333.59	146.60
P. sulcatus	22.26	11.41	869.98	21.18	338.52	26.82
P. trachylophus	23.40	12.77	905.44	15.74	254.49	42.22

(CV%=8.61). A few of these bioclimatic factors have some extreme values, which are likely important in niche differentiation of species in this genus. Taxonomic studies of *Pilocarpus* have found certain species to be present in certain vegetation types (Skorupa, 1996). This includes Tropical Rain Forest/Ombrophyllous Forest (areas of high precipitation and average temperatures with very tall trees, and storied canopy), Restinga (coastal occurrence on sandy soils, facing strong salty winds, and small scrubby plants), Seasonally Deciduous Forest (forest with a defined alternation of wet

Table 3.3. Vegetation Types of *Pilocarpus* species

Tropical	Restinga/ Coastal Forest	Seasonally deciduous forest	Caatinga
P. carajaensis	*P. giganteus*	*P. trachyllophus*	*P. trachylophus*
P. alatus	*P. grandiflorus*	*P. sulcatus*	*P. sulcatus*
P. jaborandi	*P. pauciflorus*	*P. spicatus*	*P. spicatus*
P giganteus	*P. spicatus*		
P. grandiflorus			
P. manuensis			
P. peruvianus			
P. microphyllus			
P. pauciflorus			
P. spicatus			
P. pennatifolius			

and dry seasons leading to a large loss of leaves ~90% found on higher grounds) and the Caatinga (very dry savanna in Northeastern Brazil with seasonal rivers) (Table 3.3).

The climatic variability hypothesis suggests that locations with constant warm temperatures and little seasonal variation would lead to a lower capacity or lower resolution for environmental plasticity (Molina-Montenegro and Naya, 2012). Analyzing the entire genus, we graphed the precipitation seasonality and temperature seasonality to find which species had the largest variation in temperature and precipitation seasonality (Figure 3.1). In terms of temperature seasonality, *P. pennatifolius, P. pauciflorus, P. giganteus,* and *P. spicatus* have the greatest variation.

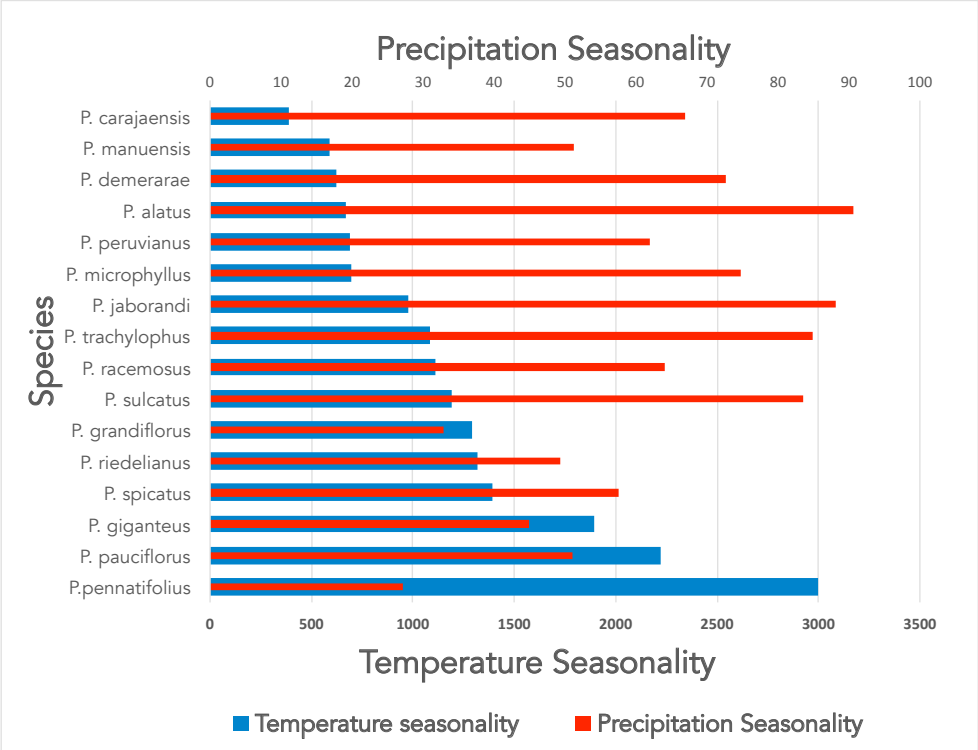


Figure 3.1 Visualization of precipitation and temperature seasonality across *Pilocarpus*. Temperature seasonality in blue has quite a large range of variation across species. Precipitation seasonality in red depicts less variation with more species experiencing high precipitation seasonality

The species with the lowest temperature seasonality are *P. carajaensis, P. manuensis, P. demerareae, P. alatus,* and *P microphyllus*. The temperature seasonality values of these species are further exemplified by their distributions in Figure 3.2, as the species with the lowest temperature seasonality tend to be distributed around the equator and Amazonian regions, whereas species with high temperature seasonality are distributed in southern regions of Brazil and along the Atlantic coast. In terms of precipitation seasonality, the greatest variation is seen in *P. alatus, P. jaborandi, P. trachylophus, P. sulcatus,* and *P. demerarae*. The species with the lowest average precipitation seasonality include *P. pennatifolius, P. grandiflorus,* and *P. giganteus.* The high precipitation seasonality is depicted for species distributions that are in northeastern Brazil and the Amazon, whereas species that are mostly found along the coast and southern inland regions have a lower precipitation seasonality. Overall, it appears that *Pilocarpus* species have either high precipitation seasonality or high temperature seasonality. This analysis can help to provide insight into species that should be further assessed for greater chemical diversity in the genus. *Pilocarpus spicatus,* the species with the lowest known diversity of alkaloids, also appears to be at the lower end of seasonality for both precipitation and temperature (Santos and Moreno, 2004; Sawaya *et al.*, 2011). This is also true when considering the alkaloid diversity of *P. grandiflorus* as well as *P. riedelianus (*Allevato Chapter 2). Interestingly, diversity of alkaloids is greatest for *P. pennatifolius, P. microphyllus,* and *P. carajaensis;* however, this does not specifically correlate with greater temperature or precipitation seasonality. Further work using phylogenetic comparative methods to assess precipitation and temperature across the tree would help to define whether there is phylogenetic signal for bioclimatic factors as well as a correlation with chemical traits.

Ancestral niche reconstruction

Vegetation types, scored as a categorical variable for each species, were reconstructed across the phylogeny using a continuous-time Markov chain model (Figure 3.2) (Revell, 2012). These reconstructions suggest a tropical origin and one major divergence, breaking the genus into two major clades. These two clades are geographically distinct.

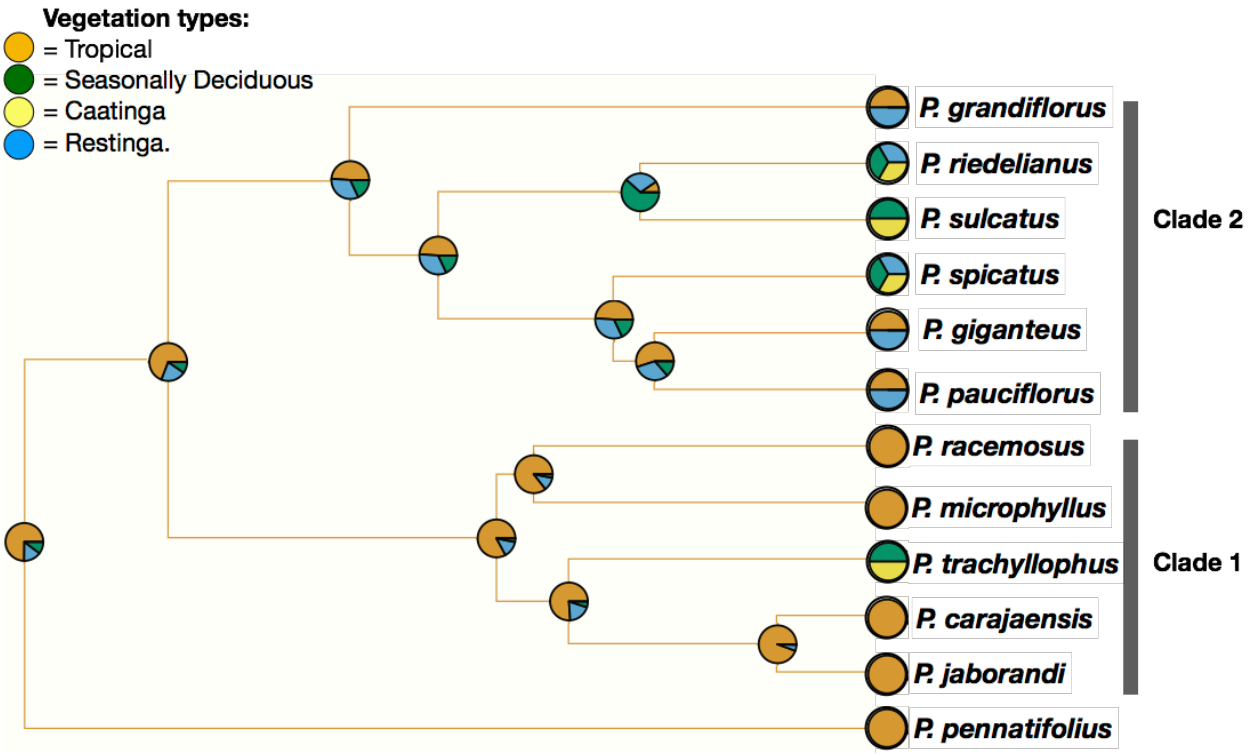


Figure 3.2 Habitat reconstruction using marginal probabilities of alternative ancestral areas through a Mk model. Each color represents a vegetation type: Orange= tropical, green= seasonally deciduous, yellow= caatinga, blue= restinga. Marginal probabilities represented by pie charts at each node.

Clade 1 is mostly distributed in the tropical northern region of Brazil, whereas Clade 2 encompasses all of the species distributed along the mid-coastal regions, the southern regions as well as around to the tropical western regions of Brazil that are north-west of the Pantanal. *Pilocarpus trachyllophus*, when compared to other members

of Clade 1 has a distribution that extends much further south, and it is exposed to much drier conditions comparatively. It is also interesting to note that *P. sulcatus,* the other species found in the caatinga growing sympatrically with *P. trachyllophus*, has the most northern distribution of all other members in Clade 2. Though these two species are distributed sympatrically they are each in different clades and their morphology is extremely different. *Pilocarpus sulcatus* is unique in the genus as it is the only species without pellucid glands. *P. trachyllophus* is the species with the greatest density of trichomes, though trichomes can be found on *P. sulcatus* and one sub-species of *P. spicatus*.

Ecological Niche modeling in *Pilocarpus*

Ecological Niche Modeling (ENM) was performed using Maxent with presence-only data from herbarium specimens of each species. The 19-bioclimatic variables from WorldClim were used; however, as Maxent uses machine learning methods, variables with insignificant contributions to the ENM were identified and not included in the final analysis. The most important contributors chosen for ENM are in Table 3.4.

Table 3.4 Variables used in Ecological Niche Model

Bioclimatic Variables	Units
Mean Diurnal Range	°C
Mean Temp Wettest Quarter	°C
Temp Seasonality	†
Precipitation Seasonality	mm
Isothermality	††
Mean Temperature Driest Quarter	°C
Temp Driest Quarter	°C
Max Temp Warmest Quarter	°C
Precipitation Coldest Quarter	mm

† = Temp Seasonality is the temperature variation over the year, that is calculated using the standard deviation of monthly average temperatures †† = Isothermality is equal to the ratio of mean diurnal range (Bio2) to annual temperature range (Bio 7= Max Temp warmest month- min temp coldest month) multiplied by 100. Value of 100 means there is less variation in a month compared to the year.

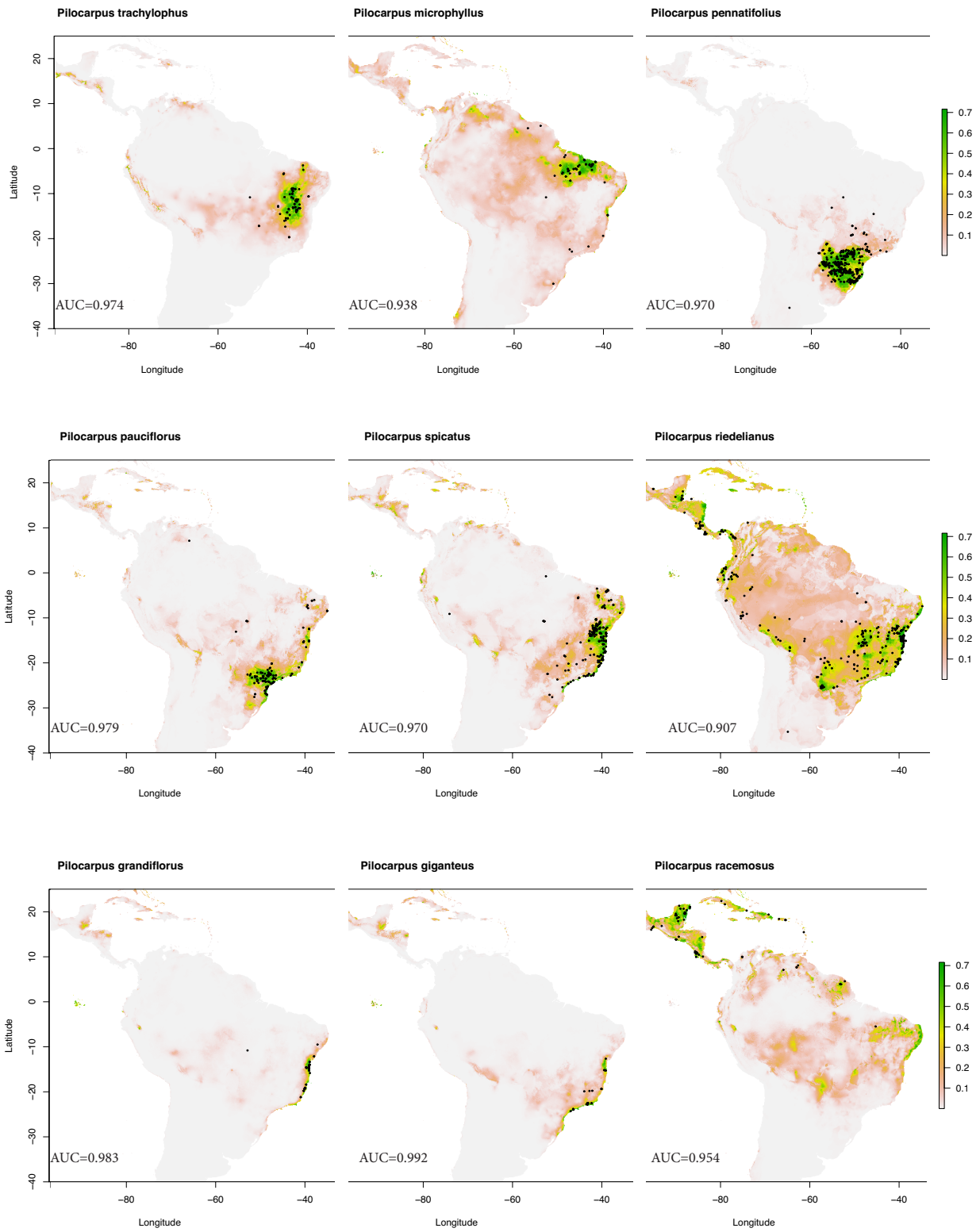


Figure 3.3 Ecological niche models for each *Pilocarpus* species. Suitability is depicted by a gradient: greatest suitability in green to lowest suitability in orange/white. The AUC value is at bottom left, with >0.90 being an excellent model. Species range from widespread to endemic.

To reduce the effect of sample size bias, only nine species were used in these analyses as each had 25 or greater field collections. This allowed the ENM of each species to be more statistically significant; this gain in statistical power was depicted by higher AUC scores. Figure 3.3 depicts the potential distributions for each species in the genus based on the ENM, and the AUC value for each species is indicated for each ENM. There is quite a diversity of niche dimensions across species in *Pilocarpus.* From the ENM's in Figure 3.3 we see that *P. grandiflorus* and *P. giganteus* have extremely restricted distributions along the coast. Although they don't have large areas of potential suitability, there are still areas, which could be important for species discovery or conservation management, as well as potential areas for *Pilocarpus* restoration. It is important to note that although a species can have a distribution in a certain area, it does not necessarily mean that it can be found there. This is due to the many barriers for dispersal including: physical (mountains and rivers) and human intervention (land use modification, human dispersal of pests and exotic plant species). The species with the greatest suitability away from its natural distribution is *P. racemosus. Pilocarpus racemosus* is found across Central America and the Caribbean, reaching areas in Colombia and Venezuela. According to our ENM there appears to be large suitable areas in the Northeastern tip of Brazil in the states of Rio Grande do Norte, Paraiba, and Pernambuco. This is another case of an area that has favorable bioclimatic factors but a clear barrier to other *Pilocarpus* populations. This barrier could be the drier caatinga and dune vegetation present in the states directly to the west/inland (States of Ceara and Piaui). When inspecting all the herbarium specimen collections, there does not appear to be any *Pilocarpus* species present on that specific coastline. Although the area has a moderate level of urbanization, it is more likely that a physical barrier prevented the dispersal of *Pilocarpus* to that region.

Analyzing the AUC values or model accuracy for the ENM of all species (Table 3.5), we discovered that the least significant models were for *P. riedelianus*, *P. racemosus*, and *P. microphyllus*. It is possible that variation in the ENM of *P. microphyllus* stems from the fact that it is the species most used for pilocarpine extraction both currently and historically. Since it has been in cultivation for a couple hundred years, many people have transported and planted it in other locations both for economic and medicinal gains. This human dispersal of *Pilocarpus* can therefore confuse the results of genetic diversity in wild populations. There have been a few population genetic studies that have found identical genotypes in individuals of *P. microphyllus* throughout three different states (Moura *et al.*, 2005; Rocha *et al.*, 2014). This is most likely due to human

Table 3.5 AUC values for each *Pilocarpus* species

AUC	Species
0.992	*P. giganteus*
0.983	*P. grandiflorus*
0.979	*P. pauciflorus*
0.974	*P. trachylophus*
0.970	*P. spicatus*
0.970	*P. pennatifolius*
0.954	*P. racemosus*
0.938	*P. microphyllus*
0.907	*P. riedelianus*

* AUC scores test model accuracy and scores of 0.5 and below are considered to be random or a poor models, >0.7 are considered good, and >0.9 are excellent models.

intervention or cultivation, since the seeds of *Pilocarpus* are mostly dispersed locally as the seeds are ejected from the dehiscent capsule. Analyzing the genetic diversity of herbarium specimens could provide an opportunity to track the potential human spread of *P. microphyllus* through space and time. These herbarium specimens could be combined with current genetic diversity in the wild for a comprehensive demographic

study.

The variation affecting the inaccuracy of the *P. racemosus* ENM may be due to the vast distributions of this species throughout Central America. Another factor that could be affecting this distribution pattern is taxonomic accuracy. *P. racemosus* has three subspecies that are typically recognized, *P. racemosus* subsp *racemosus*, *P. racemosus* subsp *viridulis*, *P. racemosus* subsp *goudotianus* (Skorupa, 1996). The inclusion of population genetic diversity statistics such as Fst and genomic markers could be used in conjunction with niche modeling studies to aid in enhancing taxonomic accuracy and detecting human intervention on species distributions. It is also likely that there are other important microclimate and/or soil dependencies that are not easily modeled but influence each subspecies distribution. Therefore, a more focused and comprehensive field collection record for these specific species could aid in refining these ecological niche models.

The large variation in *P. riedelianus* predicted suitability is possibly due to its distribution along the littoral/coastline in addition to tropical forests, and mountain ranges. The unifying factor here may be ground water, which obviates the annual rainfall values in cases of riparian habitats. In this case, a greater analysis at each site could determine other factors, perhaps even biotic factors, which are important for the distribution. As this project was done based on herbarium specimens and *P.riedelianus* is quite similar to *P.spicatus* and *P. pauciflorus*, it would also be important to look even more closely to verify the herbarium specimen identifications.

Niche Overlap in *Pilocarpus*

Niche overlap, a pairwise comparison of two species ecological niche models was calculated using two metrics: Hellinger-based I statistic (*I*) and Schoener's D (*D*) (Warren *et al.*, 2008). Our analysis of niche overlap found that *P. pauciflorus* (mean

D=0.56, mean *I*=0.64), *P. riedelianus* (mean *D*=0.4, mean *I*=0.65), and *P. spicatus* (mean *D*=0.39, mean *I*=0.67) have the greatest niche overlap with other species in the genus, whereas *P. pennatifolius* has the least similarity or overlap of niche dimensions in the genus (mean *D*=0.19, mean *I*=0.41) (Figure 3.4). The species with the highest niche overlap values are *P. trachyllophus* and *P. sulcatus* (I=0.83, D=0.575), two shrubs that are almost always found growing together in disturbed caatinga areas of northeastern Brazil.

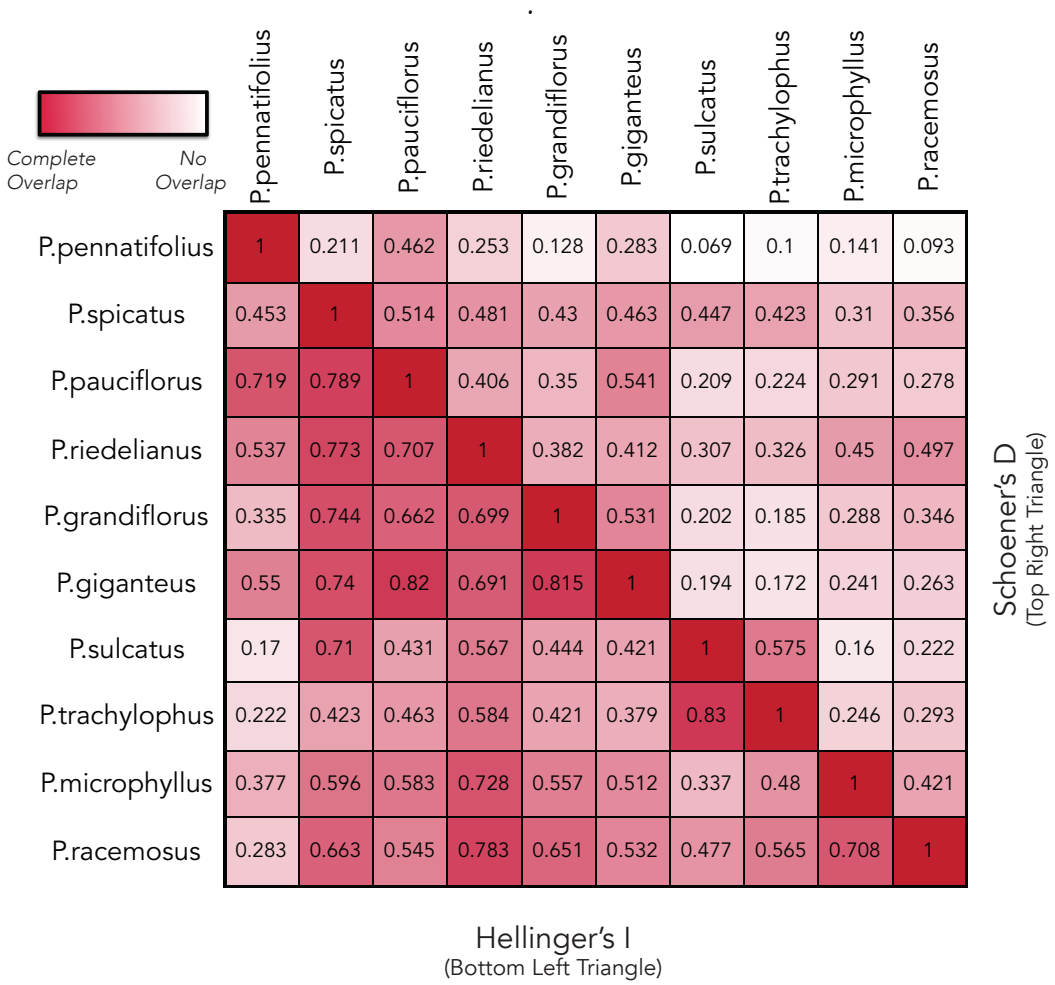


Figure 3.4 Niche overlap heatmap depicting two metrics: Hellinger's I and Schoener's D. The color gradient depicts identical niche overlap between species in dark red to low niche overlap between species in white.

In addition, *P. giganteus*, has some of the highest values for niche overlap with *P. pauciflorus* (I=0.82, D=0.541) and *P. grandiflorus* (I=0.815, D=0.531), which is further verified when comparing their bioclimatic values (Table 3.2) and vegetation types (Table 3.3). When comparing the values from each metric we see that although

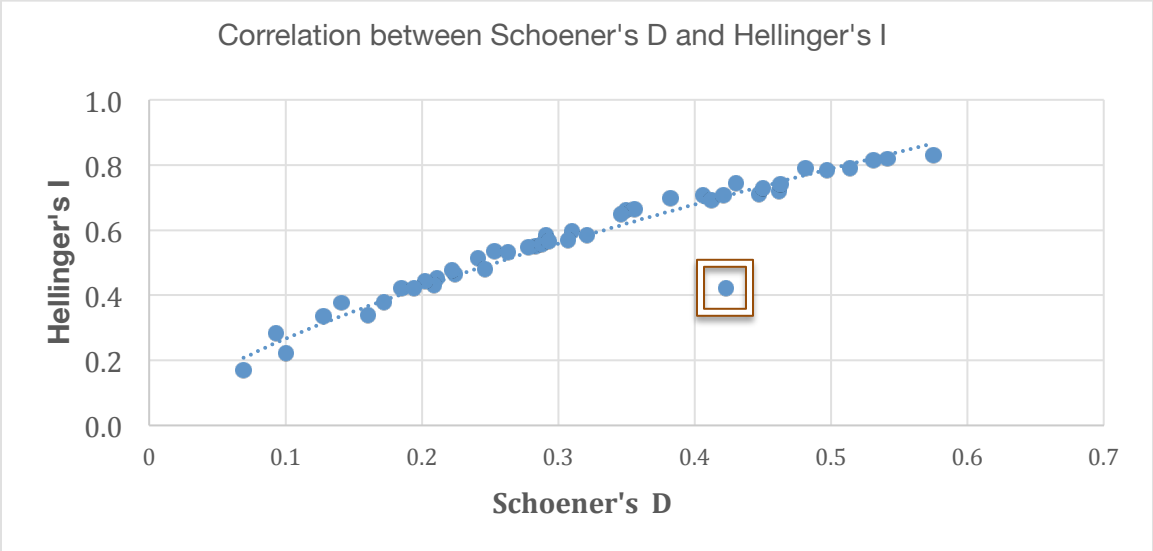


Figure 3.5 Correlation between Schoener's D and Hellinger's I (Warren et al., 2008). Each point represents the correlation of the Schoeners D metric with Hellinger's I metric for every two species comparison shown in the niche overlap heatmap. Outlier is identified by red square

Schoener's D is usually a higher number, Schoener's D and Hellinger's I are correlated. We found one outlier in the pairwise comparison, this was the niche overlap of *P. spicatus* with *P. trachyllophus* (designated by a red square in Figure 3.5). Overall the correlation of both metrics further supports our niche overlap analyses.

This analysis has confirmed field observations of the genus which have found partial sympatric distributions for: *P. sulcatus* / *P. trachylophus; P. riedelianus* / *P .spicatus; P. riedelianus* / *P .giganteus; P. pauciflorus* / *P. spicatus* (Skorupa, 1996). While in the field, I also witnessed an overlap of *P. riedelianus* / *P. grandiflorus* and *P. pennatifolius* / *P. pauciflorus,* these species combinations also have greater niche overlap values as depicted in Figure 3.4.

Conclusions

As is true with many model-based methods, particular starting assumptions can affect the analysis. Usage of occurrence records from herbarium specimens is a phenomenal resource; though, care must be used to prevent biases. These biases can include sample size of collections, sampling locations that are more easily accessible, variance in the accuracy and error of geo-referencing, and taxonomic accuracy. Our analyses attempted to reduce some of these errors, including incorrect geo-referenced samples and a reduction of duplicate samples. Some of these issues can be further ameliorated through the addition of new field collections and simulation analyses using randomly sampled herbarium samples. There can also be concerns with the accuracy of the modeled (usually interpolated) environmental variables at each location. This includes identification of relevant scales of bioclimatic factors, variation in space and time of field collections, and finally types of variables used such as categorical versus continuous. We decided that Maxent would be an appropriate model for our research question because herbarium specimens only record positive occurrences. In addition, the use of presence only data in Maxent helps to reduce errors of models using presence/absence data, and it has also been shown to be effective for smaller quantities of presence only data (Elith and Leathwick, 2009).

Ecological niche modeling of *Pilocarpus* has allowed us to visualize the distribution of species in this genus and estimate the potential (predicted ideal) niche of species, including analysis of niche overlap among species in the genus. It can be seen both by collection records and by the ENM's that there is a high likelihood of overlap in species distributions. This study has identified areas of greater species diversity, environmental variables contributing to the distribution of this genus, as well as species

exposed to a wider variation in bioclimatic factors. These results can be used to guide future field collections of *Pilocarpus* in Brazil to further assess chemical variation in a phylogenetic context. Phylogenetic comparative analyses can be used to test the Climate variability hypothesis, assessing whether areas with greater seasonality have a greater impact than phylogenetic relationships in regard to chemical diversity or other desired phenotypic traits. In addition, we have identified species of interest for further study using population genomics and environmental metabolomics techniques. Niche modeling techniques with herbarium specimens are integral to understanding suitability for species distributions, and are great resources for conservation management.

ACKNOWLEDGMENTS

We thank various individuals in assisting with this study: Dr. Paulo Mazzaferra, Dr. Eduardo Kiyota, Dr. Marcelo Caxambu, Dr. Pedro Dias, Dr. Eloy Rodriguez, Dr. Robert Raguso. We would also like to thank our funding from NSF GRFP, BSA, ASPT, SSB, GOA, Moore Funds and Endowment, and the Cornell Graduate School.

LITERATURE CITED

Braak CJFT. 1986. Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology **67**, 1167–1179.

Buckley LB, Davies TJ, Ackerly DD, *et al.* 2010. Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. Proceedings of the Royal Society B: Biological Sciences **277**, 2131–2138.

Call VB, Dilcher DL. 1995. Fossil Ptelea Samaras (Rutaceae) in North America. American Journal of Botany **82**, 1069–1073.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**, 214.

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

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. Systematic Biology **66**, 399–412.

Elith J, Leathwick JR. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics **40**, 677–697.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**, 43–57.

Etten J van, Hijmans RJ. 2010. A Geospatial Modelling Approach Integrating Archaeobotany and Genetics to Trace the Origin and Dispersal of Domesticated Plants. PLOS ONE **5**, e12060.

Franklin J. 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge Core.

Franklin J. 2013. Species distribution models in conservation biogeography: developments and challenges. Diversity and Distributions **19**, 1217–1223.

GBIF Occurrence Download. 2016. GBIF.org https://doi.org/10.15468/dl.ywhpmz.

Gernhard T. 2008. The conditioned reconstructed process. Journal of Theoretical Biology **253**, 769–778.

Godsoe W. 2010. Regional Variation Exaggerates Ecological Divergence in Niche Models. Systematic Biology **59**, 298–306.

Hijmans RJ, Etten J van, Cheng J, *et al.* 2017*a*. *raster: Geographic Data Analysis and Modeling*.

Hijmans RJ, Phillips S, Elith JL and J. 2017*b*. *dismo: Species Distribution Modeling*.

Hutchinson GE. 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**, 415–427.

IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis. 1992. *Portaria No. 06N*. Brasilia.

INCT. 2018. INCT- Herbario Virtual da Flora e dos Fungos http://inct.splink.org.br/.

Jamil T, Kruk C, Braak CJF ter. 2014. A Unimodal Species Response Model Relating Traits to Environment with Application to Phytoplankton Communities. PLOS ONE **9**, e97583.

Manchester SR, O'Leary EL. 2010. Phylogenetic Distribution and Identification of Fin-winged Fruits. The Botanical Review **76**, 1–82.

Molina-Montenegro MA, Naya DE. 2012. Latitudinal Patterns in Phenotypic Plasticity and Fitness-Related Traits: Assessing the Climatic Variability Hypothesis (CVH) with an Invasive Plant Species (F Seebacher, Ed.). PLoS ONE **7**, e47620.

Moura EF, Pinto JEBP, Santos JB dos, Lameira OA, Nunes JAR, Ferreira DF. 2005. Genetic diversity of Pilocarpus microphyllus accessions using leaf traits. Crop Breeding and Applied Biotechnology **5**, 348–354.

Nixon KC, Carpenter JM. 1993. On Outgroups. Cladistics **9**, 413–426.

Peterson AT. 2011. Ecological niche conservatism: a time-structured review of evidence. Journal of Biogeography **38**, 817–827.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. *Ecological Niches and Geographic Distributions*. Princeton, N.J: Princeton University Press.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**, 231–259.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography **31**, 161–175.

Porfirio LL, Harris RMB, Lefroy EC, Hugh S, Gould SF, Lee G, Bindoff NL, Mackey B. 2014. Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. PLOS ONE **9**, e113749.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution **3**, 217–223.

Rocha J, Silva F, Melo A, Andrade I. 2014. ISSR Primer Selection for Genetic Variability Analyses with Jaborandi (Pilocarpus microphyllus Stapf ex Wardlew., Rutaceae). Forest Research Open Access **03**.

Santos AP, Moreno PRH. 2004. Pilocarpus spp.: a survey of its chemical constituents and biological activities. Revista Brasileira de Ciências Farmacêuticas **40**, 116–137.

Sawaya ACHF, Vaz BG, Eberlin MN, Mazzafera P. 2011. Screening species of Pilocarpus (Rutaceae) as sources of pilocarpine and other imidazole alkaloids. Genetic Resources and Crop Evolution **58**, 471–480.

Schoener TW. 1970. Nonsynchronous Spatial Overlap of Lizards in Patchy Habitats. Ecology **51**, 408–418.

Skorupa LA. 1996. Revisao Taxonomica de Pilocarpus Vahl (Rutaceae). Ph.D. Thesis.

Skorupa LA. 1998. Three New Species of Pilocarpus Vahl (Rutaceae) from Brazil. Novon **8**, 447–454.

Skorupa LA. 1999. A New Species of Pilocarpus Vahl (Rutaceae) from Peru. Novon **9**, 257–259.

Warren DL, Glor RE, Turelli M. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution **62**, 2868–2883.

CHAPTER 4

Ecometabolomic analysis of wild populations of *Pilocarpus pennatifolius* using unimodal analyses

Daniella Allevato[1], Eduardo Kiyota[2], Paulo Mazzafera[2], Kevin Nixon[1]

ABSTRACT

- *Background and Aims:* Studies examining the diversity of plant specialized metabolites suggest that biotic and abiotic pressures greatly influence the qualitative and quantitative diversity found in a species. Large geographic distributions expose a species to a great variety of environmental pressures, thus providing an enormous opportunity for expression of environmental plasticity. *Pilocarpus*, a neotropical genus of Rutaceae, is rich in alkaloids, terpenoids, and coumarins, and is the only commercial source of the alkaloid pilocarpine for the treatment of glaucoma. Overharvesting of species in this genus for pilocarpine, has threatened natural populations of the species. The aim of this research was to understand how adaptation to environmental variation shapes the metabolome in multiple populations of the widespread species *Pilocarpus pennatifolius.*

- *Methods:* LCMS data from alkaloid and phenolic extracts of leaf tissue were analyzed with environmental predictors using unimodal unconstrained and constrained ordination methods for an untargeted metabolomics analysis. PLS-DA was used to further confirm the chemoecotypes of each site.

- *Key Results:* The most important variables contributing to the alkaloid variation between the sites: mean temperature of wettest quarter, as well as the soil content of

1. Section of Plant Biology, School of Plant Sciences, Cornell University, Ithaca, NY, USA
2. Departamento de Biologia Vegetal, Instituto de Biologia, UNICAMP, Universidade Estadual de Campinas, Brasil

phosphorus, magnesium, and base saturation (V%). The most important contributing to the phenolic variation between the sites: mean temperature of the wettest quarter, temperature seasonality, calcium and soil electrical conductivity (EC).

- *Conclusions:* This research will have broad implications in a variety of areas including biocontrol for pests, environmental and ecological plant physiology, and strategies for species conservation maximizing phytochemical diversity.

KEYWORDS: Ecological metabolomics, unimodal ordination, *Pilocarpus pennatifolius*, chemoecotype, ecophysiology, canonical correlation analyses

INTRODUCTION

The vast array of compounds synthesized in plants has led to much debate over the function or adaptive significance for such a diversity of compounds. When plant specialized metabolites (PSM) were first discussed, they were often seen as waste products from primary metabolism, accidents or aberrant metabolism (Haslam, 1986). Today, PSMs are believed to have multiple functions and have been shown to be as important as primary metabolites in many cases (Cipollini, 2000; Izhaki, 2002; Wink, 1999).

The qualitative and quantitative PSM diversity in plants appears to change over time (phenology) and space, encompassing many ecological roles (Moore, Andrew, Külheim, & Foley, 2014). Studies examining the diversity of plant specialized metabolites suggest that in the environment, biotic and abiotic pressures greatly influence the qualitative and quantitative diversity of the metabolome found in a species. The Climatic Variability hypothesis suggests that locations with constant warm temperatures and little seasonal variation would lead to a lower capacity for environmental plasticity (Molina-Montenegro & Naya, 2012). Large geographic distributions expose a species to a great variety of environmental pressures, thus providing an opportunity for an enormous range of chemoecotypes (Gratani, 2014; Sultan, 1987; Tack, Johnson, & Roslin, 2012).

Pilocarpus, a genus of Rutaceae in the sub-tribe Pilocarpinae, is known for its high bioactivity due to the presence of many alkaloids, terpenoids and coumarins (Santos & Moreno, 2004). Specifically, the genus has been sought after for one specific imidazole alkaloid, pilocarpine, as it is the only source of this compound that is used to treat glaucoma, as well as xerostomia. The genus *Pilocarpus* is the only known genus that contains this compound, and many of the species in this genus have become threatened

with extinction due to overharvesting of wild populations (Pinheiro, 1997, 2002). Besides two chemical variants found in the genus *Casimoroa (Rutaceae),* Imidazole alkaloids have not been found in any other genus of plants. The genus *Pilocarpus* comprises 16 species and has a widespread, but strictly Neotropical distribution from southern Mexico and the Antilles, to northern Argentina (Skorupa, 1996).

Previous studies on other species of *Pilocarpus* have found that abiotic factors such as environmental variation can affect compound expression (Abreu, Mazzafera, Eberlin, Zullo, & Sawaya, 2007; Sawaya, Vaz, Eberlin, & Mazzafera, 2011). One study examining imidazole alkaloid concentrations in leaf tissue in *Pilocarpus microphyllus* plantations in Southern Brazil has found that there is variation of compound concentrations and compound presence depending on the season. The exposure to seasonality in the South is very different from the uniform seasons where *P. microphyllus* is naturally found, in the tropical rainforest and terra firme regions of the Amazonian state of Para. The seasonal variation of alkaloid production in *P. microphyllus* has demonstrated a strong association between specific groups of compounds and seasons, and has led to the hypothesis that distinct biosynthetic pathways are active in different seasons (Abreu et al., 2007; Sawaya et al., 2011). In addition, studies looking at salt stress, wounding, and hypoxia in *Pilocarpus* reported significant reductions in certain compounds following exposure to these conditions, further supporting the plasticity of certain genotypes to environmental stresses (Avancini, Abreu, Saldaña, Mohamed, & Mazzafera, 2003). Further studies using *P. microphyllus* calluses and seedlings showed variation in pilocarpine production with pH variation, absence/excess of nutrients N, K, P and NaCl, and amino acid precursors histidine and threonine (Abreu, Sawaya, Eberlin, & Mazzafera, 2005; Andreazza, Abreu, Sawaya, Eberlin, & Mazzafera, 2009; Avancini et al., 2003).

One species in the genus, *Pilocarpus pennatifolius,* has a large latitudinal distribution as well as some of the greatest diversity of compounds present in the genus (Santos & Moreno, 2004; Sawaya et al., 2011). As *P. pennatifolius* has one of the largest distributions, one would expect that there would be a greater variation in chemistry since it is exposed to a larger variety of environmental factors. Using *P. pennatifolius* as an example of a species with a widespread latitudinal distribution, we were interested in determining which environmental factors are associated with the greatest variation among metabolomes of *P. pennatifolius* populations. We hypothesize that there are patterns of chemistry, or chemoecotypes, correlated with varying environmental conditions among wild populations of *P. pennatifolius*.

Previous plant comparative metabolomics studies have utilized linear methods to investigate correlations between PSM and the environment. (Kaplan *et al.*, 2004; Arbona *et al.*, 2013; Lankadurai *et al.*, 2013; Sampaio *et al.*, 2016). However, with the larger environmental gradient present in the wild populations of this study, the response of chemical variation in these situations follows a nonlinear and unimodal distribution (Whittaker, 1956). In situations of smaller environmental gradients, a linear response could approximate the conditions well; however, when there is a unimodal response we expect that the linear model will not suffice (Šmilauer & Lepš, 2014). PSM production in plants can display a linear or unimodal response when faced with different abiotic factors. A unimodal response can display maximum compound abundance at an optima with low compound production at the extremes (e.g. no nutrient/factor vs excessive nutrient/factor), or there can be maximum compound abundance at the extremes when a plant faces extreme environmental stressors (Figure 4.1) (Yan *et al.*, 2004; Pennycooke *et al.*, 2005; Albert *et al.*, 2009; Ncube *et al.*, 2012).

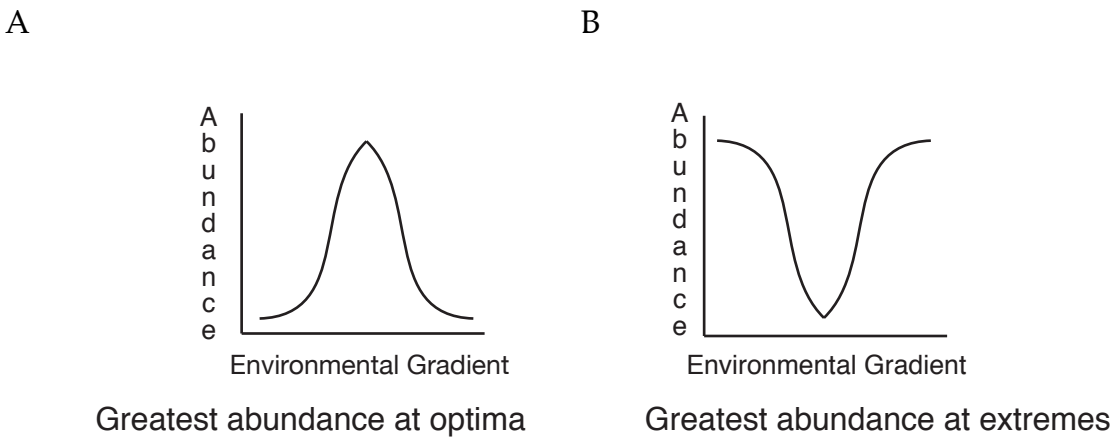


Figure 4.1. Compound abundance can follow a unimodal response with regard to environmental factors. Greatest abundance at an optima (a), or greatest abundance at environmental extremes when the plant is under stress (b).

Although unimodal methods can approximate linear responses well, the reverse is not true with linear methods. Therefore, in the case of our study with a large environmental gradient and a unimodal response of traits, only the unimodal method can accurately assess the response. Multivariate methods can address the issues related to unimodal distributions and have been used extensively in microbial ecology and community ecology, but they have not previously been applied to plant environmental metabolomics (Braak, 1986; Harbourne, 1993; Paliy & Shankar, 2016; Sardans, Penuelas, & Rivas-Ubach, 2011).

The objective of our study was to test the hypothesis that there are correlations between chemotypes and environmental conditions for wild populations of *P. pennatifolius* using multivariate methods to analyze leaf metabolomics data. This included an unconstrained ordination analysis to assess overall variation in the metabolome among sites, a constrained ordination to assess metabolomic variation attributed to environmental predictors, and discriminant analysis to assess the group classifications for chemoecotypes of the various wild populations. To avoid inaccurate estimates from linear analysis, in this study we have used the following unimodal

analyses, Correspondence Analysis (CA) (Hill, 1973, 1974) and Canonical Correspondence Analysis (CCA), to accurately assess variation in the metabolomes of wild populations of *P. pennatifolius*.

MATERIALS AND METHODS

Field collections of plant material and plant sampling design

To examine variation in alkaloid and phenolic profiles due to environmental variation, this study focused on *P. pennatifolius*. This species was chosen because it has the largest latitudinal range (greater variation of climatic variables) that is both current and historic, and it is not threatened with extinction in natural populations. Using bioclimatic data from WorldClim website (www.worldclim.org), distribution data from 2189 Brazilian herbarium specimens from SpeciesLink website (http://splink.cria.org.br), and the Ecocrop function in DIVA-GIS, localities with the greatest disparity of climatic variables were chosen for collection sites while planning our fieldwork itinerary (Dell et al., 2014).

Field collected plants were sampled from six locations for *P. pennatifolius*, obtaining as many individuals as possible from each site (Table 4.1). Specimens were collected from Parque Estadual Lago Azul, Campo Mourão Site 1, Campo Mourão Site 2, Foz do Iguaçu, Estação Ecológica St. Tereza, and Cruz do Pedro. Herbarium specimens were made for each site and can be accessed at the Bailey Hortorium Herbarium (BH) and USP Ribeirão Preto Herbarium (SPFR) (APPENDIX 6). Multiple leaves were collected from each individual. Specifically, the third leaflet down from the terminal leaflet of the compound leaves was chosen; each sample collection was obtained from a different compound leaf each time.

Soil sample field collections and analysis

At each collection locality, two soil samples were collected at 20cm depth (Camargo, Moniz, Jorge, & Valadares, 2009; Raij, 2001) and soil analysis was carried out at the Instituto Agronômico in Campinas, SP, Brazil following established methods (Raij, 2001). Fe, Mn, Zn, P. K, Ca, Mg, S, B, Na, Al, N, organic matter, pH, Electric Soil Conductivity (EC), Base saturation (V%), Cation Exchange Capacity (CEC) were determined in the soil samples (APPENDIX 7).

Table 4.1. Summary of *P. pennatifolius* population sites and final set of environmental variables

Sites in Brazil	*	Geographic coordinates	#	TS	MT WQ	MT CM	P (mg/ dm3)	Mg (mmolc /dm3)	Ca (mmolc /dm3)	V% (%)	EC (dS/ m)	H+Al (mmolc/ dm3)
A. Foz do Iguaçu, PR		-25′26′48.98″ -54′35′6.45″	3	359.78	23.73	23.73	37	27	201	94	1.3	16
B.Campo Mourao A, PR		-24′2′01.4″ -52′20′50.7″	5	291.01	23.5	16.5	100.5	6.5	83.5	39.5	1.1	139
C.Campo Mourao B, PR		-24′2′00.7″ -52′20′50.7″	4	291.01	23.5	16.5	66	38.5	362	92.5	0.9	30
D. Lago Azul, PR		-24′06′05.8″ -52′18′42.8″	9	290.26	23.35	16.38	57	35	328	93	1.5	29
E. Cruz do Pedro, SP		-21′17′42.27 -47′53′56.46″	5	204.70	23.13	18.33	32	18	73	76	0.7	29
F.Estacao Ecologica StTereza, SP		-21′13′12″ -47′50′54″	5	203.01	23.23	18.46	186	36	246	96	2.2	11

P, Mg, Ca, V% (base saturation), EC (soil electrical conductivity), and H+Al were determined by soil analysis at each site. Other bioclimatic and soil variables measured are in Supplementary Table 3.
TS= Temp Seasonality; MTWQ= Mean Temp Wettest Qtr; MTCM= Mean Temp Coldest Month
*Colors refer to color-coding for Figures 4.2 and 4.3 only.

Bioclimatic data retrieval

Bioclimatic variables (BIO1-BIO19) at 30-arcseconds were extracted from the WorldClim website (www.worldclim.org) using the longitude and latitude coordinates of each collection locality. In addition, altitude data was extracted using coordinates for each collection on the program DIVA-GIS (APPENDIX 8).

Biochemical extraction of leaf tissues for LCMS

To validate biochemical extractions of alkaloids and coumarins from field collected leaf tissue, a degradation study on two live specimens of *P. pennatifolius* from the New York Botanic Garden was conducted. Silica-dried leaf tissue on ice showed the least degradation and was not significantly different from the immediately frozen tissue. This method of preservation of leaf tissue for biochemical extractions has been used in many studies that require plant collections in remote areas, as the removal of water reduces enzymatic activity in the leaf tissue and the cold temperature helps to prevent enzymatic reactions due to handling (Fine et al., 2013).

Biochemical extractions to elucidate alkaloids and phenolics present in *Pilocarpus* leaf tissue were performed on silica-dried leaf tissue, at UNICAMP in Brazil. Alkaloid extraction protocols followed modifications made by Dr. Sawaya (Dr.Mazzafera Lab) on the method developed by Avancini et al 2003. Alkaloid extracts were done on approximately 10mg leaf tissue (normalized by exact weight in LCMS data prep). Leaf tissue was extracted with 1ml of 0.1% Formic acid and placed in an ultrasonic water bath for 20min. Supernatant was removed and 1ml of 0.1% Formic acid step was redone twice, filtered, and analyzed. The phenolic extraction protocol followed the methods developed by (Durand-Hulak et al., 2015; Vialart et al., 2012)Vialart 2012 and Durand-Hulak 2015. These phenolic extracts were done on approximately 200mg tissue (normalized by exact weight in LCMS data prep). Tissue was homogenized in 2ml 80%

ethanol for 1 min, centrifuged for 10min, and the supernatant removed to a fresh tube. Each individual plant had a total of 3 biological replicates for each extraction.

UPLC-ESI-MS methods

Alkaloids

Five microliters of each sample were injected into an Acquity UPLC coupled with a TQD triple-quadrupole mass spectrometer (Micromass-Waters, Manchester, England). Mass spectrometer conditions were: capillary 3.0 kV, cone 30 V, extractor 1V, ion source temperature 150° C, desolvation temperature 300° C and column temperature 30° C, in electrospray ionization in positive mode. The chromatographic column used was a Polaris 3 C18-A 100 x 2.0mm column (Varian) and the elution was carried out using an ammonium acetate buffer, 10mmol / L, pH 3.0 (solvent A) and acetonitrile (solvent B) in a gradient ranging from 5 to 25% of solvent B in 8 min. The retention time and m/z were used to identify alkaloid compounds in the samples.

Phenolics

Sample extracts were diluted in 80% ethanol (ethanol and water) 1:4. Four microliters of each sample were injected into an Acquity UPLC coupled with a TQD triple-quadrupole mass spectrometer (Micromass-Waters, Manchester, England). Mass spectrometer conditions were: capillary 3.0 kV, cone 35 V, extractor 1V, ion source temperature 120° C, desolvation temperature 350° C and column temperature 30° C, in electrospray ionization in positive mode. The chromatographic column used was a Waters ACQUITY C18-BEH (2.1 × 100 mm,1.7 μm) column and the elution was carried out using Milli-Q water with 0.1% formic acid (solvent A) and acetonitrile (solvent B) in a gradient ranging from 10 to 100% of solvent B in 8 min. The retention time and m/z were used to identify coumarin compounds in the samples using a SIM (single ion

monitoring) mode. For the metabolomics analysis samples were run in TIC (total ion chromatogram) mode for all phenolics.

LCMS data processing

Data processing was conducted using LCMS raw data in MZmine2 (Pluskal, Castillo, Villar-Briones, & Orešič, 2010) and XCMS in R (Smith, Want, O'Maille, Abagyan, & Siuzdak, 2006). Waters raw LCMS data were converted into .mzXML format using the MSConvert package in Proteowizard v.3.0.10051 (Chambers et al., 2012). The raw data in .mzXML format was then processed in MZmine2 following protocols described by Earll 2012: mass detection (*m*/*z* tolerance of 0.002) and filtering, chromatogram builder, peak deconvolution using Wavelets XCMS algorithm, peak alignment of all the LCMS runs, adduct search, and gap filling (Earll, 2012). The raw data was also analyzed using XCMS in R, this included peak picking, integration, and non-linear retention time correction of peaks between samples. LCMS data of all the metabolite features were exported, including EIC (extracted ion chromatograms) for manual analysis and a csv file of the processed samples containing: peak *m*/*z*, peak RT, and peak intensity areas for multivariate analyses. Some compounds were identified via standard references as well as comparison with literature (Hiserodt & Chen, 2012; Sawaya et al., 2011; Tine, Renucci, Costa, Wélé, & Paolini, 2017) (Supplemental Table 1a, 1b). All metabolite features used in the analysis are in Supplemental Data A (alkaloid extract) and Supplemental Data B (phenolic extract). LCMS data was normalized by dry weight of leaf tissue, log transformed, and pareto scaled to analyze the metabolome of individuals from these six sites.

LCMS multivariate analysis

Unsupervised multivariate methods, CA and HCA were employed in CANOCO 5. Correspondence analysis (CA) was used to reduce the dimensionality of the LCMS

data. Hierarchical cluster analysis (HCA) was utilized to create a dendrogram, utilizing Ward's method with Euclidean distances to confirm affinities from CA.

Constrained ordination is not possible with 39 environmental predictors, as the result will be unconstrained; therefore, the 39 environmental predictors were reduced to eleven environmental predictors, to reduce multicollinearity and the arch effect due to many environmental variables (Draper and Smith, 1998, Šmilauer & Lepš, 2014). Factor analysis of the environmental variables revealed five components for the 39 environmental predictors. These complex variables were made up of many environmental predictors and the variables did not load clearly; consequently, the interpretability would be lost if these complex variables were used in the constrained ordination (Draper and Smith, 1998, Šmilauer & Lepš, 2014). Evaluation of the sites on Canoco software revealed environmental variables with low contributions / arrows, which were removed first from the data set. Forward stepwise analysis was performed on Canoco software using a semi-automated procedure, determining the eleven most informative variables that explained the residual variation in chemical traits. This stepwise analysis also analyzed correlated variables, assessing the difference in eigenvalues of species-environmental correlations to determine if wrong variables or too many variables were removed from correlated sets. The final set of included variables is presented in Table 4.1; however, for both analyses only 4 variables had significant contributions to the chemical variation.

Canonical Correspondence Analysis (CCA) of environmental predictors was used to assess variation of metabolic profiles, associated with environmental variables. The three environmental variables with greatest significant variation affecting profiles were depicted on CCA.

Supervised multivariate analyses such as Partial Least Squares Discriminant Analysis (PLS-DA) using the R pls package through MetaboAnalyst, and Random Forest analysis using the randomForest package in R through MetaboAnalyst, were used to confirm classification of chemotype groups (Xia & Wishart, 2002). PLS-DA is excellent at managing both noisy data and multicollinearity, and it was used to identify significant variables in chemotypes through the determination of the regression coefficient, loading plots and variable importance on projection (VIP). Random Forest Analysis was also used for class prediction and provides error rates for each class as well as outlier measures, OOB error, and variable importance measures.

Validation of Multivariate Analyses

The significance of the class discriminations described by PLS-DA models were statistically validated using performance accuracy measures R2 (sum of squares captured by the model) and Q2 (cross-validated R2), cross-validation with different numbers of components and a permutation test using the optimal number of components from the cross-validation.

RESULTS

***Pilocarpus pennatifolius* sample sites and environmental variables**

The choice of environmental variables to use in the model can affect the analysis; therefore, forward selection of environmental predictors in Canoco was used to reduce the environmental variables for the analysis (Table 4.1). Leaf tissue underwent two separate extraction protocols, an alkaloid and a phenolic extraction, and was normalized by dry weight of leaf tissue and log transformed, to analyze the metabolome of individuals from these six sites. All identified and unidentified peaks were used for the metabolomic analyses (Supplemental Data A and B).

Correspondence Analysis to assess overall variation among sites

The relationship between the compound abundance across samples, versus the measured values of the environmental variables was evaluated. As many of our environmental variables encompass a broad range, we analyzed the distribution in SPSS and found that very few variables followed a linear response; therefore, it was preferable to use unimodal analysis methods for our analysis (Šmilauer & Lepš, 2014). For our indirect gradient analysis we used the unimodal method of correspondence analysis.

Correspondence analysis (CA), also known as reciprocal averaging, is based on the weighted averaging of compound scores to estimate the latent variables that best predict values for all compounds in a model (Lepš & Šmilauer, 2003). CA of metabolomics data from the populations of the six sites depicts about four major groups of variation in the alkaloid extract (Figure 4.2a) and about three major groups of variation in the phenolic extract (Figure 4.2c). The explained variation of each axis of the CA ordination is shown in Table 4.2. The first few axes of CA are usually attributed to environmental variables; however, as this is a secondary comparison to environmental gradients, it is an indirect gradient analysis and is viewed as the overall chemical variation between samples. When each compound is graphed on the CA diagram we can see how each grouping of samples is differentiated by a greater abundance of select compounds in the chemical profile (Supplemental Figure 4.1a, 4.1b).

Canonical Correspondence Analysis to assess variation due to environmental factors

To directly test the environmental gradient, Canonical Correspondence Analysis (CCA), a multivariate extension of CA was applied. Here the axes are restricted or

constrained to be linear combinations of environmental variables, and therefore the axes represent those variables that present the maximum separation or variation in the

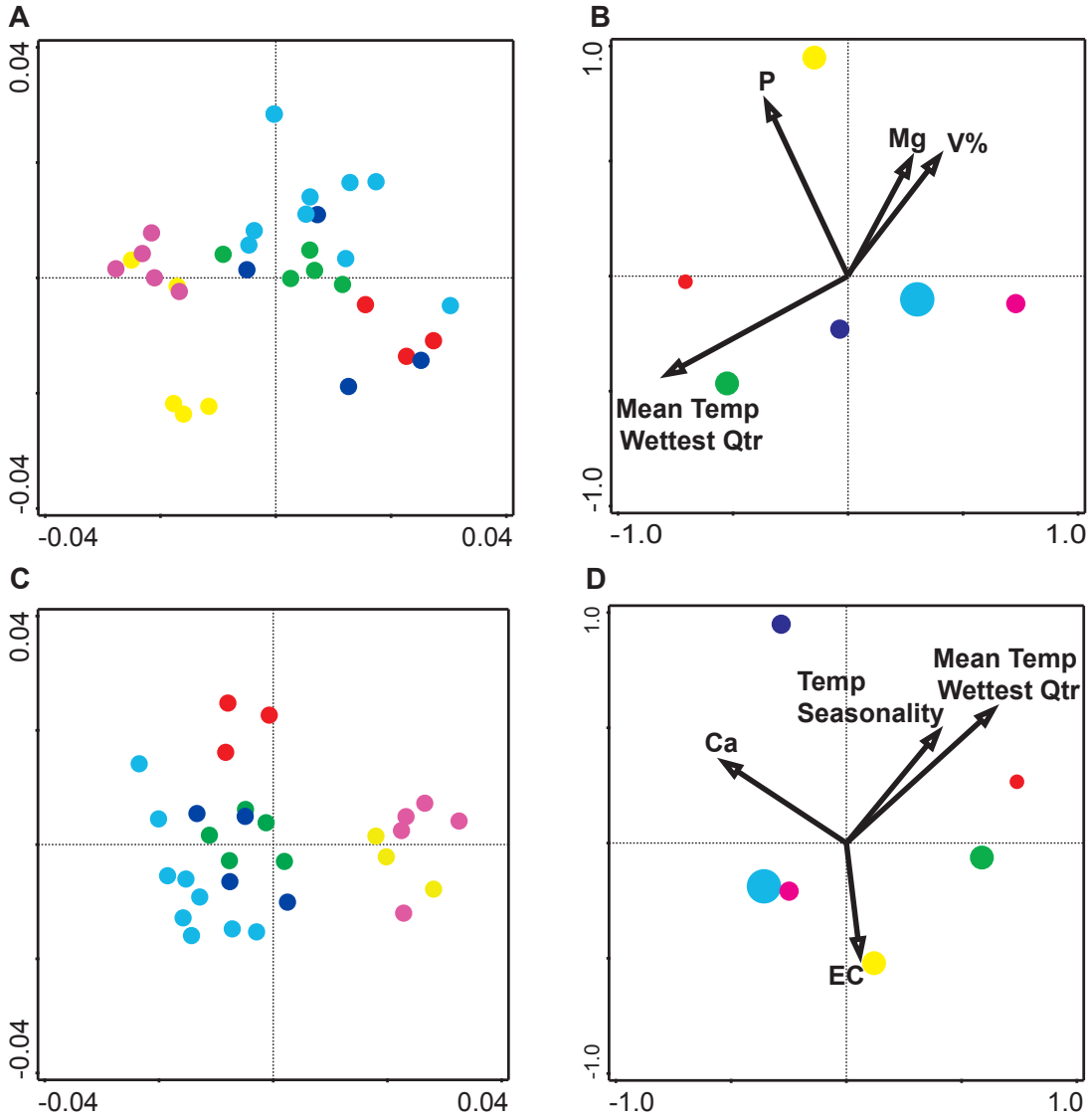


Figure 4.2. CA and CCA for the Alkaloid and Phenolic Extractions. (a) CA ordination plot of alkaloid extraction at the 6 sites with each dot representing an individual plant (average of 3 biological replicates) at a site (b) CCA ordination biplot of alkaloid extractions displaying environmental variables (arrows) and each site (circle with size representing the number of individuals collected) (c) CA ordination plot of phenolic extraction at the 6 sites with each dot representing an individual plant (average of 3 biological replicates) at a site (d) CCA ordination biplot of phenolic extraction displaying environmental variables (arrows) and each site (circle with size representing the number of individuals collected). Number of individuals collected and site location (A-F) is also noted in Table 4.1. Numerical results of CA and CCA are in Table 4.2. Site A is Red, Site B is Green, Site C is Dark Blue, Site D is Light Blue, Site E is Magenta, Site F is Yellow.

samples. The total environmental variables collected for each site consisted of 19 bioclimatic variables from the WorldClim website (www.worldclim.org) and 20 soil variables analyzed from soil samples at each site. Using all 39 environmental variables for the CCA would create an ordination diagram that appears more comparable to an unconstrained CA than to a constrained CCA. Therefore, after removing collinear variables we performed a forward selection analysis of explanatory variables in CANOCO, to test which variables would improve the fit (Tables 4.1 and 4.3). The CCA ordination diagram of the alkaloid extraction depicts the four most important variables contributing to the chemical variation between the sites as: mean temperature of wettest quarter as well as the content of phosphorus, magnesium, and V% - base saturation in the soil (Figure 4.2b). The CCA ordination diagram of the phenolic extraction depicts the four most important variables contributing to this chemical variation between the sites as: mean temperature of the wettest quarter, temperature seasonality, calcium and EC - soil electrical conductivity (Figure 4.2d). The explained variation of each axis of the CCA ordination diagrams is shown in Table 4.2. Next, we tested the constrained axes to verify the significance of the axes, and for both alkaloid and phenolic extractions we found only the first axis to be significant (Table 4.4).

Table 4.2. Explained variation from axes in the CA and CCA ordination diagrams

Alkaloid Extraction	Axis 1	Axis 2	Axis 3	Axis 4
Unconstrained (CA) explained variation (cumulative)	30.95	47.65	58.04	65.61
Constrained (CCA) explained variation (cumulative)	11.23	18.14	21.41	23.33
Pseudo-canonical correlation	0.629	0.715	0.658	0.586

Phenolic Extraction	Axis 1	Axis 2	Axis 3	Axis 4
Unconstrained (CA) explained variation (cumulative)	29.09	41.28	50.48	58.84
Constrained (CCA) explained variation (cumulative)	13.98	20.47	24.70	27.83
Pseudo-canonical correlation	0.751	0.845	0.746	0.649

Table 4.3: Interactive forward selection of environmental variables

Alkaloid Extraction	Explains %	Contribution %	P-value
Mean Temp Wettest Qtr.	11.1	35.4	0.008
P	8.9	28.4	0.018
V%-Base Saturation	4.5	14.4	0.377
Mg	3.5	11.1	0.664

Phenolic Extraction	Explains %	Contribution %	P-value
Mean Temp Wettest Qtr.	8.5	28.6	0.032
Ca	7.4	24.8	0.073
EC	6.0	20.1	0.167
Temp Seasonality	4.7	15.8	0.397

Table 4.4: Significance of constrained axes of CCA diagrams

Alkaloid Extraction	Axis 1	Axis 2	Axis 3	Axis 4
Explained by constrained axis	14.71%	8.60%	3.06%	1.53%
P-value	0.007	0.03	0.68	0.894

Phenolic Extraction	Axis 1	Axis 2	Axis 3	Axis 4
Explained by constrained axis	13.95%	6.48%	4.21%	1.86%
P-value	0.013	0.163	0.4	0.867

Classification of chemoecotypes

To further confirm the classification of chemoecotypes by site we used Partial Least Squares-Discriminant Analysis and Random Forests Analysis. PLS-DA was used on the log-transformed LCMS data. Possible hypotheses for the chemical variation at sites include: no variation between sites leading to one chemotype for all sites (Supplementary Figure 4.2a), visible groupings or chemoecotypes identified per site (Supplementary Figure 4.2b), or random chemical variation within and among sites (Supplementary Figure4.2c). For both alkaloid and phenolic extractions we were able to confirm group classifications / chemoecotypes for each site. For the alkaloid extraction, we note a clear grouping of individuals from each site in the PLS-DA 3-D scores plot,

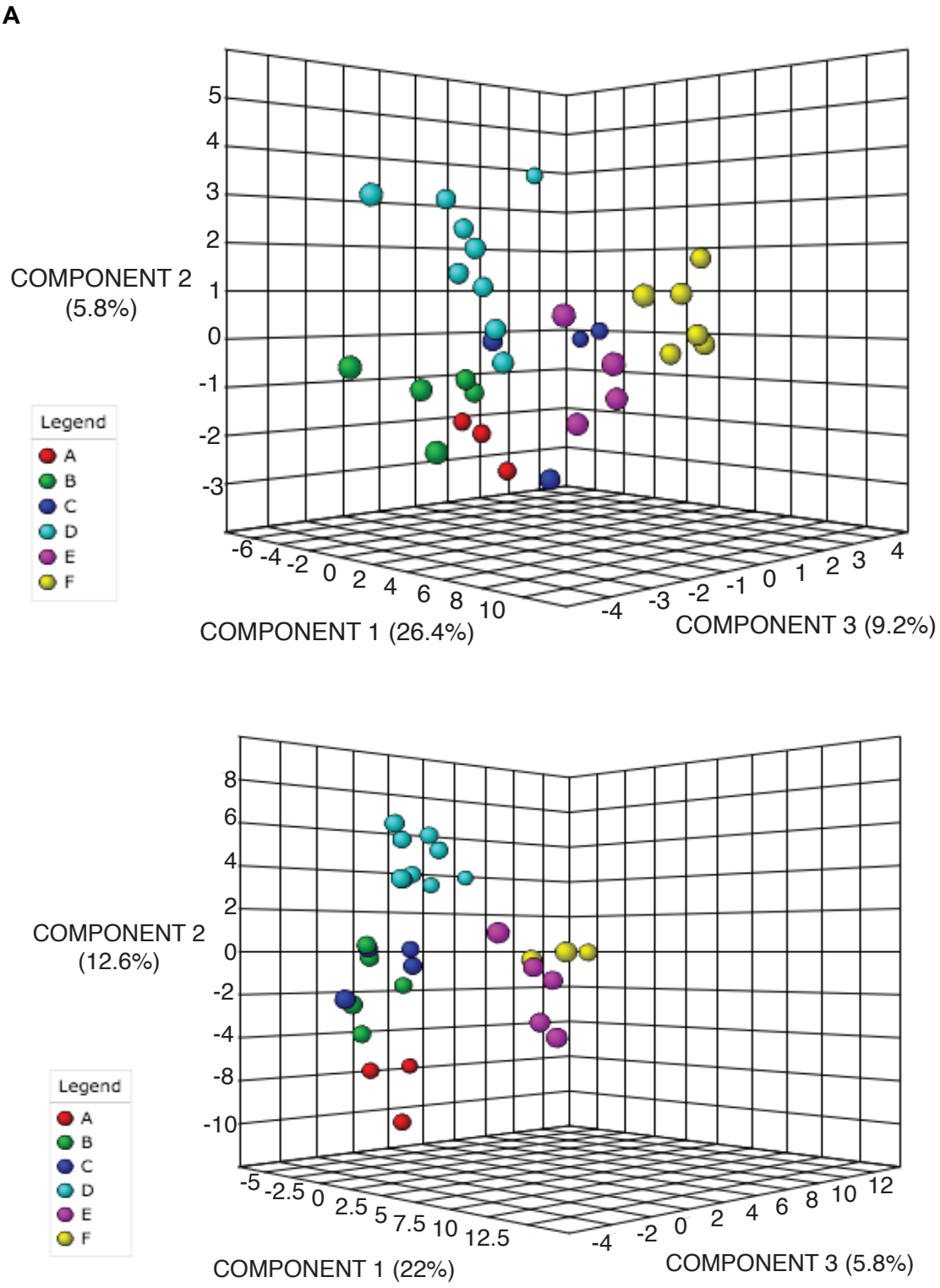


Figure 4.3 PLS-DA of Alkaloid and Phenolic extractions depicts differentiation between sites. 3D scores plot between selected PCs. The explained variances are shown in brackets. Site locations (A-F) are color-coded and ID's refer to Table 4.1. (a) PLS-DA for Alkaloid extraction (b) PLS-DA for phenolic extraction.

though we note that Site C has 1 outlier (Figure 4.3a). In the phenolic extraction, we also see a grouping of individuals from each site on the PLS-DA 3-D scores plot (Figure 4.3). PLS-DA 3-D scores plot of all biological replicates at all sites was also completed, and a dense grouping at each site is noted. (Supplemental Figure 4.3a, 4.3b). Statistical validations of the class separations of PLS-DA were done through the cross-validation of different numbers of components and permutation tests of randomly assigned class labels using the optimal number of components (Supplemental Figure 4.4, Supplemental Table 2).

Random Forest Analysis was also utilized for class prediction. In Figure 4.4a/4.4b we see that the classification of sites based on the alkaloid extraction (Figure 3a) is more defined than the phenolic extraction (Figure 4.4b). This is also confirmed by the OOB error and error rates for each class seen in Table 4.5. The variation in certain groups or sites could be due to sample size; however, it is not possible to acquire many individuals of this species at every location.

Table 4.5: Random Forests classification performance matrix

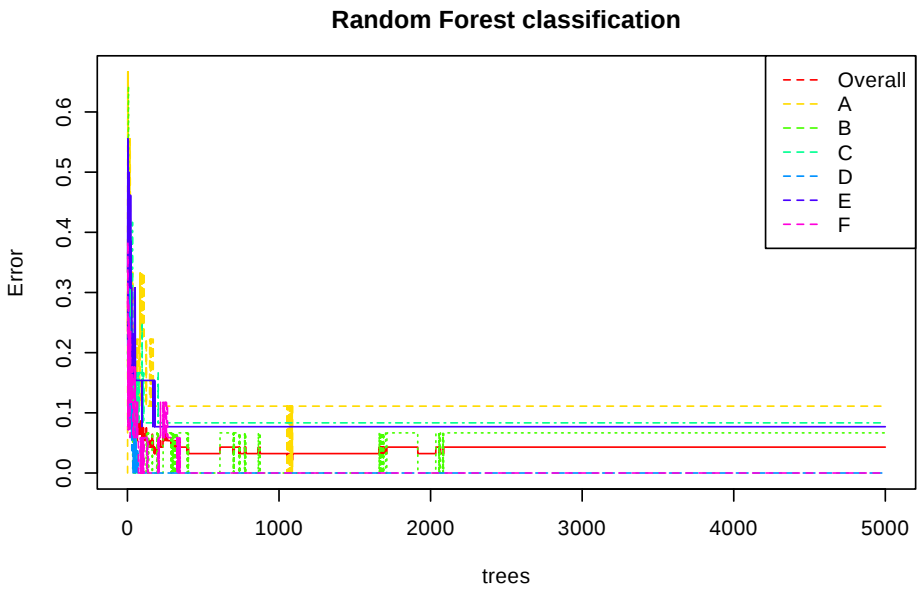
Alkaloid Extraction: OOB Error = 0.0323

	A	B	C	D	E	F	Class Error
A	9	0	0	0	0	0	0
B	0	14	0	1	0	0	0.0667
C	0	0	11	1	0	0	0.0833
D	0	0	0	27	0	0	0
E	0	0	0	0	12	1	0.0769
F	0	0	0	0	0	17	0

Phenolic Extraction: OOB Error = 0.0909

	A	B	C	D	E	F	Class Error
A	9	0	0	0	0	0	0
B	0	13	0	2	0	0	0.133
C	0	0	12	0	0	0	0
D	0	0	1	26	0	0	0.037
E	0	0	0	0	15	0	0
F	0	1	0	0	4	5	0.5

A



B

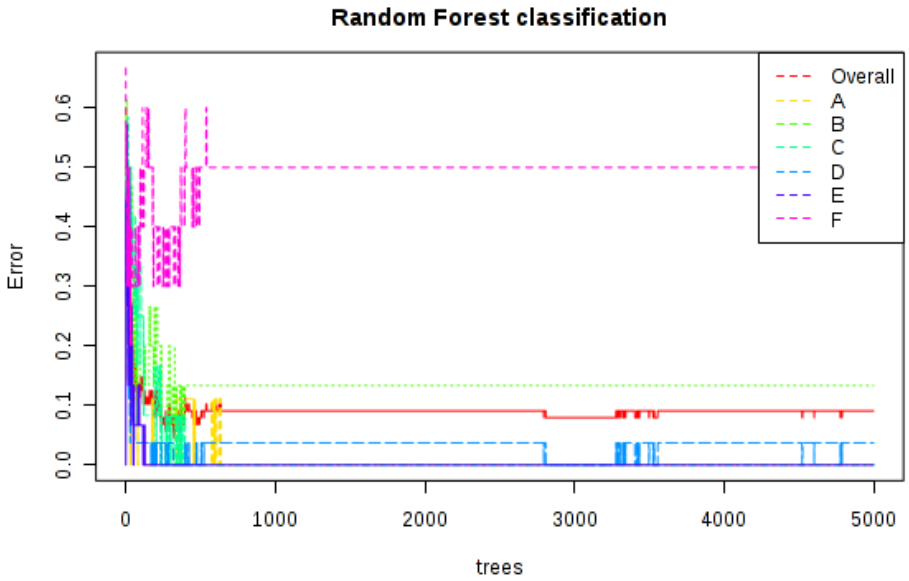


Figure 4.4 Random Forests (RF) Classification of the differentiation between sites for alkaloid and phenolic extractions. Each color represents the cumulative error rates for the RF classification of each site as a unique class /chemoecotype. A site with a smaller error rate that is stabilized with a fewer number of decision trees is considered to be more accurately distinguished as a unique chemoecotype. The overall error rate for chemoecotype classification is shown as a red line. Site locations are color coded by the legend on this graph and labeled with ID's (A-F) as found on Table 4.1. The alkaloid extraction has the lowest overall error rate for classification of chemoecotypes, therefore chemoecotype groups are more easily distinguished or identified (a) alkaloid extraction (b) phenolic extraction

DISCUSSION

Just as it is possible to determine species growing in an area based on environmental conditions present at a location (Braak, 1986; Harbert & Nixon, 2015), it is also possible to predict which compounds or compound profiles will be more abundant at environmental optima through the use of direct gradient analyses. Metabolomic expression across the distribution of a plant species can follow a non-linear, non-monotonic relationship with environmental variables, and this can be verified in situations of large environmental gradients (Whittaker, 1956). To accurately analyse non-linear responses, unimodal methods of ordination diagrams are preferred.

In our study, we analyzed the metabolite profiles of *P. pennatifolius* populations using two extraction protocols, alkaloid and phenolic. For the alkaloid extract, we found that the most significant environmental variables contributing to the variation in alkaloid profiles were mean temperature of the wettest quarter, as well as the soil content of phosphorus, magnesium, and V% (base saturation). Temperature and precipitation are known to affect alkaloid content in plants, and thus variations in these bioclimatic factors could provide variations in chemical profiles (Harborne, 1993; Waller, 1978). It is usually expected that nitrogen availability has a strong effect on alkaloid production as alkaloids are N containing compounds, but this is not always true. Previous stress studies irrigating nutrient solutions on *P. microphyllus* seedlings found that N omission in nutrient solutions did not affect pilocarpine production as much as K omission, as is further shown in our study (Avancini et al., 2003). This low response could be due to a reduced availability of nitrogen in an accessible form or other limiting nutrients that are needed in conjunction with nitrogen for a balanced nutrition. Base saturation (V%) indicates the amount of basic cations in the soil, and is essential for determining soil fertility as it specifies the percentage of sodium, calcium,

potassium, and magnesium that is a part of the cation exchange capacity (CEC). Sites A and B have high CEC values (250.50 and 232.05); however, their respective V% values (94% and 39.5%) demonstrate the great difference in fertility of these soils. The poor soil fertility of Site B is also verified by the lower pH and higher H+Al values. Minerals such as phosphorus and magnesium are very important in various enzymatic processes throughout the plant, and they have large roles in promoting plant development and growth. Studies examining the effect of increases in phosphorous and magnesium have had contradictory results in regards to the effect on alkaloids, ranging from no effect to both positive and negative effects (Bramble & Graves, 1991; Mazzafera, 1999; Waller, 1978). A study that examined the effect of Phosphorus omission on *P. microphyllus* found no effect on the pilocarpine content; however, the authors did not look at the effect on the overall alkaloid profile (Abreu et al., 2005; Avancini et al., 2003).

The most significant environmental variables contributing to the variation in the phenolic profiles in the phenolic extracts were mean temperature in the wettest quarter, temperature seasonality, as well as the soil content of calcium and the electrical conductivity of the soil (EC). Temperature seasonality has been known to have a strong effect on compound expression, since areas of greater seasonality are faced with more variation and could lead to greater variation in chemistry (Molina-Montenegro & Naya, 2012). Specifically, work on the synthesis of phenolics has found that they are regulated not only by developmental signals, but also signals from the environment (A. Cristina Figueiredo, 2008; Carbone et al., 2009; Jaakola, 2013). The potential role of calcium is vast as it is essential for membrane permeability, in addition to being a messenger for pathways in development, responses to environmental stimuli and in response to plant defense (Hepler, 2005; Lecourieux, Ranjeva, & Pugin, 2006; Zhang, Du, & Poovaiah, 2014). In fact, various studies have shown that calcium treatment has led to an increase

in phenolics, upregulating important genes in phenolic compound metabolism (Ahmad et al., 2016; Xu et al., 2014; Zhang et al., 2014). On the other hand, it appears that Phenylalanine ammonia-lyase (PAL), the first enzyme of the phenylpropanoid biosynthetic pathway, is negatively affected by high concentrations of calcium; therefore, it is likely that a variation in calcium could affect the phenolic profile in *P. pennatifolius* (Xu et al., 2014; Teixeira, Andrade, Ferrarese-Filho, & Ferrarese, 2006; Chishaki & Horiguchi, 1997; Tomás-Barberán, Gil, Castañer, Artés, & Saltveit, 1997). PAL is indispensible for the synthesis of a range of metabolites, and it has been found that PAL activity also varies when faced with environmental stimuli such as thermal stress (Christie, Alfenito, & Walbot, 1994; Hunter, Malcolm, & Hartley, 1996; Payyavula, Navarre, Kuhl, Pantoja, & Pillai, 2012; Rivero et al., 2001). Soil electrical conductivity (EC) is also important for plant defense and stress management. EC is the ability of the soil to conduct electricity, and it is determined by a variety of factors including soil moisture, clay content, soil temperature and salinity (including salts such as Na, Ca, K, Mg, Cl, SO4, CO3) (Hanlon, 1993). Various studies have found that higher soil conductivity can indicate greater fertility when there is a higher CEC or a higher sum of bases (Ca, Mg, and K)(C. W. Fraisse, K. A. Sudduth, & N. R. Kitchen, 2001; Moral, Terrón, & Silva, 2010; Officer et al., 2004). This greater availability of nutrients can increase the fertility of the soil, thereby promoting growth and subsequently metabolites produced, or it could also present a decrease in metabolites as per the growth-defense trade-off (Coley, Bryant, & Chapin, 1985; Endara & Coley, 2011; Fine et al., 2006; Herms & Mattson, 1992; Hunter et al., 1996). In this study the sum of bases is highly positively correlated with CEC (Pearson's $r = 0.94$); however, EC has a moderately positive correlation with CEC (Pearson's $r = 0.28$) and EC has a moderately positive correlation with the sum of bases (Pearson's $r = 0.34$). As the correlation of EC

and CEC is weak, it is possible that other factors are affecting the value of EC in these soil samples. These factors could be related to soil moisture/temperature or the clay soil type present, which in combination can affect the soil microbiome, thus affecting the phenolic profile present (Badri, Zolla, Bakker, Manter, & Vivanco, 2013; Sudduth, Kitchen, Bollero, Bullock, & Wiebold, 2003). Further work is required to assess soil texture and water drainage contributions to the EC values at these sites.

There are many factors that can affect a plant's metabolome and the response of each plant depends on its genetic profile, its environment, as well as the interaction between genotype and the environment. Plasticity is the term used to describe how one genotype can have multiple phenotypes, depending on the environment (Thoday, 1953). Local adaptation depends on both genetic and environmental factors to determine phenotypes with the greatest fitness; however, there is a gradient of importance for each quality (Spichtig & Kawecki, 2004). Several studies have found that areas of stable environmental factors, such as constant temperatures (i.e., the tropics), are correlated with reduced plasticity of plant specialized metabolites (Gratani, 2014; Molina-Montenegro & Naya, 2012; Sultan, 1987; Tack et al., 2012). As plants are immobile, variation in abiotic and biotic interactions leads to an adaptation to the environment, through the modification of a plant's chemical defense. Consequently, stability in the environment would reduce the overall stresses facing the plant, decreasing the selective forces favouring plasticity of the plant to change its metabolite phenotype for survival, which reduces its metabolite variation. There are many instances of a single plant species having different phytochemical profiles in different locations (Andrade-Neto, Cunha, Mafezoli, & Silveira, 2002; Endara & Coley, 2011; Hu et al., 2007; Hunter et al., 1996; Koricheva, 1999; Moore et al., 2014; Moustafa, Hesham, Quraishi, & Alrumman, 2016; O'Reilly-Wapstra et al., 2013). This could be due to

different chemotypes (distinct genetically determined phytochemical profiles that are not evolutionary plastic), plasticity of a single genotype, or gradients of plasticity with genetic variance (Gratani, 2014; Keefover-Ring, Thompson, & Linhart, 2009; Santesson, 1968). Determining the plasticity of a species is difficult without a common garden or transplant experiment.

As our study used only ordination methods to associate environmental variables with variation in chemical profiles of *P. pennatifolius*, and correlation does not imply causation, there are still many laboratory and greenhouse experiments needed, such as seedling stress tests that can be done to further confirm the effect of environmental variables on the chemical profile. In addition, further identification of the compounds found to be significantly different between chemoecotypes through CA, CCA, PLS-DA, and RF could lead to a more targeted analysis and understanding of biological pathways. Common garden experiments, planting different genotypes in the same location, could also be very useful to detect whether differences in chemical variation are indeed due to environmental factors or genetic differences. Either result would be beneficial for breeding as one could breed individuals for specific genes, or one could add/omit nutrients and stressors to modify compound yields. In the case of *P. pennatifolius*, a rare and slow growing tree, we are unable to run a common garden experiment. Therefore this exploratory analysis in wild populations is advantageous and essential as it has reduced the environmental variables, and will allow for a more guided experimental analysis.

Conclusion

Analysis of wild populations of *P. pennatifolius* has provided us with potential environmental variables that should be followed up with greenhouse and in-field experiments to determine their importance in alkaloid and phenolic biosynthesis. Of

course, the expression of compound profiles in plants is related to a variety of conditions, plant-environment interactions, plant genotype and genotype-environment interactions. Future work in our laboratory will examine population genetic differences among the various sites to assess variation that can be attributed to specific genotypes.

As the field of metabolomics is expanding into larger ecological studies examining wild populations, it is important to determine the best way to assess the response curves of compound profiles present in samples of a large population gradient. As there are non-monotonic and non-linear relationships between compound profiles and environmental variables, the use of unimodal methods (CA, CCA) instead of linear methods (PCA, RDA) will be much more accurate to describe the variation in ecometabolomic studies. As such our study has used exploratory data analysis methods common in community and microbial ecology studies, to perform an ecological metabolomics study of wild populations of *P. pennatifolius.* These unimodal methods can also be used in studies involving analysis of morphology, behavior, pollination, or physiological responses in wild populations across large gradients.

FUNDING SOURCES

This work was supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1650441); Harold E. Moore Jr. and Robert T. Clausen Endowment Fund; BSA; SSB; ASPT; Garden Club of America; and the Cornell Graduate School.

ACKNOWLEDGMENTS

We thank Professor Milton Groppo (USP-RP), Professor Pedro Dias and Dr. Rubens Coelho (USP-SP), and Professor Marcelo Caxambu (Universidade Tecnologica Federal do Parana) for helping with the collection of *Pilocarpus* specimens; Iguaçu Itaipu and Parque Estadual Lago Azul, PR for allowing the collection of specimens from their property; Dr. Alexandra Christine Helena Frankland Sawaya for providing the updated extraction protocol for imidazole alkaloids; Instituto Agronomico in Campinas, SP, Brazil for analyzing the soil samples and the statistical consulting group of Cornell University for confirming the statistical analyses. PM thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brasil) for a research fellowship and Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp - grant 2008/58035-6) for research support.

Plants collected in public locations were collected under ICMBio SISBIO permit #52758-1, allowing for the collection of botanical, fungal and microbiological specimens. In addition, plants were collected in collaboration with and under the permits of Professor Milton Groppo, Professor Pedro Dias, and Professor Marcelo Caxambu.

LITERATURE CITED

A. Cristina Figueiredo JGB. **2008**. Factors affecting secondary metabolite production in plants: volatile components and essential oils. Flavour Fragr J. *Flavour and Fragrance Journal* **23**: 213–226.

Abreu I.N., Mazzafera P, Eberlin MN, Zullo MAT, Sawaya ACHF. **2007**. Characterization of the variation in the imidazole alkaloid profile of Pilocarpus microphyllus in different seasons and parts of the plant by electrospray ionization mass spectrometry fingerprinting and identification of novel alkaloids by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **21**: 1205–1213.

Abreu I.N., Sawaya ACHF, Eberlin MN, Mazzafera P. **2005**. Production of pilocarpine in callus of jaborandi (Pilocarpus microphyllus stapf). *In Vitro Cellular & Developmental Biology - Plant* **41**: 806–811.

Ahmad P, Latef A, A A, *et al.* 2016. Calcium and Potassium Supplementation Enhanced Growth, Osmolyte Secondary Metabolite Production, and Enzymatic Antioxidant Machinery in Cadmium-Exposed Chickpea (Cicer arietinum L.). *Frontiers in Plant Science* **7**.

Albert A, Sareedenchai V, Heller W, Seidlitz HK, Zidorn C. 2009. Temperature Is the key to Altitudinal Variation of Phenolics in Arnica montana L. cv. ARBO. Oecologia **160**, 1–8.

Andrade-Neto M, Cunha U de A, Mafezoli J, Silveira ER. **2002**. Volatile Constituents of Different Populations of Pilocarpus spicatus Saint Hill. (Rutaceae) from the Northeast of Brazil. *Journal of Essential Oil Research* **14**: 319–324.

Andreazza NL, Abreu IN, Sawaya ACHF, Eberlin MN, Mazzafera P. **2009**. Production of imidazole alkaloids in cell cultures of jaborandi as affected by the medium pH. *Biotechnology Letters* **31**: 607–614.

Arbona V, Manzi M, Ollas C, Gómez-Cadenas A. 2013. Metabolomics as a Tool to Investigate Abiotic Stress Tolerance in Plants. International Journal of Molecular Sciences **14**, 4885–4911.

Avancini G, Abreu IN, Saldaña MDA, Mohamed RS, Mazzafera P. **2003**. Induction of pilocarpine formation in jaborandi leaves by salicylic acid and methyljasmonate. *Phytochemistry* **63**: 171–175.

Badri DV, Zolla G, Bakker MG, Manter DK, Vivanco JM. **2013**. Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. *The New Phytologist* **198**: 264–273.

Braak CJFT. **1986**. Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology* **67**: 1167–1179.

Bramble J, Graves D. **1991**. Calcium and Phosphate Effects on Growth and Alkaloid Production in Coffea arabica: Experimental Results and Mathematical model. *Biotechnology and Bioengineering* **37**: 859–868.

C. W. Fraisse, K. A. Sudduth, N. R. Kitchen. **2001**. Delineation of Site-Specific Management Zones by Unsupervised Classification of Topographic Attributes and Soil Electrical Conductivity. *Transactions of the ASAE* **44**: 155–166.

Camargo OA, Moniz AC, Jorge JA, Valadares JMAS. **2009**. *Métodos de Análise Química, Mineralógica e Física de Solos do Instituto Agronômico de Campinas*. Campinas, SP: Instituto Agronomico.

Carbone F, Preuss A, De V, *et al.* 2009. Developmental, genetic and environmental factors affect the expression of flavonoid genes, enzymes and metabolites in strawberry fruits. *Plant, Cell and Environment* **32**: 1117–1131.

Chambers MC, Maclean B, Burke R, *et al.* 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **30**: 918–920.

Chishaki N, Horiguchi T. **1997**. Responses of secondary metabolism in plants to nutrient deficiency In: Developments in Plant and Soil Sciences. *Plant Nutrition for Sustainable Food Production and Environment*. Springer, Dordrecht, 341–345.

Christie PJ, Alfenito MR, Walbot V. **1994**. Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: Enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta* **194**: 541–549.

Cipollini ML. **2000**. Secondary metabolites of vertebrate-dispersed fruits: evidence for adaptive functions. *Revista chilena de historia natural* **73**: 421–440.

Coley PD, Bryant JP, Chapin FS. **1985**. Resource Availability and Plant Antiherbivore Defense. *Science* **230**: 895–899.

Dell M, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME, others. **2014**. Targeting environmental adaptation in the monocot model Brachypodium distachyon: a multi-faceted approach. *BMC genomics* **15**: 801.

Draper NR, Smith H. 1998. *Applied Regression Analysis*. New York: Wiley-Interscience.

Durand-Hulak M, Dugrand A, Duval T, *et al.* 2015. Mapping the genetic and tissular diversity of 64 phenolic compounds in Citrus species using a UPLC-MS approach. *Annals of Botany* **115**: 861–877.

Earll M. **2012**. MZmine Tutorial

Endara M-J, Coley PD. **2011**. The resource availability hypothesis revisited: a meta-analysis: Revisiting the resource availability hypothesis. *Functional Ecology* **25**: 389–398.

Fine PVA, Metz MR, Lokvam J, *et al.* 2013. Insect herbivores, chemical innovation, and the evolution of habitat specialization in Amazonian trees. *Ecology* **94**: 1764–1775.

Fine PVA, Miller ZJ, Mesones I, *et al.* 2006. The growth-defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* **87**: S150-162.

Gratani L. **2014**. Plant Phenotypic Plasticity in Response to Environmental Factors. *Advances in Botany* **2014**: 1–17.

Hanlon EA. **1993**. *Soil pH and Electrical Conductivity: A County Extension Soil Laboratory Manual*. UF/IFAS Extension, University of Florida.

Harbert RS, Nixon KC. **2015**. Climate reconstruction analysis using coexistence likelihood estimation (CRACLE): A method for the estimation of climate using vegetation. *American Journal of Botany* **102**: 1277–1289.

Harborne JB, Jeffrey B. 1993. *Introduction to ecological biochemistry*. London ; San Diego: Academic Press.

Haslam E. **1986**. Secondary metabolism – fact and fiction. *Natural Product Reports* **3**: 217–249.

Hepler PK. **2005**. Calcium: A Central Regulator of Plant Growth and Development. *The Plant Cell* **17**: 2142–2155.

Herms DA, Mattson WJ. **1992**. The Dilemma of Plants: To Grow or Defend. *The Quarterly Review of Biology* **67**: 283–335.

Hill MO. **1973**. Reciprocal Averaging: An Eigenvector Method of Ordination. *Journal of Ecology* **61**: 237–249.

Hill MO. **1974**. Correspondence Analysis: A Neglected Multivariate Method. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **23**: 340–354.

Hiserodt R, Chen L. **2012**. An LC/MS/MS Method for the Analysis of Furocoumarins in Citrus Oil In: ACS Symposium Series. *Recent Advances in the Analysis of Food and Flavors*. American Chemical Society, 71–88.

Hu Y, Zhang Q, Xin H, *et al.* 2007. Association between chemical and genetic variation of Vitex rotundifolia populations from different locations in China: its implication for quality control of medicinal plants. *Biomedical chromatography: BMC* **21**: 967–975.

Hunter MD, Malcolm SB, Hartley SE. **1996**. Population-level variation in plant secondary chemistry, and the population biology of herbivores. *CHEMOECOLOGY* **7**: 45–56.

Izhaki I. **2002**. Emodin – a secondary metabolite with multiple ecological functions in higher plants. *New Phytologist* **155**: 205–217.

Jaakola L. **2013**. New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science* **18**: 477–483.

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL. 2004. Exploring the temperature-stress metabolome of Arabidopsis. Plant physiology **136**, 4159–4168.

Keefover-Ring K, Thompson JD, Linhart YB. **2009**. Beyond six scents: defining a seventh Thymus vulgaris chemotype new to southern France by ethanol extraction. *Flavour and Fragrance Journal* **24**: 117–122.

Koricheva J. **1999**. Interpreting phenotypic variation in plant allelochemistry: problems with the use of concentrations. *Oecologia* **119**: 467–473.

Lankadurai BP, Nagato EG, Simpson MJ. 2013. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. Environmental Reviews **21**, 180–205.

Lecourieux D, Ranjeva R, Pugin A. **2006**. Calcium in plant defence-signalling pathways. *The New Phytologist* **171**: 249–269.

Lepš J, Šmilauer P. **2003**. *Multivariate Analysis of Ecological Data using CANOCO*. Cambridge, UK ; New York: Cambridge University Press.

Mazzafera P. **1999**. Mineral nutrition and caffeine content in coffee leaves. *Bragantia* **58**: 387–391.

Molina-Montenegro MA, Naya DE. **2012**. Latitudinal Patterns in Phenotypic Plasticity and Fitness-Related Traits: Assessing the Climatic Variability Hypothesis (CVH) with an Invasive Plant Species (F Seebacher, Ed.). *PLoS ONE* **7**: e47620.

Moore BD, Andrew RL, Külheim C, Foley WJ. **2014**. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist* **201**: 733–750.

Moral FJ, Terrón JM, Silva JRM da. **2010**. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research* **106**: 335–343.

Moustafa MF, Hesham AE-L, Quraishi MS, Alrumman SA. **2016**. Variations in genetic and chemical constituents of Ziziphus spina-christi L. populations grown at various altitudinal zonation up to 2227m height. *Journal of Genetic Engineering and Biotechnology* **14**: 349–362.

Ncube B, Finnie JF, Van Staden J. 2012. Quality from the field: The impact of environmental factors as quality determinants in medicinal plants. South African Journal of Botany **82**, 11–20.

Officer SJ, Kravchenko A, Bollero GA, *et al.* 2004. Relationships between soil bulk electrical conductivity and the principal component analysis of topography and soil fertility values. *Plant and Soil* **258**: 269–280.

O'Reilly-Wapstra JM, Miller AM, Hamilton MG, Williams D, Glancy-Dean N, Potts BM. **2013**. Chemical Variation in a Dominant Tree Species: Population Divergence, Selection and Genetic Stability across Environments. *PLOS ONE* **8**: e58416.

Paliy O, Shankar V. **2016**. Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology* **25**: 1032–1057.

Payyavula RS, Navarre DA, Kuhl JC, Pantoja A, Pillai SS. **2012**. Differential effects of environment on potato phenylpropanoid and carotenoid expression. *BMC Plant Biology* **12**: 39.

Pennycooke JC, Cox S, Stushnoff C. 2005. Relationship of cold acclimation, total phenolic content and antioxidant capacity with chilling tolerance in petunia (Petunia x hybrida). Environmental and Experimental Botany **53**, 225–232.

Pinheiro CU. **1997**. Jaborandi (Pilocarpus sp., rutaceae): A wild species. *Economic Botany* **51**: 49–58.

Pinheiro CU. **2002**. Extrativismo, cultivo e privatização do jaborandi (Pilocarpus microphyllus Stapf ex Holm.; Rutaceae) no Maranhão, Brasil. *Acta Botanica Brasilica* **16**: 141–150.

Pluskal T, Castillo S, Villar-Briones A, Orešič M. **2010**. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**: 395.

Raij B van. **2001**. *Análise química para avaliação da fertilidade de solos tropicais*. Instituto Agronômico.

Rivero RM, Ruiz JM, García PC, López-Lefebre LR, Sánchez E, Romero L. **2001**. Resistance to cold and heat stress: accumulation of phenolic compounds in tomato and watermelon plants. *Plant Science* **160**: 315–321.

Sampaio BL, Edrada-Ebel R, Costa FBD. 2016. Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for environmental metabolomics of plants. Scientific Reports **6**, 29265.

Santesson R. **1968**. *Svensk Naturvetenskap*. Stockholm, Sweden: Allf.

Santos AP, Moreno PRH. **2004**. Pilocarpus spp.: a survey of its chemical constituents and biological activities. *Revista Brasileira de Ciências Farmacêuticas* **40**: 116–137.

Sardans J, Penuelas J, Rivas-Ubach A. **2011**. Ecological metabolomics: overview of current developments and future challenges. *Chemoecology* **21**: 191–225.

Sawaya ACHF, Vaz BG, Eberlin MN, Mazzafera P. **2011**. Screening species of Pilocarpus (Rutaceae) as sources of pilocarpine and other imidazole alkaloids. *Genetic Resources and Crop Evolution* **58**: 471–480.

Skorupa LA. **1996**. Revisao Taxonomica de Pilocarpus Vahl (Rutaceae). Ph.D. Thesis.

Šmilauer P, Lepš J. **2014**. *Multivariate Analysis of Ecological Data using CANOCO 5*. Cambridge, United Kingdom ; New York: Cambridge University Press.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. **2006**. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **78**: 779–787.

Spichtig M, Kawecki TJ. **2004**. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. *The American Naturalist* **164**: 70–84.

Sudduth KA, Kitchen NR, Bollero GA, Bullock DG, Wiebold WJ. **2003**. Comparison of Electromagnetic Induction and Direct Sensing of Soil Electrical Conductivity. *Agronomy Journal* **95**: 472–482.

Sultan S. **1987**. Evolutionary Implications of Phenotypic Plasticity in Plants. *Evolutionary Biology* **21**.

Tack AJM, Johnson MTJ, Roslin T. **2012**. Sizing up community genetics: it's a matter of scale. *Oikos* **121**: 481–488.

Teixeira AF, Andrade A de B, Ferrarese-Filho O, Ferrarese M de LL. **2006**. Role of Calcium on Phenolic Compounds and Enzymes Related to Lignification in Soybean (*Glycine max L.*) Root Growth. *Plant Growth Regulation* **49**: 69–76.

Thoday JM. **1953**. Components of fitness In: *Symposium of the Society for Experimental Biology*.

Tine Y, Renucci F, Costa J, Wélé A, Paolini J. **2017**. A Method for LC-MS/MS Profiling of Coumarins in Zanthoxylum zanthoxyloides (Lam.) B. Zepernich and Timler Extracts and Essential Oils. *Molecules (Basel, Switzerland)* **22**.

Tomás-Barberán FA, Gil MI, Castañer M, Artés F, Saltveit ME. **1997**. Effect of Selected Browning Inhibitors on Phenolic Metabolism in Stem Tissue of Harvested Lettuce. *Journal of Agricultural and Food Chemistry* **45**: 583–589.

Vialart G, Hehn A, Olry A, *et al.* 2012. A 2-oxoglutarate-dependent dioxygenase from Ruta graveolens L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. *The Plant Journal: For Cell and Molecular Biology* **70**: 460–470.

Waller GR. **1978**. *Alkaloid biology and metabolism in plants*. New York: Plenum Press.

Whittaker RH. **1956**. Vegetation of the Great Smoky Mountains. *Ecological Monographs* **26**: 1–80.

Wink M (Ed.). **1999**. *Biochemistry of Plant Secondary Metabolism*. Sheffield: WileyBlackwell.

Xia J, Wishart DS. **2002**. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis In: *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., .

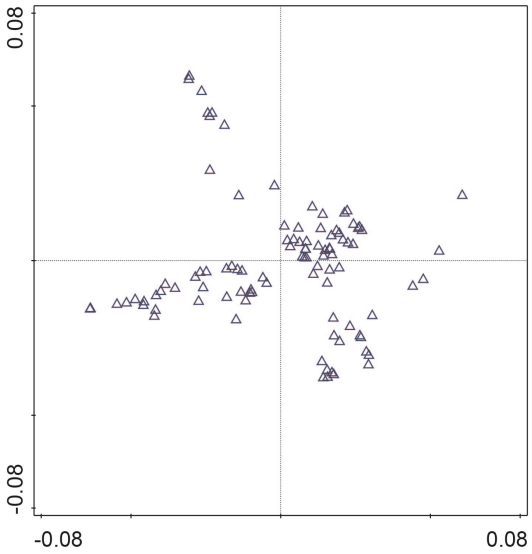
Xu W, Peng H, Yang T, *et al.* 2014. Effect of calcium on strawberry fruit flavonoid pathway gene expression and anthocyanin accumulation. *Plant physiology and biochemistry: PPB* **82**: 289–298.

Yan X, Wu S, Wang Y, Shang X, Dai S. 2004. Soil nutrient factors related to salidroside production of Rhodiola sachalinensis distributed in Chang Bai Mountain. Environmental and Experimental Botany **52**, 267–276.

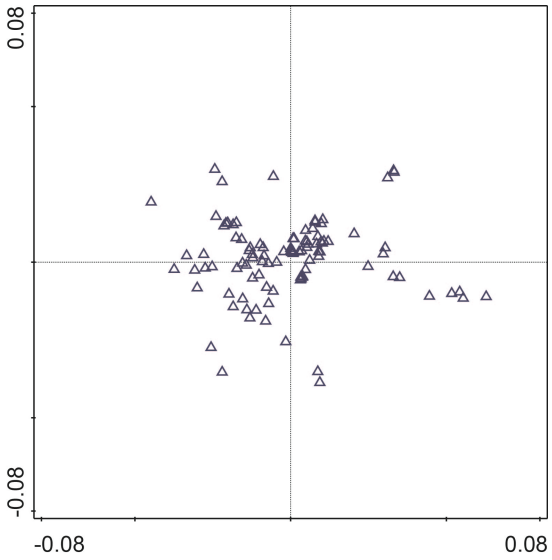
Zhang L, Du L, Poovaiah BW. **2014**. Calcium signaling and biotic defense responses in plants. *Plant Signaling & Behavior* **9**.

SUPPLEMENTARY MATERIALS

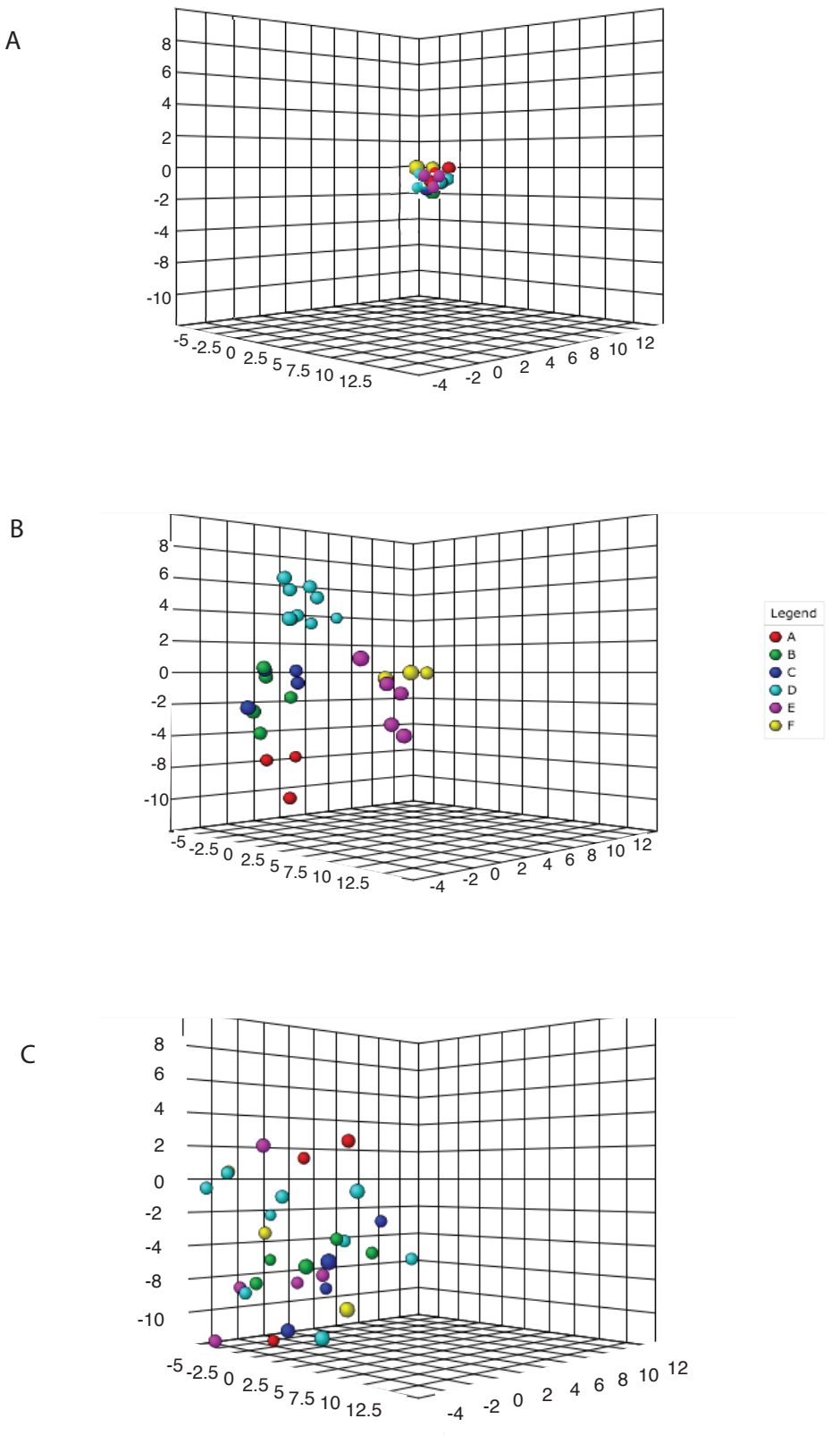
A



B

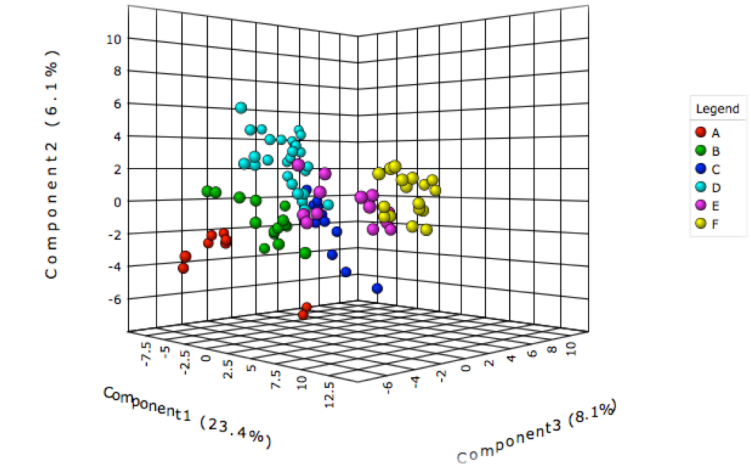


Supplemental Figure 4.1 Correspondence Analysis depicting compound optimas. This plot depicts the variation in compounds across the six sites. Each triangle represents the optima of each compound. (a) alkaloid extraction (b) phenolic extraction

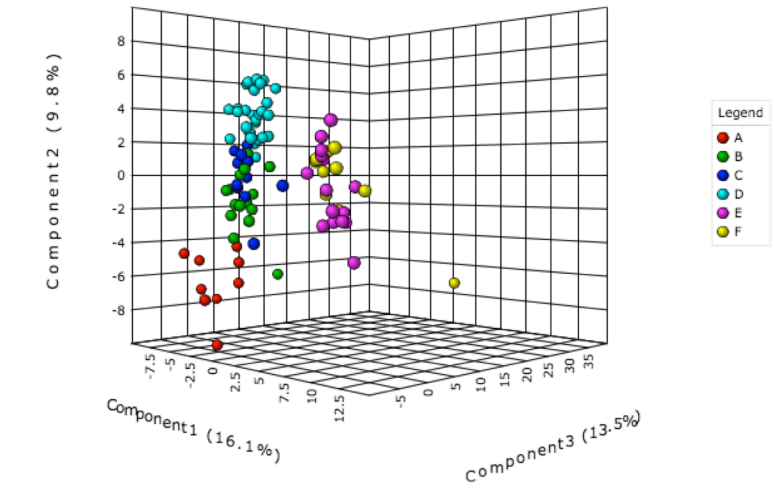


Supplementary Figure 4.2. Possible hypotheses for chemical variation at sites. (a) same chemistry at all sites (b)different chemistry at each site (c)random chemistry within/among sites

A

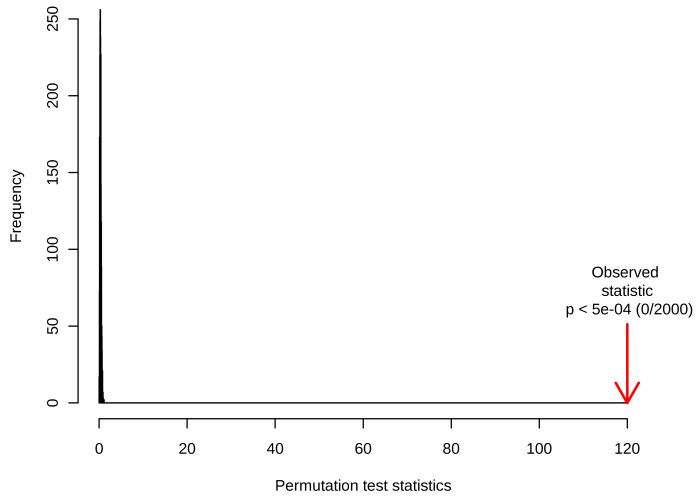


B

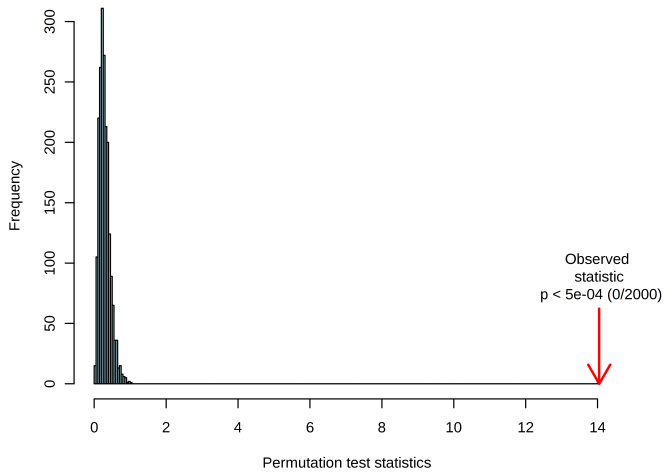


Supplemental Figure 4.3 PLS-DA of all biological replicates for both extractions. 3D scores plot between selected PCs. The explained variances are shown in brackets. (a) represents all biological replicates of the alkaloid extraction (b) represents all biological replicates of the phenolic extraction

A



B



Supplemental Figure 4.4: PLS-DA model validation by permutation test based on separation distance (B/W) displayed as a histogram. Separation distance is based on the ratio of between-group sum of squares and within-group sum of squares. P-value based on permutation test is p<5e^-04 (0/2000). (a) alkaloid extraction (b) phenolic extraction

Supplemental Table 1: Compounds identified in extractions. (a) alkaloids (b) coumarins

A

ID	m/z	RT (min)	Molecular Formula	Compound Name
M193T81	193	1.36	C10H12N2O2	13-nor-7(11) dehydropilocarpine
M179T84	179	1.41	C9H10N2O2	4-(3H-imidazol-4-ylmethyl)-3-methyl-5H-furan-2-one
M195T160	195	2.67	C10H14N2O2	Pilocarpidine / isopilocarpidine
M209T167_2	209	2.79	C11H16N2O2	Pilocarpine / isopilocarpine
M273T193	273	3.21	C15H16N2O3	3-(Hydroxy-phenyl-methyl)-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one
M273T266	273	4.43	C15H16N2O3	3-(Hydroxy-phenyl-methyl)-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one
M259T369	259	6.15	C15H18N2O2	3-Hydroxymethyl-4-(3-methyl-3H-imidazol-4-yl)-1-phenylbutan-1-one
M255T396	255	6.60	C15H14N2O2	3- Benzylidene-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one
M257T407	257	6.79	C15H16N2O2	3-Benzyl-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one
M257T409	257	6.81	C15H16N2O2	3-Benzyl-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one
M255T451	255	7.52	C15H14N2O2	3- Benzylidene-4-(3H-imidazol-4-ylmethyl)-dihydro-furan-2-one

B

ID	m/z	RT (min)	Molecular Formula	Compound Name
M193T45	193	0.75	C10H8O4	Scopoletin
M187T96	187	1.6	C11H6O3	Psoralen
M147T96	147	1.6	C9H6O2	Coumarin
M217T369	217	6.15	C12H8O4	Xanthotoxin
M247T382	247	6.36	C13H10O5	Isopimpinellin
M245T382	245	6.36	C15H16O3	Osthol
M271T433	271	7.22	C16H14O4	Imperatorin
M203T434	203	7.23	C11H6O4	Imperatorin [271-C5H8]

Supplemental Table 2: Performance measures for prediction accuracies using cross-validation of PLS-DA using different numbers of components

A. Alkaloid

Measure	1 comps	2 comps	3 comps
Accuracy	0.48387	0.45161	0.54839
R2	0.6495	0.88856	0.94244
Q2	0.50384	0.60991	0.70436

B. Phenolic

Measure	1 comps	2 comps	3 comps
Accuracy	0.46667	0.46667	0.60000
R2	0.59686	0.82203	0.91282
Q2	0.33854	0.43717	0.4544

CHAPTER 5

Population structure and genetic diversity of three *Pilocarpus* species in Brazil

Daniella M. Allevato [1], Milton Groppo[2], Eduardo Kiyota[3],

Paulo Mazzafera[3], Kevin C. Nixon [1]

ABSTRACT

Pilocarpus, the only source of the drug pilocarpine for the treatment of glaucoma, is a Neotropical genus that is both rare and endangered in the wild. Population genomic studies of this important medicinal plant are essential due to its conservation status, and will be beneficial for both conservation management and breeding programs of this group. The utilization of next-generation sequencing technologies through methods such as double-digest restriction-site associated DNA sequencing (ddRADseq), provides a multitude of SNPS that can be used to assess evolutionary dynamics (drift, selection, mutation, and recombination) in population structure and diversity in non-model species. In this study we provide the first instance of next-generation sequencing of *Pilocarpus,* as we assess the population diversity of three species in the genus: *P. pennatifolius, P. spicatus,* and *P. riedelianus*. We found significant differences among the species in terms of their population structure and diversity, ranging from extremely admixtured populations of *P. pennatifolius* to regionally structured populations of *P. riedelianus* and *P. spicatus*.

KEYWORDS: *Pilocarpus,* Rutaceae, ddRADseq, ecological genomics, population genomics

---

[1]L.H. Bailey Hortorium, Section of Plant Biology, School of Plant Sciences, Cornell University, Ithaca, NY, USA
[2] Departamento de Biologia, Universidade de Sao Paulo, Ribeirão Preto, Brazil
[3] Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Brazil

INTRODUCTION

Pilocarpus is a Neotropical genus in the citrus family (Rutaceae) that is well known as the only source of pilocarpine, an imidazole alkaloid drug used to treat glaucoma and xerostomia. This genus has a wide variety of chemical diversity, including alkaloids, terpenoids, and coumarins, though most interest is in the imidazole alkaloids (Santos and Moreno, 2004, Allevato Chapter 2). Various species of *Pilocarpus* have been extensively overharvested in the wild for the extraction of pilocarpine from the leaves. These practices caused a few species to become threatened in wild populations, which led to the creation of agricultural farms of *P. microphyllus,* a small shrub that has high yields of pilocarpine (IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis), 1992). Overall, the reduction of the range of many *Pilocarpus* species is most likely due to a combination of urbanization and the low germination viability of the seeds (Pinheiro, 1997, 2002; Caldeira *et al.*, 2017). Though the leaves of *P. microphyllus* are now harvested in agricultural farms, over the past twenty years there has also been a decline in leaf production, and consequently pilocarpine yields per hectare (Pinheiro, 1997). Although there has been extensive research to understand the biosynthesis of imidazole alkaloids, unfortunately, there is still much to be learned. Therefore, a better understanding of the genetic diversity and population structure of species in this genus is essential to manage conservation programs, unravel the biosynthesis of imidazole alkaloids, and assist with plant breeding of *Pilocarpus.*

Population genetic studies examining the evolutionary and demographic changes of populations have been revolutionized by the development of next generation sequencing techniques. The access to a large scale of SNPs across the whole genome provides a finer detail for population genomic studies, allowing for the analysis

of non-model systems (Weigel and Nordborg, 2015; Tabima *et al.*, 2017). There are a variety of techniques used to acquire these SNPs including whole-genome sequencing, RNAseq, Exon capture, and reduced representation library sequencing (such as RADseq, GBS, and ddRADseq) (Baird *et al.*, 2008; Ng and Kirkness, 2010; Davey and Blaxter, 2010; Elshire *et al.*, 2011; Grewe *et al.*, 2017). Overall, the methods of reduced representation library sequencing involve cutting the genome at various sites with a specific restriction enzyme or enzymes, followed by high throughput sequencing of the fragments. After filtering SNPs across samples, one can utilize these SNPs to analyze population structure and diversity, demographic processes, and genes under selection. Additionally, ddRADseq (double-digest restriction-site associated DNA sequencing) has been utilized to construct phylogenies and aid in better estimates for patterns in plant population genetics and phylogeography (Emerson *et al.*, 2010; Davey and Blaxter, 2010; Davey *et al.*, 2011, 2013; Peterson *et al.*, 2012).

Previous studies in our laboratory have found strong phenotypic associations between populations of *Pilocarpus* species and their respective geographic locations (Allevato Chapter 3, 4). This is expected as populations tend to diverge and adapt to their environments; however, it is also possible that as this is a medicinal plant, local populations can be affected by human intervention (Odling-Smee *et al.*, 1996; Nicotra *et al.*, 2010; Weinig *et al.*, 2014; Weigel and Nordborg, 2015). Human intervention was observed in a study on *P. microphyllus*, in which RAPD fragments of 93 individuals from 12 populations in two states were used as population genetic markers (Moura *et al.*, 2005). Moura *et al.* (2005) found there were two cases of genetically indistinguishable individuals in two different locations, implying that clonal or mono-typic source plants had been planted there. In another study of *P. microphyllus* using RAPD fragments, populations appeared to have possible associations between the presence/absence of

certain genetic markers and plants with greater quantities of pilocarpine (Sandhu *et al.*, 2006). Since RADseq SNPs are random and not specific to genes related to alkaloid biosynthesis, it is difficult to say whether genotypes developed from the RADseq approach will be highly associated with specific chemotypes.

Our objective was to compare the population structure and diversity of three species in the genus (*P. pennatifolius, P. spicatus,* and *P. riedelianus),* to help advance knowledge on breeding optimization and assist in elucidating the biosynthesis of imidazole alkaloids. We chose these three species because of their variation in chemistry: *P. pennatifolius* with a large diversity of imidazole alkaloids, and *P. spicatus* and *P. riedelianus* with a low diversity of imidazole alkaloids. These species were also chosen because they have larger latitudinal ranges (greater variation of climatic variables) that are both current and historic, and they are not threatened with extinction in natural populations. To assess population structure and summary statistics we utilized SNP data from ddRADseq data in *R* packages *vcfR, adegenet*, and *popper* (Jombart, 2008; Baird *et al.*, 2008; Peterson *et al.*, 2012; Kamvar *et al.*, 2014; Knaus and Grünwald, 2017). Based on our field observations we expected less genetic variation within populations for *P. riedelianus* and *P. spicatus,* as the growth habit of these species appeared to flourish in denser groupings. In addition, since these species are used medicinally by herbalists, both in the past and today, we expected there to be a possibility of detecting greater gene flow or the influence of human intervention throughout populations in different regions.

MATERIALS AND METHODS

Field collections of plant material

Field collected plants were sampled from six locations for *P. pennatifolius*, three locations for *P. spicatus,* and three locations for *P. riedelianus.* As many individuals as possible (3-10 individuals) were obtained from each site. Locations and numbers of individuals for each species are listed in Table 5.1 and can be visualized on the map in Supplemental Figure 5.1. Herbarium specimens were made for each site and are

Table 5.1. Locations and number of individuals per species populations

Species	State	City	ID	# Individuals
P. pennatifolius	Sao Paulo	Piracicaba		1*
	Parana	Maringa		1*
	Parana	Foz do Iguaçu	A	3
	Parana	Campo Mourao	B	5
	Parana	Campo Mourao B	C	4
	Sao Paulo	Cruz do Pedro	D	5
	Sao Paulo	EE St. Theresa	E	5
	Parana	Lago Azul	F	5**
	Parana		F	3**
	Parana		F	1**
	Sao Paulo	Campinas		1*
		TOTAL		**34**
P. spicatus subsp. *spicatus*	Sao Paulo	Santa Rita	G	5
P. spicatus subsp. *spicatus*	Sao Paulo	Cruz do Pedro	H	5
P. spicatus subsp. *longeracemosus*	Bahia	Rio de Contas	I	5
P. spicatus subsp. *aracatensis*	Bahia	Rio de Contas	J	6
		TOTAL		**21**
P. riedelianus	Bahia	Entre Rios	K	5
	Espirito Santo	Linhares	L	10
	Bahia	Morro do Chapeu	M	9
		TOTAL		**24**

*These sites only contained one individual and are therefore not included in the analyses
**These collections were found at three locations in Lago Azul park, but at distances of about 5-10 km from each other. In the analyses, these three locations were combined for the Lago Azul site

deposited at the Bailey Hortorium Herbarium (BH) at Cornell University and USP Ribeirão Preto Herbarium (SPFR) (APPENDIX 9, APPENDIX 1). Ten leaves were collected from each individual. Specifically, for compound leaves (*P. pennatifolius)* the third leaflet down from the terminal leaflet was chosen, and for simple leaves (*P. spicatus, P. riedelianus)* the third leaf down from the terminal axis was chosen. *Pilocarpus spicatus* has a large latitudinal distribution, but it is believed to be composed of 3 subspecies and 2 varieties. In this study we sampled three subspecies: one present at both sites in Sao Paulo and the other two were growing sympatrically at a site in Bahia.

DNA extraction and sequencing

DNA extractions were done with a few modifications of Barry *et al.*, (2005) . Modifications include the addition of RNase A before incubation in the water bath, the use of ice-cold ethanol, and the use of ice-cold isopropanol. Genome size (2C=5.05pg) for optimization of sequencing was determined using Flow Cytometry at Cornell (APPENDIX 10). DNA extractions were sent to University of Minnesota Genomics Center, Minneapolis, Minnesota, USA for enzyme optimization, library preparation, and ddRADseq / SBG Sequencing-based Genotyping (Baird *et al.*, 2008; Peterson *et al.*, 2012). Enzyme optimization determined that a double-digest of Psti-HF and BtgI was ideal for obtaining the greatest number of SNPs with no cutting in repetitive regions of the genome. The samples were sequenced to 1million reads/sample on a Nextseq 500 using 1X150-bp SR run. Mean quality scores for all libraries were greater than or equal to Q30 and all barcodes were detected. Fastq files were demultiplexed using Illumina bcl2fastq software, and adapter sequences were trimmed from the 5` and 3` ends using cutadapt. Stacks v. 1.46 denovo_map.pl was used to make raw VCF files using the following parameters: ustacks -m 3; ustacks -M 5; cstacks -n 5. The raw VCF file was next filtered using VCFtools, removing variants with minor allele frequency < 1%,

variants with genotype rates < 95%, and samples with genotype rates < 95%(Langmead *et al.*, 2009; Catchen *et al.*, 2013).

Population summary statistics and analysis

The *Stacks* pipeline was used on aligned RAD sequences to identify loci, SNPs and genotypes of individuals from each population, and to create VCF files (Catchen *et al.*, 2013). The VCF files were next analyzed in R using package *vcfR* to visualize and filter data further (Knaus and Grünwald, 2017).

Population summary statistics (Nei's Gst, Hendrik's G'st, Ht, Simpson's Index, and Shannon-Weiner index of MLG diversity) were calculated in *adegenet*, *poppr*, and *vcfR* (Jombart, 2008; Jombart and Ahmed, 2011; Kamvar *et al.*, 2014; Knaus and Grünwald, 2017). Analysis of Molecular Variance (AMOVA) was next utilized in R package *poppr* to assess population differentiation between and among populations, including 1000 permutations to verify significance (Kamvar *et al.*, 2014). As AMOVA does not make assumptions about Hardy-Weinberg equilibrium it can also be used for clonal populations and allow for determinations of genetic similarity.

The SNP dataset for populations of each species was next used to create Minimum spanning networks in the R package *poppr* and composite bar plots of posterior probability assignments from the Discriminant Analysis of Principle Components (DAPC) in the R package *adegenet* (Jombart *et al.*, 2010; Kamvar *et al.*, 2014). The composite bar plots depict the population admixture, the interbreeding of previously isolated genotypes. Populations are defined a priori, allowing one to maximize the variance between populations before determining the degree of admixture between populations for the composite bar plots. In addition, genetic distance was also assessed with dendrograms utilizing UPGMA distance measurements (Cornuet *et al.*, 2008).

RESULTS AND DISCUSSION

SNP Variants for each species

The total number of individuals collected for each species was: 31 *P. pennatifolius*, 21 *P. spicatus*, and 24 *P. riedelianus* (Table 5.1). After filtering we removed two *P. spicatus* individuals from the Rio de Contas, BA population as there was too much missing data / not enough sequencing reads for those samples (Table 5.2). Unfiltered *P. pennatifolius* data included an average of 1,723,255 reads/sample with 383,573 SNPs, after filtering with STACKS we ended up with 9,065 SNPs with 20.87% missing, and finally filtering more rigorously in R for missing data and depth coverage, we ended up with 1,067 SNPs and 1.64% missing data. Unfiltered *P. spicatus* data included an average of 1,723,255 reads/sample with 195,100 SNPs, after filtering with STACKS we ended up with 8,816 SNPs with 22.4% missing, and finally filtering more rigorously in R for missing data and depth coverage, we ended up with 752 high quality SNPs and 3.161% missing data. Unfiltered *P. riedelianus* data included an average of 1,569,779 reads/sample with 118,654 SNPs, after filtering with STACKS we ended up with 8,702 SNPs with 20.7% missing, and finally filtering more rigorously in R for missing data and depth coverage, we ended up with 1,274 high quality SNPs and 2.446% missing data.

Population diversity

To assess the genetic differentiation of the populations we calculated the heterozygosity of each population, the total heterozygosity, Nei's Gst and Hedrick's G'st (Table 5.2). Gst and G'st are widely used in population genomics for multiple alleles and are modifications of Fst, which is commonly used in population genetics for biallelic markers (Wright, 1949, 1965, Nei, 1972, 1973; Hedrick, 2005). In addition, the heterozygosity metric works for polyploid species as well, since it takes into account the

number of alleles observed and not the number of individuals. Both Nei's Gst and Hedrick's G'st have the greatest values for *P. riedelianus,* which implies that this species is genetically more diverse than *P. pennatifolius* and *P. spicatus.* Specifically we see that the heterozygosity of the population in Entre Rios, BA is the greatest, which is further confirmed by MSN analyses (Figure 5.2). *Pilocarpus riedelianus* is most diverse, followed by *P. spicatus,* and lastly *P. pennatifolius* is the least diverse. Heterozygosity calculations were also calculated with the 20% missing 8000 SNP dataset (Supplementary Table 5.1). Both *P. spicatus* and *P. pennatifolius* had higher Hs values for all populations with 20% missing SNPs when compared to analyses of 2%missing 1000 SNP dataset (Table 5.2). This is in contrast to *P. riedelianus* which had only one population with higher Hs and two other populations with lower Hs. A greater heterozygosity value is understandable with a greater number of SNPs, since there is a greater diversity possible. One interesting thing to note is that some populations' Hs levels changed much more than others; this could be due to missing SNPs in certain populations.

From the results of the AMOVA of each species, we see there is some variation in population structure across the three species (Table 5.3). Both *P. spicatus* and *P. riedelianus* have a greater variation between populations (71.46%, 88.53%) than within the populations (28.54%, 11.46%) [p-value < 0.001]. This variation between populations could be due to specialization or adaptation to different regions. Specifically, *P. riedelianus* contains the smallest genetic variation within populations, which is further verified by observations made in the field showing a clustered growth habit. *Pilocarpus riedelianus* was found growing in very dense clusters of many individuals in Linhares, ES and Morro do Chapeu, BA, whereas the population in Entre Rios was much less dense (only five individuals) and individuals were growing in a relatively small area.

Table 5.2. Heterozygosity and genetic differentiation of populations

P. pennatifolius	
Hs A	0.107
Hs B	0.123
Hs C	0.114
Hs D	0.193
Hs E	0.157
Hs F	0.132
Ht	0.186
Gst	0.239
G'st	0.287

P. spicatus	
Hs G	0.027
Hs H	0.023
Hs I	0.125
Hs J	0.125
Ht	0.180
Gst	0.316
G'st	0.349

P. riedelianus	
Hs K	0.186
Hs L	0.125
Hs M	0.146
Ht	0.223
Gst	0.323
G'st	0.401

*Heterozygosity for each population, total heterozygosity, Nei's Gst (1972,1973), Hedrik's G'st(2005)

Table 5.3. Analysis of molecular variance (AMOVA) within / among populations of each species

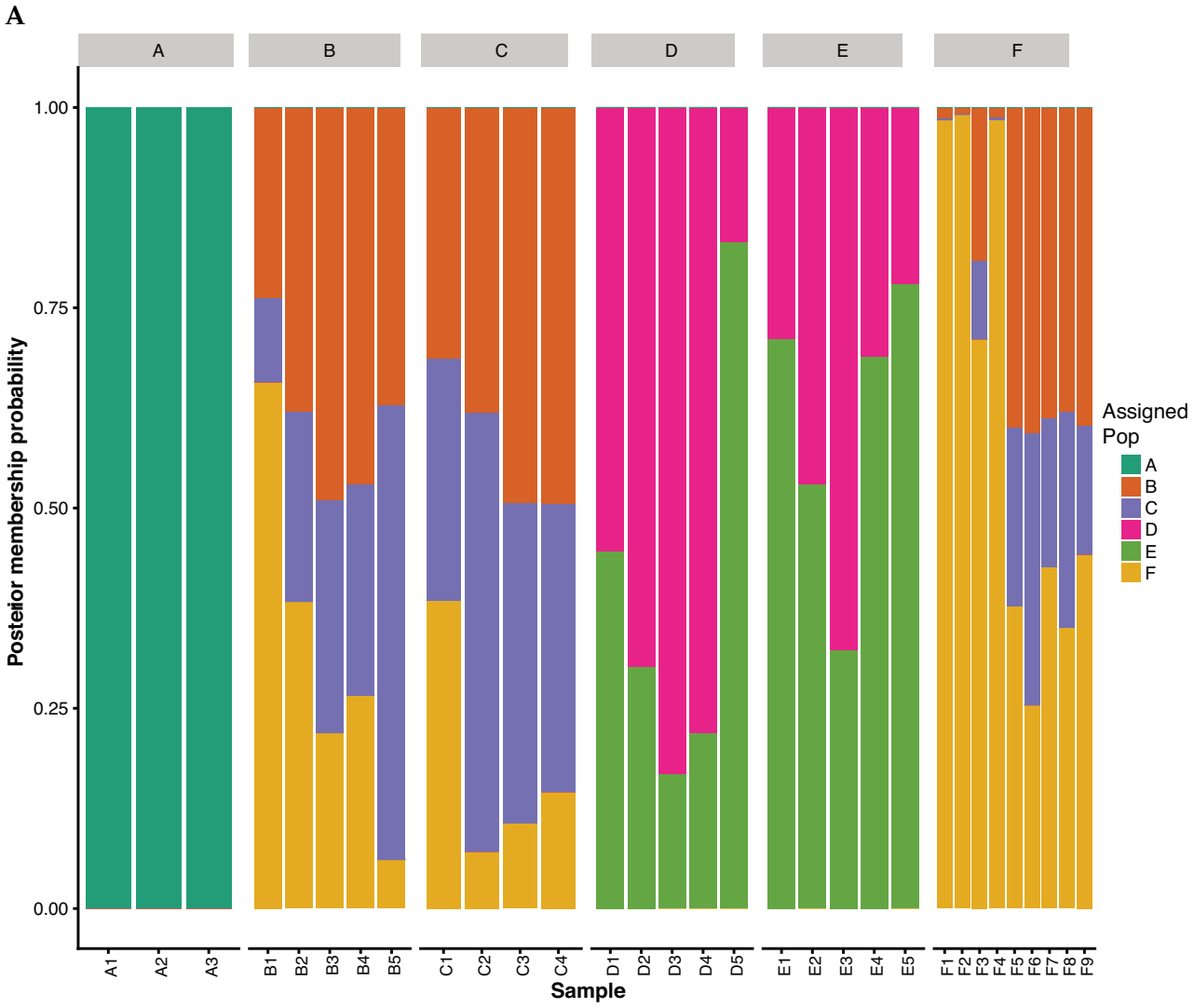
P. pennatifolius	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between populations	5	1863.23	372.65	54.77	36.08	0.361
Within populations	25	2425.81	97.03	97.03	63.92	

P. spicatus	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between populations	3	357.8	119.3	23.22	71.46	0.715
Within populations	15	139.2	9.3	9.27	28.54	

P. riedelianus	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between populations	2	2065.0722	1032.536	131.737	88.53	0.885
Within populations	21	358.344	17.064	17.064	11.46	

*Monte-Carlo significance test with 999 random permutations provided p-value<0.001 for each species; df= degrees of freedom; Sum of Squares; Mean Square; Sigma= degree of differentiation between populations; %CoV= percentage of component variance; ϕ=*F statistics*

The genetic differentiation between the three collection sites is also noticeable in Table 5.2, since Entre Rios, BA had the highest heterozygosity when compared to Linhares, ES and Morro do Chapeu, BA. Although we searched an area 10km away from the first sighting, we were unable to find any other individuals, so for these clusters we picked individuals at the edges of the population.



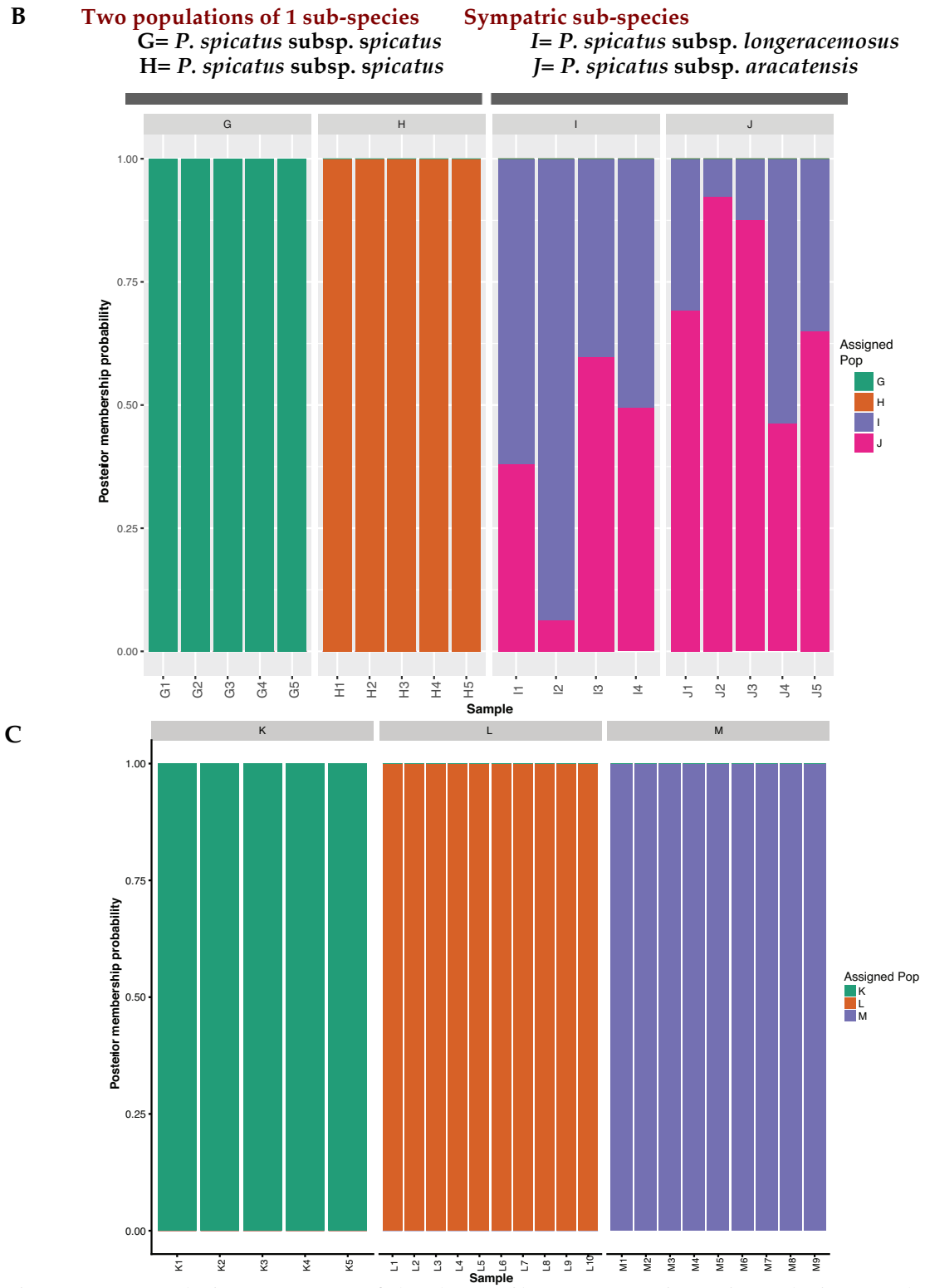


Figure 5.1. Population structure of the three *Pilocarpus* species using admixture plots. The y-axis shows membership probability to each group and x-axis is binned by site/population and each vertical is bar is one individual or MLG (a) *P. pennatifolius* has admixed individuals with membership in 6 groups (b) *P. spicatus* has little to no admixture with 4 distinct groups (c) *P. riedelianus* has no admixture and 3 distinct groups

On the other hand, when we analyze our AMOVA results for the *P. pennatifolius* populations we find that although the populations are significantly differentiated from one another with low variation between populations (36.08%), and the greatest variation is within the population (63.92%) [p<0.001]. This lower diversity between populations in *P. pennatifolius* could be due to human intervention as this particular species is well known as a medicinal plant. Of all the species in the genus *P. pennatifolius* has the most collection records, both historically and currently. In regards to the population differentiation statistic (ϕ), we see *P. riedelianus* has the highest value, implying a higher amount of differentiation between populations, which was also verified when determining Nei's Gst and Hedrick's G'st (Table 5.2,5.3). AMOVA was also calculated using the larger SNP dataset of ~8000 SNPs (Supplementary Table 5.2), and these values correspond to and are not significantly different from AMOVA calculated using higher quality ~1000 SNPs (Table 5.3).

Population structure of *Pilocarpus* species

Admixture composite bar plots

In the admixture plots of our three species, one of the most defining differences we see is the dissimilarity in the plots of *P. pennatifolius* and *P. riedelianus* (Figure 5.1 a, b). It is clear that there is extensive admixture in *P. pennatifolius*, whereas there is no admixture and instead significant structure in *P. riedelianus* populations. The extensive admixture observed in *P. pennatifolius* is likely related to gene flow between the southern populations in the state of Parana or human interference. *Pilocarpus spicatus* on the other hand has two distinct groups with no admixture in the southern populations of *P. spicatus* subsp. *spicatus*, but there is some admixture in the sympatric subspecies of

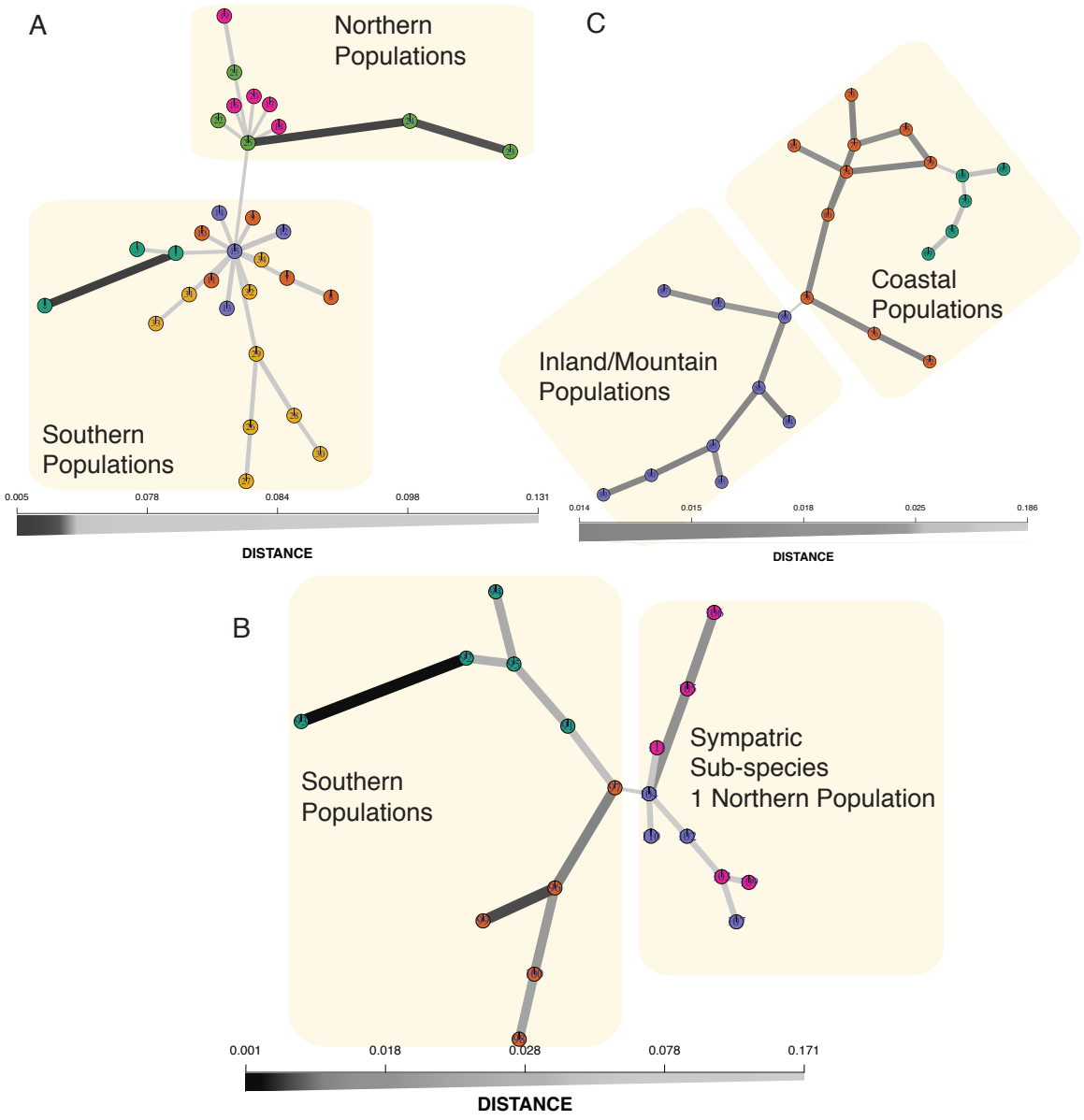


Figure 5.2. Minimum Spanning Network (MSN) of multilocus genotypes (MLGs) with clone corrected individuals of each species. Individuals appear to be mostly clustered by population. The gradient scale bar depicts Nei's genetic distance. A thin light-grey line represents a large genetic distance, whereas a thick and dark line represents closely related MLG's. (a) *P. pennatifolius [top left]* (b) *P. spicatus [bottom]* (c) *P. riedelianus [top right]*. Bi-regional differentiation of genetic distance (the edge between the yellow groups): The greatest genetic distance is between the two regional groups (Coastal vs inland) in *P. riedelianus* (thinnest light grey line), next between northern and southern populations of *P. spicatus*, and finally between northern and southern populations of *P. pennatifolius.*

the north (*P. spicatus* subsp. *aracatensis*, *P. spicatus* subsp. *longeracemosus)*. Although there is evidence of admixture between the sympatric species, it does appear that each

subspecies has a majority membership probability to their own subspecies than to the other sympatric subspecies. PCA plots of SNP data also confirm these groups and can be visualized in Supplemental Figure 5.1, in conjunction with collection localities on the map.

Minimum spanning network diagrams

From the minimum spanning network diagrams we are able to visualize the connections between the populations (Figure 5.2). In particular, the MSN diagram of *P. riedelianus* provides finer population structure when compared to its admixture plot (Figure 5.2c). *P. riedelianus* has significant geographic differentiation between Coastal and Inland/Mountain Populations that is very different from the Northern and Southern differentiation found in *P. spicatus* and *P. pennatifolius.* Even though the two coastal populations are extremely distant geographically, they are more genetically similar when compared to the population that is inland, in the valley of the mountain range of Chapada Diamantina. In fact, the branch connecting to the inland/mountain population is the thinnest and lightest connection, representing a large genetic distance. This great distance is likely due to the mountain being a large physical barrier for dispersal. In addition, when investigating the genetic distance within *P. riedelianus* populations, we observe that the branches of the Entre Rios, BA population (Green) are lighter and thinner; therefore, we have further verification that the population in Entre Rios, BA is more genetically diverse when compared to the other two *P. riedelianus* populations. This difference could possibly be due to dispersal into a new environment, followed by subsequent selection. Overall *P. riedelianus* has the greatest variation among populations, as well as the smallest variation within populations.

In *P. spicatus*, we see a general trend towards North/South regional clustering, with varying scales of geographic structure (Figure 5.2.b). There are two distinct groups

that are more distantly related: one group made up of individuals from Cruz do Pedro, Sao Paulo and Santa Rita, Sao Paulo (green and orange), and the other group is made up of two sympatric subspecies in Rio de Contas, BA (pink and purple). This is intriguing because within these groups we see that the two Sao Paulo populations have a fine-scale geographic structure, and are more differentiated from each other than the sympatric subspecies. This has led to the discussion of whether these two sympatric sub-species are indeed two sub-species. Although they appear to be morphologically distinct in the field, it is not possible to have inter-breeding sub-species. Therefore, we recommend a revision of these sub-species, to maintain accurate taxonomic distinctions. It is also interesting to note that there are some individuals in the Sao Paulo populations that are extremely genetically similar (thicker dark connecting lines).

When examining the MSN of *P. pennatifolius* (Figure 5.2.a) there appears to be a general regional cluster with sites in Sao Paulo (pink and green) together versus sites in Parana (purple, orange, green, yellow). Once again a North / South regional clustering is visualized; though, it is more informative to analyze the UPGMA dendrogram of genetic distance or the admixture plot (Supplementary Figure 5.2, Figure 5.1a). This extensive admixture suggests large gene flow between populations of *P. pennatifolius*. This could be natural; however, it is likely that this is instead the result of human interference. Compared to *P. spicatus* and *P. riedelianus*, *P. pennatifolius* has been very well known as the medicinal plant Jaborandi among the public, and this has increased the opportunity for human influence in its distribution. These genotypes could have been spread through the migrations of indigenous groups or through escape from cultivation in homes, leading to extensive admixture in wild populations.

Limitations in *Pilocarpus* study

As this study involved collections in the field of non-model species that are also in some cases rare, it was difficult to collect as many individuals as we desired to increase sample numbers for analyses. As such, it was not possible to accurately calculate other population summary statistics such as evenness, richness, Shannon-Wiener index (H), Simpson's index (Lambda) or index of association metrics (I_A and $\bar{r}_d$) for the populations in our study (Shannon, 1948; Simpson, 1949; Magurran, 2003). A common limitation in RADseq studies is the percentage of missing data; however, we attempted to reduce this effect by using high quality SNPs (Weigel and Nordborg, 2015; Tabima *et al.*, 2017). Genetic differentiation occurs for a variety of reasons including mutation, migration, genetic drift, and selection. A potential bias with SNPs is that there are varying degrees of mutation rates, with some having very high variability, which can skew measurements of heterozygosity and Gst.

Dispersal of seeds is an extremely important factor that affects the distribution of this species in wild populations. Since the fruit type most common in this group is a dry dehiscent capsule, as the fruit matures and dries out, the seeds are dispersed into the local area. This relatively limited passive dispersal is likely one of the reasons why this species is very commonly found growing in small clusters of trees. Future work is needed to investigate further the dispersal mechanisms, pollinator interactions, and reproductive diversity such as self-fertilization.

Conclusions

This research is the first large-scale SNP analysis of the medicinally important genus *Pilocarpus,* but there is still much to be learned in terms of the demographic history of species in the genus. For this project we also extracted and sequenced a few

herbarium specimens of *P. pennatifolius* and *P. spicatus.* From these preliminary analyses, some admixture relationships were observed; however, future work at a larger scale that assesses genetic diversity in herbarium specimens would be very helpful to analyze the demographic history of species in this genus. Through this research we identified the effects of a physical barrier to dispersal in wild populations of *P. riedelianus,* detected interbreeding in two *P. spicatus* subspecies leading to further taxonomic revision, and finally we revealed the influence of humans on the wild populations of the well known species *P. pennatifolius.* This study and future work developing from this project will be essential to management and conservation of this economically and medicinally important group of plants.

ACKNOWLEDGMENTS

The author(s) thank various individuals and funding institutes for their assistance throughout this project. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1650441). In addition, this work was supported by Harold E. Moore Jr. and Robert T. Clausen Endowment Fund, BSA, SSB, ASPT, Garden Club of America, Einaudi Center for International Studies, Latin American Studies Program of Cornell, and the Cornell Graduate School. We would like to acknowledge Dr. Eloy Rodriguez and Dr. Robert Raguso for thoughtful discussions throughout this project. We would also like to thank Professor Milton Groppo (USP-RP), Professor Pedro Dias (USP-SP), and Professor Marcelo Caxambu (Universidade Tecnologica Federal do Parana), Dr. Rubens Coelho (USP-SP), Alex Popovkin (Bahia, Brazil), Lukas Daneu (CEPEC Herbaria) for helping with the collection of *Pilocarpus* specimens; Iguaçu Itaipu and Parque Estadual Lago

Azul, PR for allowing the collection of specimens from their property for helping with the collection of *Pilocarpus* specimens. Dr. Adrienne Roeder for assistance with Flow Cytometry. Dr. Alexandra Christine Helena Frankland Sawaya for providing the updated extraction protocol for imidazole alkaloids; Instituto Agronomico in Campinas, SP, Brazil for analyzing the soil samples and the statistical consulting group of Cornell University for confirming the statistical analyses. PM thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brasil) for a research fellowship and Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp - grant 2008/58035-6) for research support.

Plants collected in public locations were collected under ICMBio SISBIO permit #52758-1, allowing for the collection of botanical, fungal and microbiological specimens. In addition, plants were collected in collaboration with and under the permits of Professor Milton Groppo, Professor Pedro Dias, and Professor Marcelo Caxambu.

LITERATURE CITED

Agapow P-M, Burt A. 2001. Indices of multilocus linkage disequilibrium. Molecular Ecology Notes **1**, 101–102.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers (JC Fay, Ed.). PLoS ONE **3**, e3376.

Barry CS, McQuinn RP, Thompson AJ, Seymour GB, Grierson D, Giovannoni JJ. 2005. Ethylene insensitivity conferred by the Green-ripe and Never-ripe 2 ripening mutants of tomato. Plant Physiology **138**, 267–275.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology **13**, 969–980.

Brown AHD, Feldman MW, Nevo E. 1980. Multilocus Structure of Natural Populations of Hordeum Spontaneum. Genetics **96**, 523–536.

Caldeira CF, Giannini TC, Ramos SJ, *et al.* 2017. Sustainability of Jaborandi in the eastern Brazilian Amazon. Perspectives in Ecology and Conservation **15**, 161–171.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology **22**, 3124–3140.

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics **24**, 2713–2719.

Davey JW, Blaxter ML. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics **9**, 416–423.

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology **22**, 3151–3164.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics **12**, 499–510.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE **6**, e19379.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences **107**, 16196–16200.

Foll M, Gaggiotti O. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics **180**, 977–993.

Grewe F, Huang J-P, Leavitt SD, Lumbsch HT. 2017. Reference-based RADseq resolves robust relationships among closely related species of lichen-forming fungi using metagenomic DNA. Scientific Reports **7**, 9884.

Hedrick PW. 2005. A Standardized Genetic Differentiation Measure. Evolution **59**, 1633–1638.

IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturals Renovaveis. 1992. *Portaria No. 06N*. Brasilia.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**, 1403–1405.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics **27**, 3070–3071.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**, 94.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**, e281.

Knaus BJ, Grünwald NJ. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources **17**, 44–53.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10**, R25.

Magurran AE. 2003. *Measuring Biological Diversity*. Malden, Ma: Wiley-Blackwell.

Moura EF, Pinto JEBP, Santos JB dos, Lameira OA, Nunes JAR, Ferreira DF. 2005. Genetic diversity of Pilocarpus microphyllus accessions using leaf traits. Crop Breeding and Applied Biotechnology **5**, 348–354.

Nei M. 1972. Genetic Distance between Populations. The American Naturalist **106**, 283–292.

Nei M. 1973. Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences **70**, 3321–3323.

Ng PC, Kirkness EF. 2010. Whole Genome Sequencing. Methods in Molecular Biology. Genetic Variation. Humana Press, Totowa, NJ, 215–226.

Nicotra AB, Atkin OK, Bonser SP, *et al.* 2010. Plant phenotypic plasticity in a changing climate. Trends in Plant Science **15**, 684–692.

Nimpaye E, Kaissi O, Singh TR, Vannier B, Ibrahimi A, Moussa A. 2014. MatSAM: a Matlab implementation for Significance Analysis of Microarrays. International Journal for Computational Biology (IJCB) **3**, 49–51.

Odling-Smee FJ, Laland KN, Feldman MW. 1996. Niche Construction. The American Naturalist **147**, 641–648.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species (L Orlando, Ed.). PLoS ONE **7**, e37135.

Pinheiro CU. 1997. Jaborandi (Pilocarpus sp., rutaceae): A wild species. Economic Botany **51**, 49–58.

Pinheiro CU. 2002. Extrativismo, cultivo e privatização do jaborandi (Pilocarpus microphyllus Stapf ex Holm.; Rutaceae) no Maranhão, Brasil. Acta Botanica Brasilica **16**, 141–150.

Sandhu SS, Abreu IN, Colombo CA, Mazzafera P. 2006. Pilocarpine content and molecular diversity in Jaborandi. Scientia Agricola **63**, 478–482.

Santos AP, Moreno PRH. 2004. Pilocarpus spp.: a survey of its chemical constituents and biological activities. Revista Brasileira de Ciências Farmacêuticas **40**, 116–137.

Shannon CE. 1948. A mathematical theory of communication. Bell System Technical Journal, The **27**, 379–423.

Simpson EH. 1949. Measurement of Diversity. Nature **163**, 688.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Systematic Biology **57**, 758–771.

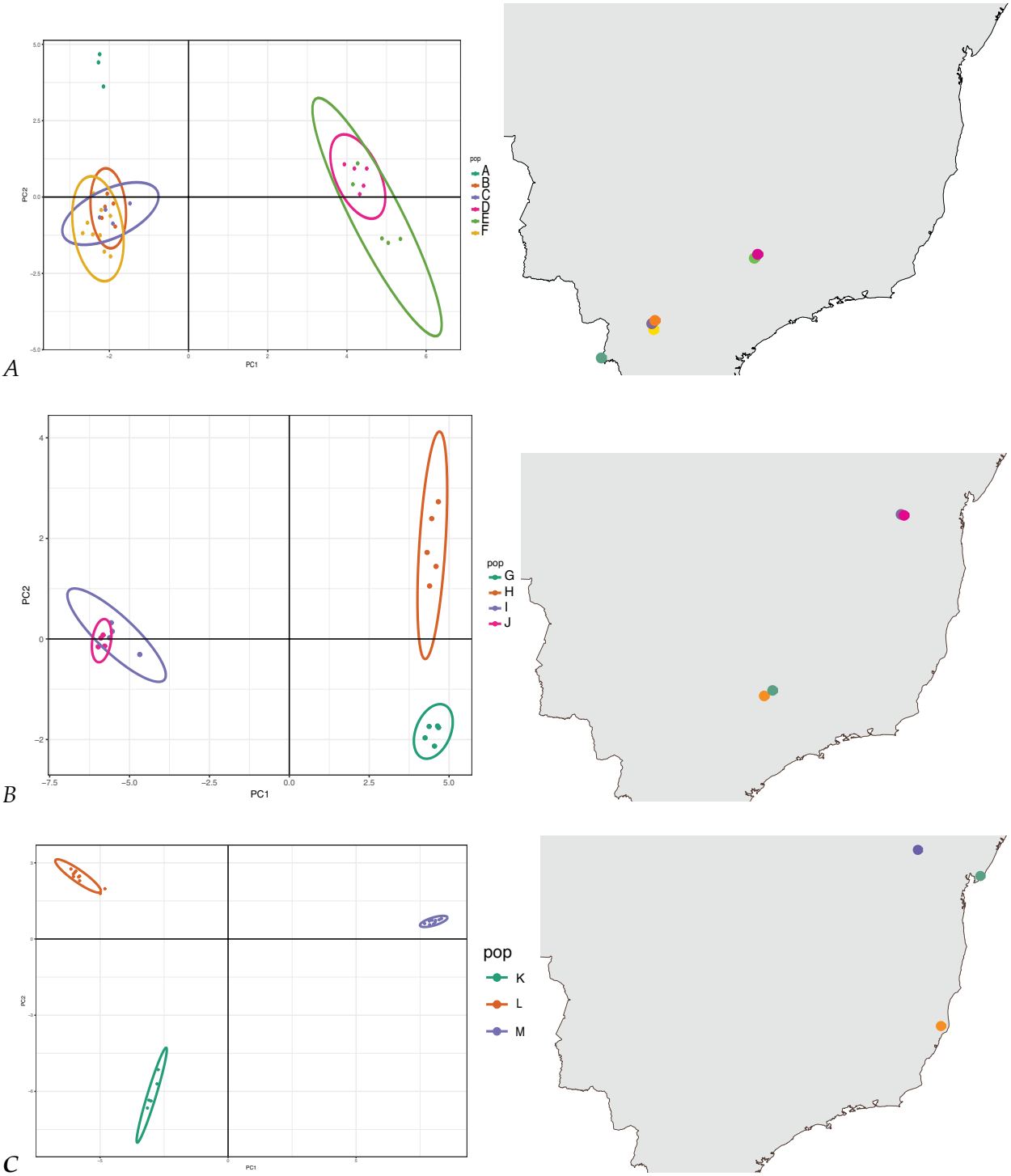
Tabima JF, Kronmiller BA, Press CM, Tyler BM, Zasada IA, Grünwald NJ. 2017. Whole Genome Sequences of the Raspberry and Strawberry Pathogens Phytophthora rubi and P. fragariae. Molecular Plant-Microbe Interactions **30**, 767–769.

Weigel D, Nordborg M. 2015. Population Genomics for Understanding Adaptation in Wild Plant Species. Annual Review of Genetics **49**, 315–338.

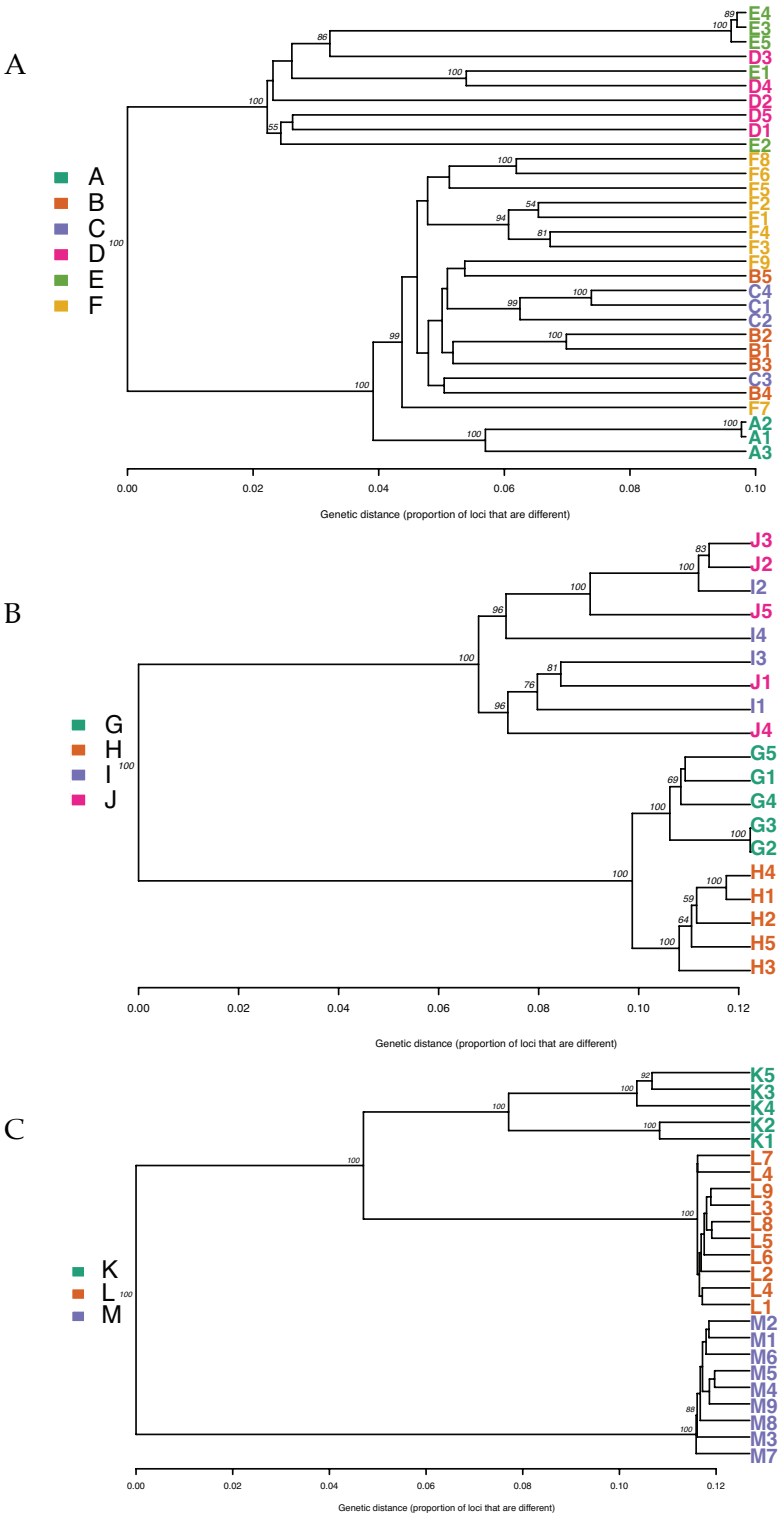
Weinig C, Ewers BE, Welch SM. 2014. Ecological genomics and process modeling of local adaptation to climate. Current Opinion in Plant Biology **18**, 66–72.

Wright S. 1949. The Genetical Structure of Populations. Annals of Eugenics **15**, 323–354.

Wright S. 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution **19**, 395–420.



Supplemental Figure 5.1. Principal components analysis of genetic variation of SNPS for individuals/populations of each species (on the left), color-coded and corresponding to map depicting collection localities for each population (on the right) (a) *P. pennatifolius (b) P. spicatus (c) P. riedelianus*



Supplemental Figure 5.2. Dendrogram of genetic distance using UPGMA distance.
(a) *P. pennatifolius [top]*, (b) *P. spicatus [middle]*, (c) *P. riedelianus [bottom]*

Supplemental Table 5.1.Heterozygosity and population genetic differentiation with ~ 8000 SNPs

P. pennatifolius	
Hs A	0.096
Hs B	0.121
Hs C	0.110
Hs D	0.176
Hs E	0.151
Hs F	0.131
Ht	0.190
Gst	0.272
G'st	0.322

P.spicatus	
Hs G	0.034
Hs H	0.027
Hs I	0.132
Hs J	0.135
Ht	0.185
Gst	0.319
G'st	0.360

P.riedelianus	
Hs K	0.194
Hs L	0.105
Hs M	0.119
Ht	0.244
Gst	0.390
G'st	0.464

*Heterozygosity for each population, total heterozygosity, Nei's Gst (1972,1973), Hedrik's G'st(2005)

Supplemental Table 5.2. Analysis of molecular variance (AMOVA) within/among populations of each species using datasets with 20% missing data approximately~8000 SNPS

P. pennatifolius	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between samples	5	9695.8	1939.2	289	37.34	0.373
Within samples	25	12121.7	484.9	484.9	62.66	

P. spicatus	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between samples	3	4855.0	1618.3	314.9	71.32	0.713
Within samples	18	1899.3	126.62	126.6	28.68	

P. riedelianus	dF	Sum Sq	Mean Sq	Sigma	% CoV	ϕ
Between samples	2	12044.82	6022.4	769.39	89.34	0.893
Within samples	21	1926.13	91.72	91.72	10.65	

*Monte-Carlo significance test with 999 random permutations; df= degrees of freedom; Sum of Squares; Mean Square; Sigma= degree of differentiation between populations; %CoV= percentage of component variance; ϕ=*F statistic*

APPENDICES

APPENDIX 1
Silica and herbarium sheets for *Pilocarpus* dissertation field sites

Collection #	Genus	Species	Location	State
1	*Metrodorea*	*nigra*	Piracicaba	SP
2	*Esenbeckia*	*leiocarpa*	Piracicaba	SP
3	*Helieta*	*sp*	Piracicaba	SP
4	*Pilocarpus*	*pennatifolius*	Maringa	PR
5	*Pilocarpus*	*pennatifolius*	Foz do Iguaçu	PR
6	*Pilocarpus*	*pennatifolius*	Foz do Iguaçu	PR
7	*Pilocarpus*	*pennatifolius*	Foz do Iguaçu	PR
8	*Balfourodendron*	*riedelianum*	Piracicaba	SP
9	*Pilocarpus*	*pauciflorus*	Piracicaba	SP
10	*Pilocarpus*	*pennatifolius*	Piracicaba	SP
11	*Pilocarpus*	*riedelianus*	Entre Rios	BA
12	*Pilocarpus*	*riedelianus*	Entre Rios	BA
13	*Pilocarpus*	*riedelianus*	Entre Rios	BA
14	*Pilocarpus*	*riedelianus*	Entre Rios	BA
15	*Pilocarpus*	*riedelianus*	Entre Rios	BA
16	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
17	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
18	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
19	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
20	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
21	*Pilocarpus*	*grandiflorus*	Entre Rios	BA
22	*Pilocarpus*	*giganteus*	Linhares	ES
23	*Pilocarpus*	*giganteus*	Linhares	ES
24	*Pilocarpus*	*giganteus*	Linhares	ES
25	*Pilocarpus*	*galipea*	Linhares	ES
26	*Pilocarpus*	*riedelianus*	Linhares	ES
27	*Pilocarpus*	*riedelianus*	Linhares	ES
28	*Pilocarpus*	*riedelianus*	Linhares	ES
29	*Pilocarpus*	*riedelianus*	Linhares	ES
30	*Pilocarpus*	*riedelianus*	Linhares	ES
31	*Pilocarpus*	*riedelianus*	Linhares	ES
32	*Pilocarpus*	*riedelianus*	Linhares	ES
33	*Pilocarpus*	*riedelianus*	Linhares	ES
34	*Pilocarpus*	*riedelianus*	Linhares	ES
35	*Pilocarpus*	*riedelianus*	Linhares	ES
36	*Pilocarpus*	*trachyllophus*	Maniacu	BA

37	*Pilocarpus*	*trachyllophus*	Maniacu	BA
38	*Pilocarpus*	*trachyllophus*	Maniacu	BA
39	*Pilocarpus*	*trachyllophus*	Maniacu	BA
40	*Pilocarpus*	*trachyllophus*	Maniacu	BA
41	*Pilocarpus*	*trachyllophus*	Maniacu	BA
42	*Pilocarpus*	*trachyllophus*	Maniacu	BA
43	*Pilocarpus*	*trachyllophus*	Maniacu	BA
44	*Pilocarpus*	*trachyllophus*	Maniacu	BA
45	*Pilocarpus*	*trachyllophus*	Maniacu	BA
46	*Pilocarpus*	*trachyllophus*	Maniacu	BA
47	*Pilocarpus*	*trachyllophus*	Maniacu	BA
48	*Pilocarpus*	*trachyllophus*	Maniacu	BA
49	*Pilocarpus*	*trachyllophus*	Maniacu	BA
50	*Pilocarpus*	*trachyllophus*	Maniacu	BA
51	*Pilocarpus*	*trachyllophus*	Maniacu	BA
52	*Pilocarpus*	*trachyllophus*	Maniacu	BA
53	*Pilocarpus*	*sulcatus*	Maniacu	BA
54	*Pilocarpus*	*sulcatus*	Maniacu	BA
55	*Pilocarpus*	*sulcatus*	Maniacu	BA
56	*Pilocarpus*	*sulcatus*	Maniacu	BA
57	*Pilocarpus*	*sulcatus*	Maniacu	BA
58	*Pilocarpus*	*sulcatus*	Maniacu	BA
59	*Pilocarpus*	*sulcatus*	Maniacu	BA
60	*Pilocarpus*	*sulcatus*	Maniacu	BA
61	*Pilocarpus*	*sulcatus*	Maniacu	BA
62	*Pilocarpus*	*sulcatus*	Maniacu	BA
63	*Pilocarpus*	*sulcatus*	Maniacu	BA
64	*Pilocarpus*	*sulcatus*	Maniacu	BA
65	*Pilocarpus*	*sulcatus*	Maniacu	BA
66	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
67	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
68	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
69	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
70	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
71	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
72	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
73	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
74	*Pilocarpus*	*riedelianus*	Morro do Chapeu	BA
75	*Pilocarpus*	*pennatifolius*	Campo Mourao	PR
76	*Pilocarpus*	*pennatifolius*	Campo Mourao	PR
77	*Pilocarpus*	*pennatifolius*	Campo Mourao	PR
78	*Pilocarpus*	*pennatifolius*	Campo Mourao	PR
79	*Pilocarpus*	*pennatifolius*	Campo Mourao	PR

80	*Pilocarpus*	*pennatifolius*	Campo Mourao B	PR
81	*Pilocarpus*	*pennatifolius*	Campo Mourao B	PR
82	*Pilocarpus*	*pennatifolius*	Campo Mourao B	PR
83	*Pilocarpus*	*pennatifolius*	Campo Mourao B	PR
84	*Pilocarpus*	*spicatus*	Rio de Contas	BA
85	*Pilocarpus*	*spicatus*	Rio de Contas	BA
86	*Pilocarpus*	*spicatus*	Rio de Contas	BA
87	*Pilocarpus*	*spicatus*	Rio de Contas	BA
88	*Pilocarpus*	*spicatus*	Rio de Contas	BA
89	*Pilocarpus*	*spicatus*	Rio de Contas	BA
90	*Pilocarpus*	*spicatus*	Rio de Contas	BA
91	*Pilocarpus*	*spicatus*	Rio de Contas	BA
92	*Pilocarpus*	*spicatus*	Rio de Contas	BA
93	*Pilocarpus*	*spicatus*	Rio de Contas	BA
94	*Pilocarpus*	*spicatus*	Rio de Contas	BA
95	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
96	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
97	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
98	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
99	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
100	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
101	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
102	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
103	*Pilocarpus*	*pennatifolius*	Lago Azul	PR
104	*Pilocarpus*	*pennatifolius*	Campinas	SP
105	*Pilocarpus*	*spicatus*	Cruz do Pedro	SP
106	*Pilocarpus*	*spicatus*	Cruz do Pedro	SP
107	*Pilocarpus*	*spicatus*	Cruz do Pedro	SP
108	*Pilocarpus*	*spicatus*	Cruz do Pedro	SP
109	*Pilocarpus*	*spicatus*	Cruz do Pedro	SP
110	*Pilocarpus*	*pennatifolius*	Cruz do Pedro	SP
111	*Pilocarpus*	*pennatifolius*	Cruz do Pedro	SP
112	*Pilocarpus*	*pennatifolius*	Cruz do Pedro	SP
113	*Pilocarpus*	*pennatifolius*	Cruz do Pedro	SP
114	*Pilocarpus*	*pennatifolius*	Cruz do Pedro	SP
115	*Pilocarpus*	*pennatifolius*	Estacao Ecologica St. Theresa	SP
116	*Pilocarpus*	*pennatifolius*	Estacao Ecologica St. Theresa	SP
117	*Pilocarpus*	*pennatifolius*	Estacao Ecologica St. Theresa	SP
118	*Pilocarpus*	*pennatifolius*	Estacao Ecologica St. Theresa	SP
119	*Pilocarpus*	*pennatifolius*	Estacao Ecologica St. Theresa	SP

120	*Pilocarpus*	*spicatus*	Santa Rita	SP
121	*Pilocarpus*	*spicatus*	Santa Rita	SP
122	*Pilocarpus*	*spicatus*	Santa Rita	SP
123	*Pilocarpus*	*spicatus*	Santa Rita	SP
124	*Pilocarpus*	*spicatus*	Santa Rita	SP
125	*Pilocarpus*	*microphyllus*	Campinas	SP
126	*Pilocarpus*	*spicatus*	Barra do Corda	MA
127	*Pilocarpus*	*jaborandi*	Barra do Corda	MA
128	*Pilocarpus*	*racemosus*	Barra do Corda	MA
129	*Pilocarpus*	*carajaensis*	Carajas	PA
130	*Pilocarpus*	*trachyllophus*	Barra do Corda	MA
131	*Pilocarpus*	*microphyllus*	Piaui	PI
132	*Pilocarpus*	*microphyllus*	Para	PA
133	*Pilocarpus*	*microphyllus*	Tocantins	TO
134	*Pilocarpus*	*microphyllus*	Barra do Corda	MA
135	*Pilocarpus*	*pennatifolius*	Sao Paulo	SP
136	*Pilocarpus*	*pennatifolius*	Campinas	SP

APPENDIX 2
Discrete and continuous chemical traits reconstructed on the phylogeny

Taxon	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14
P. pennatifolius	1	1	1	1	1	0	0	1	0	0	0	0	1	0
P. jaborandi	1	0	1	1	0	0	0	0	0	0	0	0	0	0
P. carajaensis	1	0	0	1	0	0	0	0	1	1	0	1	1	1
P. trachyllophus	1	1	0	1	0	0	0	0	0	0	0	1	1	0
P. microphyllus	1	1	0	1	1	1	1	1	1	1	1	1	1	1
P. pauciflorus	1	1	1	1	0	0	0	0	0	0	0	0	0	0
P. giganteus	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P. grandiflorus	1	1	1	1	0	0	0	0	0	0	0	0	0	0
P. sulcatus	1	1	1	1	1	0	0	0	0	0	0	0	0	0
P. racemosus	1	1	0	1	0	0	0	0	0	0	0	0	0	1
P. riedelianus	0	1	0	1	0	0	0	0	0	0	0	0	0	0
P. spicatus	1	1	0	0	0	0	0	0	0	0	0	0	1	0

Taxon	a15	a16	a17	a18	a19	a20	a21	a22	Adiv	Pilocarpine
P. pennatifolius	0	1	0	1	1	1	0	0	11	1.2444
P. jaborandi	0	0	0	0	0	0	0	0	3	0.0040
P. carajaensis	0	0	1	0	0	0	1	1	10	1.2300
P. trachyllophus	0	0	0	0	0	0	0	0	5	2.5098
P. microphyllus	1	1	1	0	0	1	1	0	18	1.3692
P. pauciflorus	0	0	0	0	0	0	0	0	4	0.0884
P. giganteus	0	0	0	0	0	0	0	0	1	0.0000
P. grandiflorus	0	0	0	0	0	0	0	0	4	2.7065
P. sulcatus	0	0	0	0	0	0	0	0	5	6.8750
P. racemosus	0	0	0	0	0	0	0	0	4	0.1149
P. riedelianus	0	0	0	0	0	0	0	0	2	0.0214
P. spicatus	0	0	1	0	0	0	0	0	4	0.0000

Taxon	c1	c2	c3	c4	c5	c6	c7	c8	Cdiv
P. pennatifolius	1	1	1	1	1	1	1	1	8
P. jaborandi	1	1	0	1	0	1	1	1	6
P. carajaensis	1	0	0	0	0	1	0	1	3
P. trachyllophus	1	1	1	1	1	1	1	1	8
P. microphyllus	1	1	0	1	0	1	0	1	5
P. pauciflorus	1	1	1	1	0	1	0	1	6
P. giganteus	1	1	1	1	1	1	1	1	8
P. grandiflorus	1	1	1	1	1	1	1	1	8
P. sulcatus	1	1	1	1	1	1	1	1	8
P. racemosus	1	1	0	0	0	1	0	0	3
P. riedelianus	1	1	1	1	1	1	0	1	7
P. spicatus	1	1	0	1	0	1	1	1	6

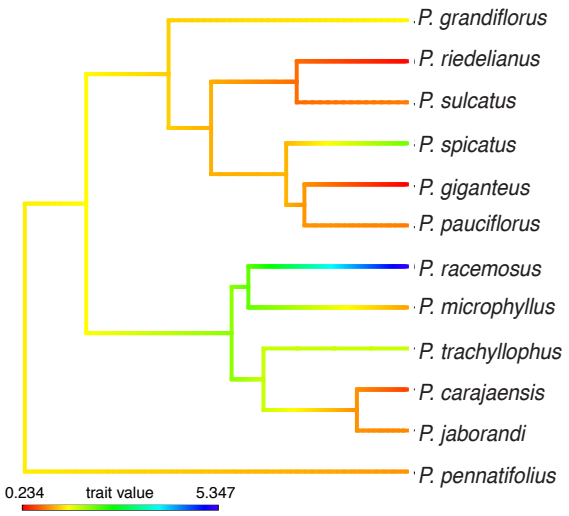
Taxon	cc1	cc2	cc3	cc4	cc5	cc6	cc7	cc8
P. pennatifolius	1.5646	14.155	0.868	49.982	0.4834	49.98	0.7839	142.45
P. jaborandi	0.8788	0.7298	0	11.777	0	14.41	0.2773	0.8022
P. carajaensis	0.5587	0	0	0	0	0.786	0	0
P. trachyllophus	1.8034	0.4539	1.121	12.549	1.3819	29.69	2.5229	16.852
P. microphyllus	0.2635	0.0861	0	0.3036	0	1.555	0	1.0424
P. pauciflorus	0.7795	0.3151	0.920	4.0284	0	27.07	0	4.9056
P. giganteus	0.2739	0.2502	0.920	0.5818	0.6564	17.34	0.3726	1.8790
P. grandiflorus	1.0321	1.4472	0.656	0.4587	0.4886	0.813	0.3696	1.3052
P. sulcatus	1.0046	0.8989	1.052	49.982	2.0365	49.98	0.2729	43.581
P. racemosus	5.3454	0.4739	0	0	0	2.04	0	0
P. riedelianus	0.2432	0.5445	0.623	0.3569	0.4737	9.436	0	1.6074
P. spicatus	1.3670	0.5296	0	15.777	0	34.83	0.7407	10.468

*a#=Presence/Absence Alkaloid
*c#= Presence/Absence Coumarin
*Adiv= Alkaloid diversity/ Total number of alkaloids
*Cdiv= Coumarin diversity/ Total number of coumarins
*Pilocarpine= pilocarpine concentration ug/mL
*cc#= Concentration of coumarins in ug/mL

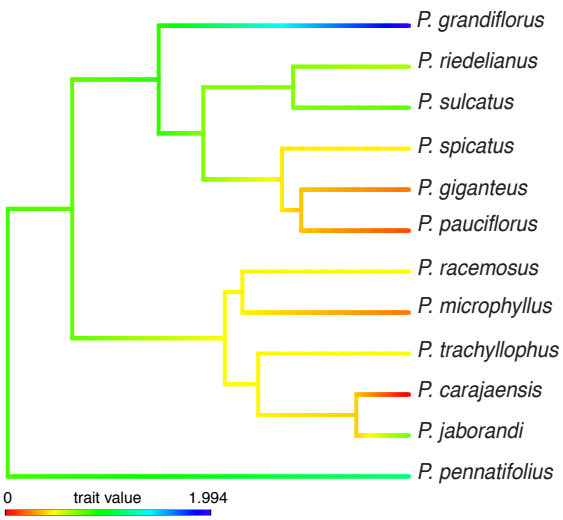
APPENDIX 3
Reconstruction of chemical traits across *Pilocarpus*

Ancestral trait reconstruction of the concentration of coumarin compounds. This trait was analyzed as a quantitative variable using an Mk model. The gradient depicts the concentration for each taxon; red (small concentration) to blue (large concentration).

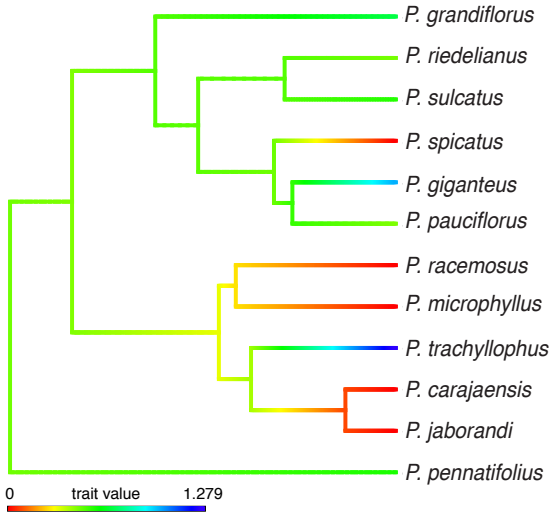
CC1: Scopoletin



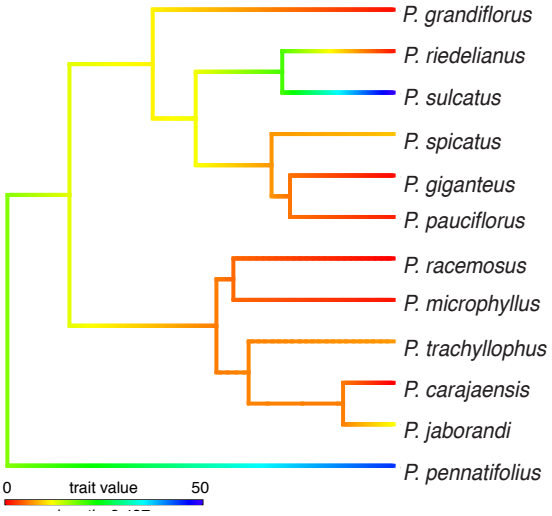
CC2: Psoralen



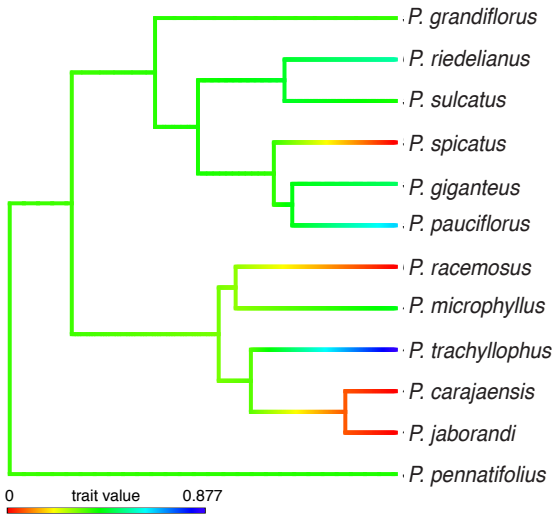
CC3: Coumarin



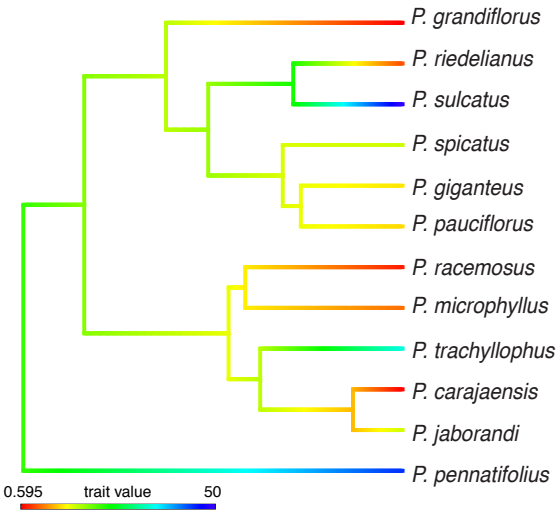
CC4: Xanthotoxin



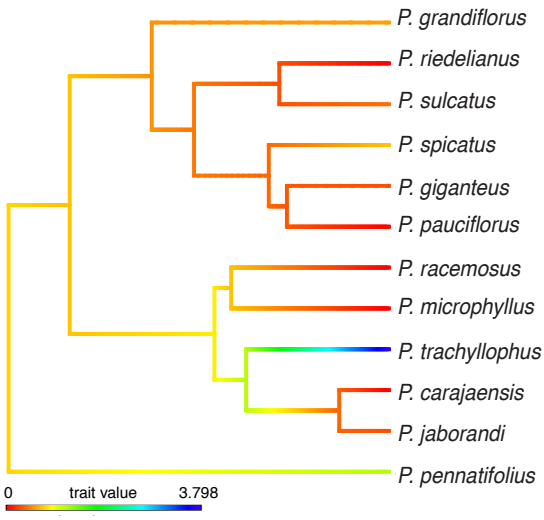
CC5: Citropten



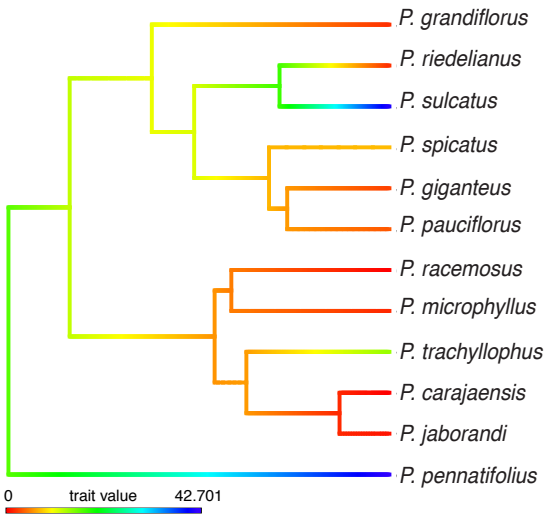
CC6: Imperatorin B



CC7: Osthol



CC8: Imperatorin



APPENDIX 4
INCT and GBIF Herbarium Specimens

REF #	Species	LON	LAT	REF #	Species	LON	LAT
4	*alatus*	-44.97	-4.43	3171	*pennatifolius*	-56.28	-25.83
107	*alatus*	-44.97	-4.43	3172	*pennatifolius*	-56.27	-25.72
627	*alatus*	-45.24	-5.51	3177	*pennatifolius*	-54.63	-25.38
829	*alatus*	-44.93	-4.40	3178	*pennatifolius*	-56.85	-26.08
1288	*alatus*	-44.92	-4.40	3180	*pennatifolius*	-56.25	-25.95
5	*carajaensis*	-50.62	-5.95	3181	*pennatifolius*	-57.63	-25.25
44	*carajaensis*	-53.09	-10.77	3182	*pennatifolius*	-55.47	-26.15
75	*carajaensis*	-49.12	-5.37	3183	*pennatifolius*	-55.78	-26.02
76	*carajaensis*	-49.67	-3.77	3185	*pennatifolius*	-54.83	-27.11
77	*carajaensis*	-50.62	-5.95	3194	*pennatifolius*	-57.08	-25.92
108	*carajaensis*	-50.29	-6.02	3195	*pennatifolius*	-56.03	-26.45
109	*carajaensis*	-50.26	-6.08	3196	*pennatifolius*	-57.13	-25.63
110	*carajaensis*	-50.55	-5.80	3198	*pennatifolius*	-54.67	-25.63
112	*carajaensis*	-50.16	-6.07	3199	*pennatifolius*	-55.98	-22.67
113	*carajaensis*	-52.64	-4.61	3200	*pennatifolius*	-54.28	-24.03
114	*carajaensis*	-50.06	-6.06	3204	*pennatifolius*	-56.45	-24.67
628	*carajaensis*	-45.24	-5.51	3205	*pennatifolius*	-56.02	-25.45
763	*carajaensis*	-52.87	-10.83	3206	*pennatifolius*	-56.52	-24.63
830	*carajaensis*	-49.90	-6.07	3207	*pennatifolius*	-57.35	-25.32
831	*carajaensis*	-50.62	-5.95	3208	*pennatifolius*	-54.35	-24.46
1439	*carajaensis*	-51.52	-6.13	3211	*pennatifolius*	-55.92	-22.65
1481	*carajaensis*	-49.90	-6.07	3213	*pennatifolius*	-57.05	-23.37
1943	*carajaensis*	-50.29	-6.02	3214	*pennatifolius*	-55.57	-23.27
1944	*carajaensis*	-50.26	-6.08	3215	*pennatifolius*	-56.95	-25.67
1945	*carajaensis*	-49.67	-3.77	3232	*pennatifolius*	-56.25	-25.83
1946	*carajaensis*	-50.16	-6.07	3234	*pennatifolius*	-57.42	-27.05
2236	*carajaensis*	-51.52	-6.13	3251	*pennatifolius*	-58.17	-23.20
2341	*carajaensis*	-50.62	-5.95	3276	*pennatifolius*	-64.92	-35.39
2345	*carajaensis*	-52.64	-4.61	3400	*pennatifolius*	-56.35	-25.72
2394	*carajaensis*	-50.06	-6.06	18	*peruvianus*	-67.62	-9.75
3219	*carajaensis*	-53.09	-10.77	48	*peruvianus*	-53.09	-10.77
78	*demerarae*	-44.97	-4.43	169	*peruvianus*	-66.84	-9.99
2079	*demerarae*	-44.97	-4.43	170	*peruvianus*	-68.60	-9.41
2404	*demerarae*	-59.04	6.05	171	*peruvianus*	-68.73	-9.20
58	*giganteus*	-39.10	-12.67	172	*peruvianus*	-62.93	-10.12
99	*giganteus*	-43.21	-22.90	173	*peruvianus*	-67.62	-9.75
115	*giganteus*	-40.07	-19.39	798	*peruvianus*	-52.87	-10.83
116	*giganteus*	-40.07	-19.39	656	*racemosus*	-45.24	-5.51

117	*giganteus*	-39.29	-15.24
118	*giganteus*	-43.18	-22.51
277	*giganteus*	-46.99	-24.32
344	*giganteus*	-39.08	-15.29
346	*giganteus*	-39.50	-15.19
457	*giganteus*	-42.14	-19.79
832	*giganteus*	-46.42	-23.85
833	*giganteus*	-46.43	-23.90
929	*giganteus*	-43.94	-19.92
965	*giganteus*	-46.33	-23.96
1085	*giganteus*	-42.78	-19.84
1265	*giganteus*	-43.45	-22.76
1290	*giganteus*	-46.40	-24.00
1535	*giganteus*	-42.30	-22.55
2386	*giganteus*	-42.30	-22.55
2700	*giganteus*	-40.07	-19.39
2708	*giganteus*	-40.07	-19.39
2709	*giganteus*	-39.29	-15.24
2739	*giganteus*	-43.18	-22.51
2845	*giganteus*	-43.05	-22.53
2999	*giganteus*	-46.42	-23.85
3002	*giganteus*	-46.43	-23.90
6	*grandiflorus*	-39.04	-14.26
7	*grandiflorus*	-39.10	-14.49
80	*grandiflorus*	-38.88	-15.86
119	*grandiflorus*	-38.95	-13.99
120	*grandiflorus*	-40.12	-19.37
122	*grandiflorus*	-39.04	-14.26
123	*grandiflorus*	-39.08	-15.29
124	*grandiflorus*	-39.70	-14.73
125	*grandiflorus*	-39.74	-14.57
126	*grandiflorus*	-40.07	-19.39
127	*grandiflorus*	-39.13	-14.49
128	*grandiflorus*	-39.10	-14.49
129	*grandiflorus*	-38.95	-13.99
130	*grandiflorus*	-39.96	-19.14
131	*grandiflorus*	-39.01	-14.10
349	*grandiflorus*	-38.93	-13.37
350	*grandiflorus*	-39.08	-14.42
351	*grandiflorus*	-39.05	-14.79
352	*grandiflorus*	-39.01	-14.10
353	*grandiflorus*	-39.10	-14.49
355	*grandiflorus*	-40.07	-19.39
1948	*racemosus*	-85.50	10.96
1949	*racemosus*	-89.98	13.82
1964	*racemosus*	-52.15	4.54
1970	*racemosus*	-85.34	10.76
1979	*racemosus*	-95.01	16.67
1985	*racemosus*	-85.35	10.79
1986	*racemosus*	-90.03	19.53
2001	*racemosus*	-93.17	16.81
2003	*racemosus*	-95.01	16.68
2007	*racemosus*	-85.73	10.90
2031	*racemosus*	-79.83	22.37
2032	*racemosus*	-86.84	21.05
2033	*racemosus*	-88.02	19.58
2035	*racemosus*	-87.85	19.69
2036	*racemosus*	-89.93	13.82
2074	*racemosus*	-89.74	18.52
2077	*racemosus*	-75.10	9.94
2078	*racemosus*	-85.25	10.75
2081	*racemosus*	-85.49	10.96
2118	*racemosus*	-75.17	9.97
2146	*racemosus*	-62.97	7.63
2147	*racemosus*	-85.72	11.18
2170	*racemosus*	-85.56	10.96
2182	*racemosus*	-84.12	14.35
2229	*racemosus*	-85.46	10.92
2230	*racemosus*	-85.49	10.96
2231	*racemosus*	-85.35	10.77
2250	*racemosus*	-85.48	10.95
2252	*racemosus*	-85.48	10.93
2265	*racemosus*	-85.35	10.78
2271	*racemosus*	-95.52	15.98
2278	*racemosus*	-85.47	10.93
2348	*racemosus*	-85.47	10.67
2349	*racemosus*	-85.48	9.93
2361	*racemosus*	-85.48	10.67
2366	*racemosus*	-85.34	10.75
2412	*racemosus*	-74.16	20.24
2427	*racemosus*	-62.97	7.63
2431	*racemosus*	-89.93	13.82
2432	*racemosus*	-90.12	19.19
2436	*racemosus*	-89.98	13.82
2437	*racemosus*	-89.39	18.74
2438	*racemosus*	-95.32	16.35

460	*grandiflorus*	-39.64	-14.71
743	*grandiflorus*	-40.27	-19.82
746	*grandiflorus*	-39.91	-19.19
766	*grandiflorus*	-52.87	-10.83
834	*grandiflorus*	-39.12	-14.92
835	*grandiflorus*	-37.19	-9.52
910	*grandiflorus*	-40.94	-21.21
1343	*grandiflorus*	-40.27	-19.82
1344	*grandiflorus*	-39.74	-18.42
1777	*grandiflorus*	-37.97	-12.17
1953	*grandiflorus*	-38.93	-13.37
2206	*grandiflorus*	-39.08	-14.42
2228	*grandiflorus*	-40.27	-19.82
2339	*grandiflorus*	-37.97	-12.17
2531	*grandiflorus*	-39.74	-18.42
2704	*grandiflorus*	-39.08	-15.30
2705	*grandiflorus*	-39.74	-14.57
2706	*grandiflorus*	-40.07	-19.39
2707	*grandiflorus*	-39.13	-14.49
2756	*grandiflorus*	-40.94	-21.21
2891	*grandiflorus*	-39.96	-19.14
3074	*grandiflorus*	-39.01	-14.11
3132	*grandiflorus*	-39.91	-19.19
59	*jaborandi*	-41.16	-11.55
61	*jaborandi*	-40.47	-11.86
64	*jaborandi*	-38.19	-8.38
629	*jaborandi*	-40.99	-3.73
630	*jaborandi*	-41.09	-3.56
633	*jaborandi*	-40.55	-3.59
767	*jaborandi*	-52.87	-10.83
836	*jaborandi*	-45.33	-5.67
837	*jaborandi*	-40.77	-3.87
838	*jaborandi*	-40.12	-3.98
1070	*jaborandi*	-40.55	-3.59
1478	*jaborandi*	-43.14	-11.09
1479	*jaborandi*	-40.05	-7.21
1702	*jaborandi*	-42.70	-9.02
1888	*jaborandi*	-45.24	-5.51
2134	*jaborandi*	-77.01	38.89
2302	*jaborandi*	-40.55	-3.59
2897	*jaborandi*	-45.33	-5.67
3154	*jaborandi*	-38.19	-8.38
8	*manuensis*	-72.57	-8.93
2439	*racemosus*	-89.86	19.17
2441	*racemosus*	-89.51	17.91
2442	*racemosus*	-87.98	19.49
2443	*racemosus*	-87.87	19.68
2455	*racemosus*	-88.10	19.59
2457	*racemosus*	-86.85	21.04
2460	*racemosus*	-86.87	20.94
2464	*racemosus*	-69.23	19.36
2478	*racemosus*	-85.50	10.96
2479	*racemosus*	-85.47	10.93
2480	*racemosus*	-89.35	14.42
2494	*racemosus*	-62.63	8.05
2523	*racemosus*	-85.47	10.93
2535	*racemosus*	-84.88	10.42
2536	*racemosus*	-85.28	10.75
2537	*racemosus*	-85.56	10.97
2539	*racemosus*	-85.24	10.78
2577	*racemosus*	-66.92	18.37
2610	*racemosus*	-89.39	18.74
2611	*racemosus*	-88.00	19.86
2788	*racemosus*	-89.26	21.30
2791	*racemosus*	-85.49	10.96
2795	*racemosus*	-85.33	10.76
2798	*racemosus*	-85.46	11.00
2801	*racemosus*	-85.46	10.99
2843	*racemosus*	-87.98	19.61
2868	*racemosus*	-87.57	20.33
2869	*racemosus*	-88.25	20.65
2977	*racemosus*	-75.10	9.92
3062	*racemosus*	-65.44	18.15
3064	*racemosus*	-65.42	18.13
3065	*racemosus*	-65.48	18.13
3103	*racemosus*	-86.88	20.82
3159	*racemosus*	-61.36	15.44
3160	*racemosus*	-88.19	18.21
3223	*racemosus*	-79.04	21.62
3225	*racemosus*	-53.24	3.92
3226	*racemosus*	-88.97	20.66
3239	*racemosus*	-65.91	7.08
3294	*racemosus*	-52.97	3.86
3345	*racemosus*	-85.48	10.89
3346	*racemosus*	-85.58	10.96
3403	*racemosus*	-83.95	10.01

9	*manuensis*	-72.71	-8.97
132	*manuensis*	-72.71	-8.97
133	*manuensis*	-72.57	-8.93
2311	*manuensis*	-71.27	-11.93
3049	*manuensis*	-71.90	-9.38
10	*microphyllus*	-49.12	-5.37
11	*microphyllus*	-47.42	-7.13
12	*microphyllus*	-49.25	-5.58
14	*microphyllus*	-49.65	-3.75
86	*microphyllus*	-49.67	-3.77
88	*microphyllus*	-45.66	-3.96
89	*microphyllus*	-49.12	-5.37
90	*microphyllus*	-47.56	-22.41
100	*microphyllus*	-47.06	-22.90
101	*microphyllus*	-48.50	-1.46
136	*microphyllus*	-45.66	-3.96
137	*microphyllus*	-47.42	-7.13
138	*microphyllus*	-44.30	-2.53
139	*microphyllus*	-44.20	-2.58
140	*microphyllus*	-39.05	-14.79
356	*microphyllus*	-39.05	-14.79
448	*microphyllus*	-44.30	-2.53
449	*microphyllus*	-47.49	-5.53
638	*microphyllus*	-45.24	-5.51
639	*microphyllus*	-45.38	-3.67
641	*microphyllus*	-42.58	-4.76
642	*microphyllus*	-42.55	-3.52
649	*microphyllus*	-43.10	-3.52
650	*microphyllus*	-39.72	-7.51
748	*microphyllus*	-40.07	-19.39
772	*microphyllus*	-52.87	-10.83
839	*microphyllus*	-48.77	-1.90
840	*microphyllus*	-47.95	-5.02
842	*microphyllus*	-47.97	-4.55
846	*microphyllus*	-47.95	-4.55
847	*microphyllus*	-44.27	-2.53
848	*microphyllus*	-43.92	-3.47
849	*microphyllus*	-42.55	-3.52
850	*microphyllus*	-45.98	-4.63
852	*microphyllus*	-46.23	-4.45
853	*microphyllus*	-47.72	-4.90
855	*microphyllus*	-45.33	-5.67
856	*microphyllus*	-48.43	-6.22
3104	*riedelianus*	-42.51	-14.73
3105	*riedelianus*	-43.89	-19.63
3110	*riedelianus*	-56.43	-21.19
3111	*riedelianus*	-43.67	-19.33
3113	*riedelianus*	-47.93	-15.78
3115	*riedelianus*	-44.73	-14.25
19	*riedelianus*	-39.00	-14.95
20	*riedelianus*	-39.40	-16.85
21	*riedelianus*	-38.95	-13.99
22	*riedelianus*	-39.52	-15.17
23	*riedelianus*	-39.08	-14.33
24	*riedelianus*	-40.77	-20.87
25	*riedelianus*	-39.08	-14.39
26	*riedelianus*	-38.99	-16.11
27	*riedelianus*	-39.37	-14.33
41	*riedelianus*	-39.05	-14.79
92	*riedelianus*	-39.08	-15.29
174	*riedelianus*	-40.77	-20.87
176	*riedelianus*	-39.08	-14.39
177	*riedelianus*	-38.99	-14.28
178	*riedelianus*	-39.52	-15.17
179	*riedelianus*	-39.53	-15.17
180	*riedelianus*	-39.06	-14.42
181	*riedelianus*	-39.05	-14.98
182	*riedelianus*	-39.08	-14.33
183	*riedelianus*	-39.52	-15.16
184	*riedelianus*	-39.52	-15.17
186	*riedelianus*	-39.10	-14.48
187	*riedelianus*	-40.43	-13.44
189	*riedelianus*	-40.77	-20.87
190	*riedelianus*	-39.30	-14.95
191	*riedelianus*	-39.08	-15.15
192	*riedelianus*	-39.05	-14.79
193	*riedelianus*	-39.06	-16.45
194	*riedelianus*	-39.03	-14.31
195	*riedelianus*	-38.95	-13.99
196	*riedelianus*	-39.38	-14.35
197	*riedelianus*	-39.96	-19.14
198	*riedelianus*	-39.41	-16.86
199	*riedelianus*	-35.87	-9.21
200	*riedelianus*	-38.99	-16.11
201	*riedelianus*	-40.07	-19.39
364	*riedelianus*	-39.60	-14.71

857	*microphyllus*	-50.93	-6.05
858	*microphyllus*	-41.78	-3.01
1002	*microphyllus*	-45.28	-5.28
1067	*microphyllus*	-46.67	-4.55
1480	*microphyllus*	-43.12	-3.67
1508	*microphyllus*	-42.37	-3.46
1510	*microphyllus*	-42.56	-3.72
1511	*microphyllus*	-43.11	-3.63
1512	*microphyllus*	-41.78	-2.90
1513	*microphyllus*	-42.45	-3.75
1514	*microphyllus*	-42.76	-4.01
1704	*microphyllus*	-43.92	-3.54
1740	*microphyllus*	-51.23	-30.03
1950	*microphyllus*	-56.93	4.52
2190	*microphyllus*	-45.28	-5.28
2291	*microphyllus*	-45.98	-4.63
2338	*microphyllus*	-46.67	-4.55
2881	*microphyllus*	-46.23	-4.45
2894	*microphyllus*	-45.33	-5.67
2908	*microphyllus*	-48.43	-6.22
3024	*microphyllus*	-41.78	-3.01
3068	*microphyllus*	-45.66	-3.97
3070	*microphyllus*	-44.30	-2.53
3071	*microphyllus*	-44.20	-2.58
3075	*microphyllus*	-39.05	-14.79
3393	*microphyllus*	-54.08	5.05
1693	*microphylus*	-43.35	-21.76
15	*pauciflorus*	-40.38	-12.25
16	*pauciflorus*	-39.08	-15.15
45	*pauciflorus*	-53.09	-10.77
46	*pauciflorus*	-51.16	-27.03
47	*pauciflorus*	-42.91	-22.48
141	*pauciflorus*	-39.08	-12.53
142	*pauciflorus*	-48.12	-22.28
143	*pauciflorus*	-39.08	-15.15
144	*pauciflorus*	-48.44	-22.88
145	*pauciflorus*	-47.55	-24.70
147	*pauciflorus*	-47.00	-23.39
148	*pauciflorus*	-46.63	-23.54
149	*pauciflorus*	-50.74	-23.73
150	*pauciflorus*	-40.35	-15.48
151	*pauciflorus*	-48.51	-25.52
152	*pauciflorus*	-48.33	-25.31
365	*riedelianus*	-38.98	-14.30
366	*riedelianus*	-39.12	-14.50
367	*riedelianus*	-39.08	-14.42
369	*riedelianus*	-39.30	-14.96
371	*riedelianus*	-38.95	-15.68
373	*riedelianus*	-38.52	-15.18
374	*riedelianus*	-41.51	-17.86
375	*riedelianus*	-39.50	-15.19
377	*riedelianus*	-39.08	-14.33
378	*riedelianus*	-39.00	-14.28
380	*riedelianus*	-39.08	-15.15
381	*riedelianus*	-39.07	-14.78
389	*riedelianus*	-39.52	-15.15
390	*riedelianus*	-40.11	-13.53
391	*riedelianus*	-38.95	-13.99
392	*riedelianus*	-39.37	-14.33
393	*riedelianus*	-39.37	-17.89
394	*riedelianus*	-34.81	-7.47
550	*riedelianus*	-39.97	-13.53
552	*riedelianus*	-40.07	-19.39
553	*riedelianus*	-42.66	-14.77
595	*riedelianus*	-41.02	-12.45
596	*riedelianus*	-41.09	-13.53
597	*riedelianus*	-40.82	-12.77
749	*riedelianus*	-40.33	-19.27
799	*riedelianus*	-52.87	-10.83
871	*riedelianus*	-39.52	-15.17
873	*riedelianus*	-40.11	-13.53
874	*riedelianus*	-39.08	-16.45
914	*riedelianus*	-40.80	-19.89
1007	*riedelianus*	-38.50	-12.97
1008	*riedelianus*	-39.47	-12.85
1009	*riedelianus*	-39.52	-12.93
1010	*riedelianus*	-39.10	-12.67
1011	*riedelianus*	-40.47	-11.85
1014	*riedelianus*	-39.60	-13.02
1018	*riedelianus*	-39.71	-14.18
1020	*riedelianus*	-42.68	-14.77
1071	*riedelianus*	-37.35	-11.10
1261	*riedelianus*	-40.17	-11.96
1347	*riedelianus*	-40.29	-20.33
1348	*riedelianus*	-39.69	-19.29
1350	*riedelianus*	-40.27	-20.27

153	*pauciflorus*	-48.16	-22.56
154	*pauciflorus*	-49.46	-24.11
278	*pauciflorus*	-47.64	-22.72
279	*pauciflorus*	-47.71	-22.43
282	*pauciflorus*	-48.02	-22.73
283	*pauciflorus*	-49.32	-23.00
284	*pauciflorus*	-48.36	-21.60
285	*pauciflorus*	-47.92	-24.38
286	*pauciflorus*	-50.58	-23.91
287	*pauciflorus*	-48.51	-25.52
288	*pauciflorus*	-48.18	-23.03
289	*pauciflorus*	-49.55	-22.29
291	*pauciflorus*	-50.69	-22.24
292	*pauciflorus*	-46.51	-22.84
357	*pauciflorus*	-40.25	-15.25
358	*pauciflorus*	-48.12	-22.28
359	*pauciflorus*	-39.08	-15.15
468	*pauciflorus*	-48.57	-25.88
477	*pauciflorus*	-39.08	-15.29
478	*pauciflorus*	-48.33	-25.31
480	*pauciflorus*	-48.71	-25.43
481	*pauciflorus*	-50.71	-24.10
482	*pauciflorus*	-50.41	-24.51
483	*pauciflorus*	-49.90	-23.50
484	*pauciflorus*	-49.91	-23.57
485	*pauciflorus*	-52.34	-22.85
486	*pauciflorus*	-51.33	-24.57
593	*pauciflorus*	-48.75	-24.26
602	*pauciflorus*	-55.46	-13.15
603	*pauciflorus*	-92.53	-83.77
604	*pauciflorus*	-48.13	-22.17
605	*pauciflorus*	-47.80	-21.17
606	*pauciflorus*	-47.90	-21.30
651	*pauciflorus*	-39.41	-7.23
652	*pauciflorus*	-39.30	-7.93
653	*pauciflorus*	-39.55	-7.63
780	*pauciflorus*	-52.87	-10.83
859	*pauciflorus*	-40.35	-15.48
860	*pauciflorus*	-45.07	-23.43
861	*pauciflorus*	-46.63	-23.54
862	*pauciflorus*	-48.44	-22.88
865	*pauciflorus*	-41.37	-21.06
911	*pauciflorus*	-40.68	-19.93
1351	*riedelianus*	-40.20	-20.18
1352	*riedelianus*	-40.27	-19.82
1398	*riedelianus*	-42.48	-14.07
1450	*riedelianus*	-42.73	-19.71
1789	*riedelianus*	-40.49	-12.30
1790	*riedelianus*	-40.54	-12.42
1791	*riedelianus*	-39.47	-12.87
1792	*riedelianus*	-38.05	-12.02
1795	*riedelianus*	-39.17	-17.66
1796	*riedelianus*	-37.97	-12.17
1797	*riedelianus*	-41.51	-12.87
1798	*riedelianus*	-39.59	-14.87
1799	*riedelianus*	-40.29	-12.01
1800	*riedelianus*	-40.41	-13.41
1801	*riedelianus*	-40.11	-13.53
1802	*riedelianus*	-40.49	-12.30
1923	*riedelianus*	-38.98	-14.13
1947	*riedelianus*	-85.49	10.96
1966	*riedelianus*	-39.08	-14.39
2015	*riedelianus*	-39.47	-12.85
2021	*riedelianus*	-40.49	-12.30
2022	*riedelianus*	-40.54	-12.42
2038	*riedelianus*	-40.29	-20.33
2040	*riedelianus*	-41.02	-12.45
2184	*riedelianus*	-39.47	-12.87
2191	*riedelianus*	-39.69	-19.29
2196	*riedelianus*	-39.12	-14.50
2199	*riedelianus*	-39.08	-14.42
2221	*riedelianus*	-40.80	-19.89
2238	*riedelianus*	-38.05	-12.02
2253	*riedelianus*	-42.68	-14.77
2333	*riedelianus*	-40.27	-20.27
2335	*riedelianus*	-39.17	-17.66
2371	*riedelianus*	-37.97	-12.17
2384	*riedelianus*	-40.47	-11.85
2527	*riedelianus*	-41.51	-12.87
2544	*riedelianus*	-37.35	-11.10
2562	*riedelianus*	-39.40	-16.85
2632	*riedelianus*	-40.29	-12.01
2684	*riedelianus*	-38.99	-14.28
2693	*riedelianus*	-39.10	-14.48
2694	*riedelianus*	-40.43	-13.44
2697	*riedelianus*	-39.05	-14.79

934	*pauciflorus*	-35.02	-8.53
935	*pauciflorus*	-35.02	-8.52
936	*pauciflorus*	-43.21	-22.90
973	*pauciflorus*	-47.08	-22.87
974	*pauciflorus*	-49.32	-23.00
1004	*pauciflorus*	-39.08	-12.53
1103	*pauciflorus*	-50.75	-23.08
1104	*pauciflorus*	-47.47	-21.85
1105	*pauciflorus*	-51.19	-23.00
1109	*pauciflorus*	-51.16	-23.31
1239	*pauciflorus*	-51.34	-27.63
1242	*pauciflorus*	-48.55	-27.16
1243	*pauciflorus*	-48.52	-27.12
1291	*pauciflorus*	-47.55	-24.70
1303	*pauciflorus*	-47.25	-24.55
1307	*pauciflorus*	-46.99	-24.32
1444	*pauciflorus*	NA	NA
1486	*pauciflorus*	-40.35	-15.48
1505	*pauciflorus*	-49.83	-22.25
1527	*pauciflorus*	-48.63	-26.99
1540	*pauciflorus*	-48.25	-23.10
1542	*pauciflorus*	-48.98	-22.46
1550	*pauciflorus*	-47.43	-20.22
1562	*pauciflorus*	-48.21	-24.23
1566	*pauciflorus*	-47.47	-20.25
1567	*pauciflorus*	-48.77	-23.51
1568	*pauciflorus*	-48.75	-23.55
1570	*pauciflorus*	-49.06	-24.07
1571	*pauciflorus*	-49.07	-24.07
1572	*pauciflorus*	-47.22	-23.65
1573	*pauciflorus*	-49.06	-24.07
1574	*pauciflorus*	-47.01	-24.34
1575	*pauciflorus*	-52.16	-22.53
1607	*pauciflorus*	-48.87	-26.82
1608	*pauciflorus*	-48.96	-26.28
1719	*pauciflorus*	-52.86	-22.66
1782	*pauciflorus*	-48.02	-22.73
1885	*pauciflorus*	-49.94	-22.21
1892	*pauciflorus*	-49.38	-23.39
1896	*pauciflorus*	-48.16	-22.56
1904	*pauciflorus*	-49.36	-23.01
1957	*pauciflorus*	-49.38	-23.39
1959	*pauciflorus*	-39.08	-12.53
2698	*riedelianus*	-39.06	-16.45
2699	*riedelianus*	-39.03	-14.31
2736	*riedelianus*	-38.52	-15.18
2758	*riedelianus*	-38.98	-14.13
2825	*riedelianus*	-41.09	-13.53
2872	*riedelianus*	-40.82	-12.77
2893	*riedelianus*	-42.83	-22.43
2904	*riedelianus*	-39.38	-14.35
2914	*riedelianus*	-40.41	-13.41
2916	*riedelianus*	-40.11	-13.53
2930	*riedelianus*	-39.08	-15.15
2945	*riedelianus*	-39.07	-14.78
2952	*riedelianus*	-40.27	-19.82
2954	*riedelianus*	-38.95	-13.99
2967	*riedelianus*	-39.96	-19.14
2969	*riedelianus*	-39.41	-16.86
3033	*riedelianus*	-39.52	-15.15
3034	*riedelianus*	-39.52	-12.93
3044	*riedelianus*	-39.71	-14.18
3076	*riedelianus*	-40.07	-19.39
3091	*riedelianus*	-40.49	-12.30
3093	*riedelianus*	-39.37	-14.33
3389	*riedelianus*	-38.95	-13.98
28	*spicatus*	-41.75	-13.31
29	*spicatus*	-37.43	-7.19
30	*spicatus*	-40.57	-15.26
31	*spicatus*	-39.58	-16.92
32	*spicatus*	-39.08	-15.15
34	*spicatus*	-41.78	-13.55
35	*spicatus*	-41.27	-10.45
36	*spicatus*	-39.57	-15.13
37	*spicatus*	-39.71	-14.64
38	*spicatus*	-40.48	-20.62
49	*spicatus*	-53.09	-10.77
50	*spicatus*	-42.91	-22.48
51	*spicatus*	-42.00	-12.18
62	*spicatus*	-41.16	-11.55
93	*spicatus*	-39.07	-13.37
104	*spicatus*	-43.21	-22.90
105	*spicatus*	-41.77	-12.42
202	*spicatus*	-40.99	-11.63
203	*spicatus*	-40.21	-18.37
204	*spicatus*	-40.77	-20.87

2028	*pauciflorus*	-48.21	-24.23
2198	*pauciflorus*	-46.51	-22.84
2346	*pauciflorus*	-48.02	-22.73
2376	*pauciflorus*	-48.87	-26.82
2411	*pauciflorus*	-47.43	-20.22
2444	*pauciflorus*	-51.87	-23.35
2601	*pauciflorus*	-49.83	-22.25
2630	*pauciflorus*	-48.77	-23.51
2712	*pauciflorus*	-48.12	-22.28
2714	*pauciflorus*	-48.44	-22.88
2715	*pauciflorus*	-47.55	-24.70
2717	*pauciflorus*	-47.00	-23.39
2718	*pauciflorus*	-46.63	-23.54
2719	*pauciflorus*	-48.51	-25.52
2720	*pauciflorus*	-48.33	-25.31
2721	*pauciflorus*	-48.16	-22.56
2722	*pauciflorus*	-49.46	-24.12
2732	*pauciflorus*	-50.71	-24.10
2733	*pauciflorus*	-48.75	-23.55
2741	*pauciflorus*	-49.06	-24.07
2742	*pauciflorus*	-49.07	-24.07
2804	*pauciflorus*	-49.06	-24.07
2809	*pauciflorus*	-47.01	-24.34
2811	*pauciflorus*	-35.02	-8.53
2835	*pauciflorus*	-48.96	-26.28
2850	*pauciflorus*	-50.74	-23.73
2876	*pauciflorus*	-48.52	-27.12
2901	*pauciflorus*	-35.02	-8.52
2980	*pauciflorus*	-49.36	-23.01
2987	*pauciflorus*	-48.75	-24.26
3029	*pauciflorus*	-40.68	-19.93
3080	*pauciflorus*	-48.02	-22.73
3098	*pauciflorus*	-48.18	-23.03
3128	*pauciflorus*	-39.75	-16.56
3220	*pauciflorus*	-53.09	-10.77
3221	*pauciflorus*	-51.16	-27.03
3222	*pauciflorus*	-42.91	-22.48
3235	*pauciflorus*	-65.91	7.08
1741	*penatiflolius*	-53.48	-23.08
487	*pennatfolius*	-50.71	-24.10
17	*pennatifolius*	-53.53	-26.70
63	*pennatifolius*	-49.52	-27.06
69	*pennatifolius*	-50.48	-23.54
205	*spicatus*	-39.58	-16.96
206	*spicatus*	-40.42	-10.94
207	*spicatus*	-39.08	-15.15
209	*spicatus*	-39.21	-17.57
210	*spicatus*	-41.75	-13.25
211	*spicatus*	-45.07	-23.43
213	*spicatus*	-47.55	-24.70
214	*spicatus*	-39.08	-15.29
215	*spicatus*	-41.83	-13.22
216	*spicatus*	-40.38	-12.25
218	*spicatus*	-41.75	-13.31
220	*spicatus*	-40.57	-15.26
221	*spicatus*	-43.67	-19.33
222	*spicatus*	-38.62	-3.98
223	*spicatus*	-41.02	-11.65
224	*spicatus*	-41.80	-13.58
226	*spicatus*	-41.78	-13.55
227	*spicatus*	-41.27	-10.45
228	*spicatus*	-41.77	-13.15
229	*spicatus*	-40.52	-11.18
233	*spicatus*	-39.57	-15.14
234	*spicatus*	-40.48	-20.62
235	*spicatus*	-39.01	-14.10
236	*spicatus*	-43.21	-22.90
237	*spicatus*	-52.16	-22.53
240	*spicatus*	-39.22	-17.50
242	*spicatus*	-39.23	-17.72
247	*spicatus*	-39.07	-13.37
312	*spicatus*	-41.97	-12.41
313	*spicatus*	-41.88	-22.75
314	*spicatus*	-41.66	-13.25
396	*spicatus*	-40.54	-12.39
397	*spicatus*	-39.20	-17.52
399	*spicatus*	-40.43	-13.44
402	*spicatus*	-40.52	-11.18
406	*spicatus*	-40.60	-19.94
407	*spicatus*	-39.58	-17.00
408	*spicatus*	-42.48	-14.07
409	*spicatus*	-42.70	-13.02
410	*spicatus*	-40.15	-13.53
412	*spicatus*	-41.75	-13.31
413	*spicatus*	-39.11	-13.60
415	*spicatus*	-39.06	-16.45

91	*pennatifolius*	-47.47	-21.85
103	*pennatifolius*	-47.29	-22.77
155	*pennatifolius*	-54.59	-25.55
156	*pennatifolius*	-50.58	-29.45
157	*pennatifolius*	-52.65	-27.37
158	*pennatifolius*	-50.91	-23.07
159	*pennatifolius*	-53.81	-25.67
160	*pennatifolius*	-54.09	-25.30
161	*pennatifolius*	-52.42	-25.41
162	*pennatifolius*	-53.53	-26.70
163	*pennatifolius*	-49.64	-27.21
293	*pennatifolius*	-56.48	-21.12
294	*pennatifolius*	-49.55	-22.29
296	*pennatifolius*	-49.13	-22.41
297	*pennatifolius*	-52.16	-22.53
299	*pennatifolius*	-47.64	-22.72
302	*pennatifolius*	-52.68	-29.60
303	*pennatifolius*	-51.05	-23.27
305	*pennatifolius*	-54.59	-25.55
306	*pennatifolius*	-52.57	-25.98
307	*pennatifolius*	-47.06	-22.90
308	*pennatifolius*	-50.41	-22.66
309	*pennatifolius*	-47.77	-22.51
362	*pennatifolius*	-53.72	-25.92
439	*pennatifolius*	-46.63	-23.54
440	*pennatifolius*	-53.76	-27.37
442	*pennatifolius*	-53.44	-29.57
450	*pennatifolius*	-47.49	-5.53
489	*pennatifolius*	-49.27	-25.43
490	*pennatifolius*	-48.68	-27.96
491	*pennatifolius*	-52.34	-22.73
492	*pennatifolius*	-54.06	-24.56
493	*pennatifolius*	-52.42	-25.41
494	*pennatifolius*	-52.44	-23.77
497	*pennatifolius*	-51.94	-23.43
498	*pennatifolius*	-49.26	-24.82
501	*pennatifolius*	-53.49	-23.74
502	*pennatifolius*	-50.37	-23.11
504	*pennatifolius*	-50.65	-23.18
505	*pennatifolius*	-51.66	-25.70
509	*pennatifolius*	-52.62	-27.10
511	*pennatifolius*	-52.91	-23.01
512	*pennatifolius*	-49.73	-29.34
416	*spicatus*	-39.01	-14.10
417	*spicatus*	-41.56	-12.53
418	*spicatus*	-40.07	-15.12
419	*spicatus*	-39.53	-15.95
420	*spicatus*	-39.23	-17.72
421	*spicatus*	-39.22	-17.50
422	*spicatus*	-45.07	-23.43
426	*spicatus*	-41.27	-10.45
427	*spicatus*	-41.78	-13.55
428	*spicatus*	-41.02	-11.65
429	*spicatus*	-41.80	-13.58
430	*spicatus*	-39.57	-15.13
443	*spicatus*	-44.75	-15.51
454	*spicatus*	-43.21	-22.90
555	*spicatus*	-41.30	-13.81
556	*spicatus*	-42.87	-15.16
561	*spicatus*	-49.95	-23.78
564	*spicatus*	-40.62	-15.25
565	*spicatus*	-47.47	-20.25
568	*spicatus*	-40.51	-20.66
571	*spicatus*	-52.16	-22.53
575	*spicatus*	-48.83	-25.48
598	*spicatus*	-42.88	-15.15
599	*spicatus*	-41.58	-12.53
626	*spicatus*	-40.68	-5.18
673	*spicatus*	-38.69	-3.89
674	*spicatus*	-37.91	-6.09
675	*spicatus*	-38.44	-6.26
676	*spicatus*	-40.77	-5.53
677	*spicatus*	-39.55	-7.63
679	*spicatus*	-41.09	-3.56
682	*spicatus*	-40.85	-5.08
686	*spicatus*	-45.24	-5.51
690	*spicatus*	-39.94	-7.78
692	*spicatus*	-39.01	-4.29
698	*spicatus*	-39.01	-4.17
699	*spicatus*	-38.92	-4.23
701	*spicatus*	-39.72	-7.51
708	*spicatus*	-40.60	-11.43
754	*spicatus*	-39.95	-19.19
755	*spicatus*	-40.07	-19.39
802	*spicatus*	-52.87	-10.83
875	*spicatus*	-41.82	-12.43

513	*pennatifolius*	-48.28	-18.92
514	*pennatifolius*	-51.91	-27.46
515	*pennatifolius*	-50.84	-23.37
516	*pennatifolius*	-53.13	-25.62
517	*pennatifolius*	-53.61	-25.48
518	*pennatifolius*	-50.36	-23.41
519	*pennatifolius*	-50.41	-24.51
521	*pennatifolius*	-51.61	-27.34
522	*pennatifolius*	-51.80	-27.63
524	*pennatifolius*	-54.09	-25.30
525	*pennatifolius*	-53.41	-23.90
527	*pennatifolius*	-51.66	-22.76
528	*pennatifolius*	-51.03	-22.85
532	*pennatifolius*	-48.51	-25.52
533	*pennatifolius*	-52.44	-24.73
537	*pennatifolius*	-51.98	-23.92
539	*pennatifolius*	-54.26	-24.08
542	*pennatifolius*	-55.95	-22.19
546	*pennatifolius*	-51.35	-29.23
594	*pennatifolius*	-50.98	-23.54
607	*pennatifolius*	-55.46	-13.15
608	*pennatifolius*	-92.53	-83.77
609	*pennatifolius*	-47.80	-21.17
610	*pennatifolius*	-49.69	-22.42
611	*pennatifolius*	-47.90	-21.30
613	*pennatifolius*	-51.16	-23.31
614	*pennatifolius*	-51.56	-27.26
615	*pennatifolius*	-52.10	-25.64
616	*pennatifolius*	-51.22	-27.40
617	*pennatifolius*	-50.17	-29.48
618	*pennatifolius*	-49.85	-29.37
624	*pennatifolius*	-51.61	-27.35
733	*pennatifolius*	-48.54	-25.82
789	*pennatifolius*	-52.87	-10.83
866	*pennatifolius*	-50.76	-28.20
913	*pennatifolius*	-47.06	-22.91
931	*pennatifolius*	-51.99	-29.51
932	*pennatifolius*	-54.87	-29.19
933	*pennatifolius*	-43.21	-22.90
944	*pennatifolius*	-52.08	-23.97
945	*pennatifolius*	-52.38	-24.05
946	*pennatifolius*	-51.97	-23.87
947	*pennatifolius*	-51.97	-23.88
876	*spicatus*	-42.35	-13.86
877	*spicatus*	-42.76	-13.60
878	*spicatus*	-41.73	-18.23
879	*spicatus*	-41.82	-12.43
880	*spicatus*	-45.33	-5.67
881	*spicatus*	-39.58	-17.83
882	*spicatus*	-41.77	-12.42
883	*spicatus*	-39.03	-13.52
884	*spicatus*	-38.62	-3.98
885	*spicatus*	-38.68	-3.89
886	*spicatus*	-40.52	-11.18
887	*spicatus*	-41.83	-13.22
889	*spicatus*	-41.79	-22.37
890	*spicatus*	-47.82	-15.75
891	*spicatus*	-40.52	-19.55
892	*spicatus*	-40.85	-19.55
893	*spicatus*	-40.40	-11.15
894	*spicatus*	-40.85	-19.92
899	*spicatus*	-40.77	-20.87
915	*spicatus*	-40.68	-19.81
916	*spicatus*	-40.68	-19.81
919	*spicatus*	-40.16	-18.30
922	*spicatus*	-40.73	-20.86
928	*spicatus*	-50.03	-20.69
1022	*spicatus*	-41.02	-11.68
1023	*spicatus*	-41.00	-11.68
1024	*spicatus*	-39.69	-12.19
1025	*spicatus*	-40.50	-11.33
1027	*spicatus*	-41.65	-14.20
1029	*spicatus*	-41.00	-11.62
1030	*spicatus*	-41.51	-12.87
1031	*spicatus*	-39.61	-14.18
1032	*spicatus*	-41.72	-13.29
1033	*spicatus*	-41.01	-11.67
1034	*spicatus*	-41.52	-12.37
1035	*spicatus*	-42.21	-13.99
1036	*spicatus*	-40.57	-15.26
1037	*spicatus*	-40.42	-13.43
1038	*spicatus*	-41.01	-11.68
1039	*spicatus*	-41.02	-11.67
1040	*spicatus*	-41.55	-12.52
1041	*spicatus*	-40.52	-11.17
1043	*spicatus*	-39.07	-13.37

948	*pennatifolius*	-52.88	-23.87
949	*pennatifolius*	-52.84	-24.43
950	*pennatifolius*	-54.47	-25.63
952	*pennatifolius*	-52.12	-24.10
954	*pennatifolius*	-52.50	-23.93
957	*pennatifolius*	-52.30	-24.10
959	*pennatifolius*	-53.19	-24.53
960	*pennatifolius*	-52.35	-24.03
961	*pennatifolius*	-52.35	-24.03
962	*pennatifolius*	-52.99	-26.85
978	*pennatifolius*	-47.06	-22.91
979	*pennatifolius*	-47.09	-22.86
982	*pennatifolius*	-47.08	-22.87
983	*pennatifolius*	-47.04	-22.80
1058	*pennatifolius*	-50.78	-23.52
1061	*pennatifolius*	-53.06	-25.73
1074	*pennatifolius*	-55.08	-25.31
1077	*pennatifolius*	-54.00	-25.24
1080	*pennatifolius*	-53.48	-23.08
1082	*pennatifolius*	-53.77	-25.72
1084	*pennatifolius*	-53.96	-25.31
1110	*pennatifolius*	-51.04	-23.06
1125	*pennatifolius*	-51.37	-23.31
1128	*pennatifolius*	-51.19	-23.00
1129	*pennatifolius*	-50.84	-23.04
1131	*pennatifolius*	-50.56	-23.43
1151	*pennatifolius*	-50.52	-23.15
1168	*pennatifolius*	-53.35	-25.63
1170	*pennatifolius*	-51.70	-22.69
1173	*pennatifolius*	-51.15	-23.15
1175	*pennatifolius*	-51.74	-29.17
1176	*pennatifolius*	-51.28	-23.28
1181	*pennatifolius*	-51.42	-23.42
1182	*pennatifolius*	-50.80	-23.20
1198	*pennatifolius*	-49.38	-23.19
1205	*pennatifolius*	-50.77	-19.15
1211	*pennatifolius*	-50.58	-29.45
1212	*pennatifolius*	-54.76	-27.66
1214	*pennatifolius*	-53.39	-27.36
1215	*pennatifolius*	-50.42	-28.67
1217	*pennatifolius*	-55.15	-29.05
1218	*pennatifolius*	-54.26	-28.30
1219	*pennatifolius*	-54.75	-27.95
1073	*spicatus*	-41.16	-11.55
1200	*spicatus*	-42.60	-18.37
1202	*spicatus*	-40.44	-11.09
1268	*spicatus*	-44.04	-22.96
1283	*spicatus*	-51.13	-27.69
1284	*spicatus*	-51.73	-27.18
1314	*spicatus*	-41.83	-12.70
1315	*spicatus*	-48.17	-21.79
1326	*spicatus*	-44.25	-19.47
1355	*spicatus*	-40.18	-20.15
1356	*spicatus*	-40.70	-20.83
1357	*spicatus*	-39.75	-19.44
1359	*spicatus*	-40.72	-20.84
1360	*spicatus*	-39.73	-18.59
1361	*spicatus*	-39.85	-18.96
1453	*spicatus*	-48.38	-18.49
1454	*spicatus*	-46.99	-24.32
1456	*spicatus*	-42.47	-16.95
1537	*spicatus*	-47.47	-21.70
1598	*spicatus*	-46.39	-23.96
1600	*spicatus*	-39.75	-15.93
1696	*spicatus*	-48.19	-18.65
1707	*spicatus*	-39.30	-7.93
1709	*spicatus*	-39.41	-7.23
1710	*spicatus*	-35.94	-8.97
1803	*spicatus*	-41.98	-12.41
1804	*spicatus*	-41.58	-12.54
1805	*spicatus*	-39.08	-12.53
1806	*spicatus*	-41.50	-12.33
1807	*spicatus*	-41.19	-11.44
1809	*spicatus*	-41.31	-11.83
1810	*spicatus*	-40.38	-12.25
1811	*spicatus*	-39.00	-13.58
1815	*spicatus*	-41.32	-10.37
1817	*spicatus*	-39.08	-15.15
1818	*spicatus*	-42.06	-11.81
1819	*spicatus*	-42.05	-11.80
1820	*spicatus*	-40.33	-13.46
1821	*spicatus*	-41.76	-13.60
1822	*spicatus*	-41.12	-12.56
1823	*spicatus*	-41.49	-12.41
1825	*spicatus*	-40.92	-5.14
1827	*spicatus*	-41.76	-13.60

1220	*pennatifolius*	-53.28	-27.63
1221	*pennatifolius*	-53.43	-27.40
1222	*pennatifolius*	-54.72	-29.34
1223	*pennatifolius*	-54.69	-29.50
1224	*pennatifolius*	-52.27	-27.63
1227	*pennatifolius*	-52.73	-27.53
1233	*pennatifolius*	-52.99	-27.73
1248	*pennatifolius*	-50.45	-29.64
1250	*pennatifolius*	-49.30	-28.71
1251	*pennatifolius*	-53.52	-26.73
1252	*pennatifolius*	-52.46	-26.63
1255	*pennatifolius*	-51.67	-27.57
1256	*pennatifolius*	-51.13	-27.69
1257	*pennatifolius*	-53.71	-27.17
1258	*pennatifolius*	-49.09	-26.88
1260	*pennatifolius*	-53.07	-28.06
1281	*pennatifolius*	-50.81	-27.49
1282	*pennatifolius*	-50.51	-27.94
1285	*pennatifolius*	-50.83	-17.17
1311	*pennatifolius*	-44.57	-22.64
1312	*pennatifolius*	-47.56	-22.41
1323	*pennatifolius*	-50.58	-23.91
1325	*pennatifolius*	-50.08	-17.73
1345	*pennatifolius*	-51.52	-29.17
1378	*pennatifolius*	-51.82	-23.98
1391	*pennatifolius*	-48.68	-23.94
1393	*pennatifolius*	-53.34	-25.21
1411	*pennatifolius*	-51.18	-29.17
1416	*pennatifolius*	-49.68	-27.54
1417	*pennatifolius*	-52.38	-27.09
1423	*pennatifolius*	-51.41	-28.99
1448	*pennatifolius*	NA	NA
1449	*pennatifolius*	-43.51	-20.29
1473	*pennatifolius*	-56.64	-21.26
1494	*pennatifolius*	-48.19	-18.65
1500	*pennatifolius*	-47.67	-19.15
1507	*pennatifolius*	-51.11	-19.31
1532	*pennatifolius*	-51.61	-27.27
1543	*pennatifolius*	-48.55	-22.29
1587	*pennatifolius*	-49.07	-24.06
1588	*pennatifolius*	-53.97	-25.31
1589	*pennatifolius*	-53.99	-25.24
1590	*pennatifolius*	-54.11	-25.30
1828	*spicatus*	-41.93	-13.52
1829	*spicatus*	-41.76	-13.60
1830	*spicatus*	-41.16	-12.93
1831	*spicatus*	-41.30	-13.83
1832	*spicatus*	-41.55	-12.53
1833	*spicatus*	-41.58	-12.52
1834	*spicatus*	-41.76	-13.62
1835	*spicatus*	-41.82	-12.43
1836	*spicatus*	-41.57	-12.57
1837	*spicatus*	-41.76	-13.62
1839	*spicatus*	-42.76	-13.60
1841	*spicatus*	-41.82	-12.43
1842	*spicatus*	-41.15	-11.75
1843	*spicatus*	-40.99	-11.63
1845	*spicatus*	-40.91	-11.56
1846	*spicatus*	-41.31	-10.42
1849	*spicatus*	-42.21	-13.25
1852	*spicatus*	-40.52	-12.47
1853	*spicatus*	-40.48	-12.34
1854	*spicatus*	-42.64	-14.07
1855	*spicatus*	-38.80	-12.25
1940	*spicatus*	-39.71	-14.64
1942	*spicatus*	-40.21	-18.37
1954	*spicatus*	-42.21	-13.99
1958	*spicatus*	-40.42	-13.43
2006	*spicatus*	-40.54	-12.39
2017	*spicatus*	-41.00	-11.68
2020	*spicatus*	-41.02	-11.68
2024	*spicatus*	-41.98	-12.41
2026	*spicatus*	-41.58	-12.54
2029	*spicatus*	-41.27	-10.45
2048	*spicatus*	-39.69	-12.19
2053	*spicatus*	-41.15	-11.75
2152	*spicatus*	-40.68	-19.81
2178	*spicatus*	-40.99	-11.63
2179	*spicatus*	-39.08	-12.53
2188	*spicatus*	-42.21	-13.25
2195	*spicatus*	-41.50	-12.33
2209	*spicatus*	-40.68	-19.81
2210	*spicatus*	-41.19	-11.44
2213	*spicatus*	-41.52	-12.37
2216	*spicatus*	-39.95	-19.19
2220	*spicatus*	-41.31	-11.83

1592	*pennatifolius*	-53.87	-25.39
1601	*pennatifolius*	-51.44	-27.14
1602	*pennatifolius*	-49.30	-28.70
1603	*pennatifolius*	-52.68	-29.60
1604	*pennatifolius*	-49.47	-28.75
1605	*pennatifolius*	-49.42	-28.60
1606	*pennatifolius*	-49.37	-28.68
1609	*pennatifolius*	-53.28	-26.69
1610	*pennatifolius*	-51.61	-27.49
1611	*pennatifolius*	-51.48	-27.54
1613	*pennatifolius*	-52.20	-27.13
1617	*pennatifolius*	-51.48	-27.13
1618	*pennatifolius*	-51.89	-27.36
1619	*pennatifolius*	-53.46	-26.96
1620	*pennatifolius*	-52.52	-26.77
1621	*pennatifolius*	-49.86	-28.98
1622	*pennatifolius*	-49.78	-27.86
1623	*pennatifolius*	-49.23	-28.35
1624	*pennatifolius*	-49.32	-28.35
1625	*pennatifolius*	-49.77	-26.91
1627	*pennatifolius*	-49.95	-29.05
1628	*pennatifolius*	-49.99	-26.83
1630	*pennatifolius*	-49.68	-26.82
1631	*pennatifolius*	-53.50	-26.83
1632	*pennatifolius*	-49.90	-26.90
1633	*pennatifolius*	-49.92	-26.87
1635	*pennatifolius*	-49.24	-25.44
1636	*pennatifolius*	-53.86	-27.26
1637	*pennatifolius*	-52.38	-27.00
1638	*pennatifolius*	-53.28	-26.46
1639	*pennatifolius*	-49.85	-27.20
1640	*pennatifolius*	-49.78	-27.42
1642	*pennatifolius*	-51.39	-26.91
1643	*pennatifolius*	-51.30	-26.91
1645	*pennatifolius*	-50.21	-29.68
1647	*pennatifolius*	-50.85	-29.31
1656	*pennatifolius*	-52.77	-27.36
1658	*pennatifolius*	-51.38	-29.59
1659	*pennatifolius*	-51.13	-29.68
1662	*pennatifolius*	-51.46	-29.69
1663	*pennatifolius*	-54.48	-27.87
1666	*pennatifolius*	-51.23	-30.03
1672	*pennatifolius*	-50.82	-29.37
2227	*spicatus*	-48.38	-18.49
2251	*spicatus*	-40.18	-20.15
2283	*spicatus*	-39.75	-15.93
2284	*spicatus*	-41.75	-13.31
2300	*spicatus*	-41.16	-11.55
2318	*spicatus*	-40.38	-12.25
2325	*spicatus*	-39.75	-19.44
2334	*spicatus*	-39.00	-13.58
2340	*spicatus*	-41.02	-11.65
2373	*spicatus*	-39.23	-17.72
2429	*spicatus*	-40.72	-20.84
2459	*spicatus*	-39.73	-19.35
2476	*spicatus*	-39.85	-18.96
2483	*spicatus*	-39.58	-16.92
2487	*spicatus*	-37.43	-7.19
2529	*spicatus*	-41.00	-11.62
2532	*spicatus*	-42.64	-14.07
2560	*spicatus*	-40.85	-5.08
2592	*spicatus*	-40.16	-18.30
2603	*spicatus*	-41.32	-10.37
2605	*spicatus*	-41.87	-22.76
2606	*spicatus*	-43.62	-23.05
2608	*spicatus*	-41.03	-11.93
2614	*spicatus*	-41.51	-12.37
2620	*spicatus*	-41.72	-12.55
2625	*spicatus*	-40.91	-11.56
2649	*spicatus*	-40.42	-10.94
2651	*spicatus*	-41.77	-12.42
2652	*spicatus*	-39.21	-17.57
2655	*spicatus*	-45.07	-23.43
2657	*spicatus*	-47.55	-24.70
2658	*spicatus*	-39.08	-15.30
2663	*spicatus*	-41.77	-13.16
2664	*spicatus*	-40.52	-11.18
2668	*spicatus*	-39.01	-14.11
2669	*spicatus*	-43.21	-22.91
2670	*spicatus*	-52.16	-22.53
2679	*spicatus*	-39.07	-13.37
2749	*spicatus*	-39.01	-4.29
2764	*spicatus*	-45.67	-20.35
2767	*spicatus*	-41.95	-22.83
2817	*spicatus*	-39.01	-4.17
2818	*spicatus*	-40.73	-20.86

1676	*pennatifolius*	-51.01	-29.64
1677	*pennatifolius*	-50.52	-29.82
1683	*pennatifolius*	-53.84	-29.08
1684	*pennatifolius*	-53.17	-26.76
1722	*pennatifolius*	-51.93	-23.43
1724	*pennatifolius*	-52.13	-23.85
1725	*pennatifolius*	-53.27	-22.77
1726	*pennatifolius*	-51.87	-23.35
1727	*pennatifolius*	-51.94	-23.39
1728	*pennatifolius*	-51.87	-23.35
1730	*pennatifolius*	-51.87	-23.34
1731	*pennatifolius*	-51.87	-23.35
1732	*pennatifolius*	-51.94	-23.39
1733	*pennatifolius*	-53.22	-22.82
1738	*pennatifolius*	-52.86	-22.66
1743	*pennatifolius*	-51.46	-27.67
1744	*pennatifolius*	-49.85	-29.37
1745	*pennatifolius*	-53.50	-28.29
1746	*pennatifolius*	-50.97	-29.61
1747	*pennatifolius*	-50.94	-29.60
1749	*pennatifolius*	-51.11	-29.38
1752	*pennatifolius*	-55.48	-29.59
1910	*pennatifolius*	-47.15	-22.76
1911	*pennatifolius*	-53.45	-26.95
1913	*pennatifolius*	-49.06	-22.31
1914	*pennatifolius*	-55.23	-23.10
1937	*pennatifolius*	-50.98	-23.54
1965	*pennatifolius*	-53.72	-25.92
1974	*pennatifolius*	-56.26	-25.80
1975	*pennatifolius*	-57.36	-22.53
2005	*pennatifolius*	-51.61	-27.27
2018	*pennatifolius*	-51.22	-27.40
2042	*pennatifolius*	-53.28	-26.69
2051	*pennatifolius*	-51.61	-27.49
2056	*pennatifolius*	-54.82	-27.05
2057	*pennatifolius*	-52.38	-27.09
2060	*pennatifolius*	-57.23	-25.15
2061	*pennatifolius*	-56.83	-26.08
2063	*pennatifolius*	-56.25	-25.67
2064	*pennatifolius*	-56.25	-25.92
2066	*pennatifolius*	-56.42	-25.75
2069	*pennatifolius*	-50.17	-29.48
2070	*pennatifolius*	-56.83	-26.17
2870	*spicatus*	-42.35	-13.86
2871	*spicatus*	-42.76	-13.60
2873	*spicatus*	-41.73	-18.23
2874	*spicatus*	-41.82	-12.43
2895	*spicatus*	-45.33	-5.67
2907	*spicatus*	-41.31	-10.42
2911	*spicatus*	-42.27	-13.84
2915	*spicatus*	-40.33	-13.46
2919	*spicatus*	-41.76	-13.60
2920	*spicatus*	-41.12	-12.56
2922	*spicatus*	-41.49	-12.41
2964	*spicatus*	-38.62	-3.99
2966	*spicatus*	-40.92	-5.14
2968	*spicatus*	-41.72	-13.29
2971	*spicatus*	-41.76	-13.60
2973	*spicatus*	-41.93	-13.52
2978	*spicatus*	-41.76	-13.60
2981	*spicatus*	-52.53	-0.83
2989	*spicatus*	-41.16	-12.93
2991	*spicatus*	-41.02	-11.67
2993	*spicatus*	-41.01	-11.68
3004	*spicatus*	-41.55	-12.53
3008	*spicatus*	-40.99	-18.57
3009	*spicatus*	-41.58	-12.52
3011	*spicatus*	-40.13	-14.41
3013	*spicatus*	-40.21	-14.36
3014	*spicatus*	-41.76	-13.62
3030	*spicatus*	-39.57	-15.13
3038	*spicatus*	-41.01	-11.67
3043	*spicatus*	-40.52	-12.47
3045	*spicatus*	-40.58	-18.68
3046	*spicatus*	-40.59	-18.71
3048	*spicatus*	-41.57	-12.57
3056	*spicatus*	-41.97	-12.41
3084	*spicatus*	-41.76	-13.62
3097	*spicatus*	-41.82	-12.43
3118	*spicatus*	-40.48	-12.34
3129	*spicatus*	-42.44	-22.93
3166	*spicatus*	-74.14	-9.18
3227	*spicatus*	-42.00	-12.18
3229	*spicatus*	-53.09	-10.77
3230	*spicatus*	-42.91	-22.48
2	*sulcatus*	-42.49	-13.55

2071	*pennatifolius*	-57.15	-25.22
2080	*pennatifolius*	-57.17	-25.42
2082	*pennatifolius*	-56.07	-22.63
2084	*pennatifolius*	-56.87	-25.65
2085	*pennatifolius*	-57.53	-22.62
2086	*pennatifolius*	-54.63	-25.27
2087	*pennatifolius*	-54.63	-25.13
2094	*pennatifolius*	-57.12	-25.42
2095	*pennatifolius*	-57.13	-25.87
2101	*pennatifolius*	-57.43	-25.63
2102	*pennatifolius*	-57.43	-25.67
2103	*pennatifolius*	-56.25	-25.75
2106	*pennatifolius*	-57.47	-25.33
2111	*pennatifolius*	-57.47	-25.33
2114	*pennatifolius*	-56.67	-25.25
2119	*pennatifolius*	-57.13	-25.92
2122	*pennatifolius*	-55.53	-24.13
2125	*pennatifolius*	-56.17	-25.92
2130	*pennatifolius*	-56.57	-24.20
2140	*pennatifolius*	-57.25	-25.13
2141	*pennatifolius*	-56.35	-25.72
2148	*pennatifolius*	-57.18	-25.58
2149	*pennatifolius*	-56.78	-26.08
2160	*pennatifolius*	-52.20	-27.13
2162	*pennatifolius*	-56.83	-26.05
2164	*pennatifolius*	-57.17	-25.87
2172	*pennatifolius*	-51.48	-27.13
2174	*pennatifolius*	-51.89	-27.36
2177	*pennatifolius*	-53.46	-26.96
2181	*pennatifolius*	-49.85	-29.37
2189	*pennatifolius*	-52.08	-23.97
2201	*pennatifolius*	-52.52	-26.77
2208	*pennatifolius*	-51.93	-23.43
2222	*pennatifolius*	-50.97	-29.61
2232	*pennatifolius*	-51.97	-23.87
2239	*pennatifolius*	-54.12	-25.97
2241	*pennatifolius*	-49.86	-28.98
2245	*pennatifolius*	-51.97	-23.88
2277	*pennatifolius*	-50.94	-29.60
2292	*pennatifolius*	-52.84	-24.43
2303	*pennatifolius*	-54.47	-25.63
2307	*pennatifolius*	-55.57	-27.28
2315	*pennatifolius*	-55.49	-27.65
39	*sulcatus*	-42.39	-13.82
53	*sulcatus*	-42.00	-12.18
66	*sulcatus*	-53.09	-10.77
252	*sulcatus*	-42.39	-13.82
253	*sulcatus*	-42.49	-13.55
254	*sulcatus*	-42.47	-14.06
255	*sulcatus*	-42.87	-15.16
316	*sulcatus*	-42.39	-13.82
317	*sulcatus*	-42.87	-15.16
318	*sulcatus*	-42.48	-14.07
432	*sulcatus*	-42.66	-14.77
709	*sulcatus*	-40.51	-11.16
820	*sulcatus*	-52.87	-10.83
900	*sulcatus*	-42.48	-13.53
1044	*sulcatus*	-42.47	-14.07
1045	*sulcatus*	-42.21	-13.99
1046	*sulcatus*	-40.97	-12.99
1047	*sulcatus*	-42.68	-14.77
1330	*sulcatus*	-42.38	-13.80
1331	*sulcatus*	-42.39	-13.82
1697	*sulcatus*	-41.16	-11.55
1857	*sulcatus*	-42.28	-11.68
1860	*sulcatus*	-42.38	-13.76
1861	*sulcatus*	-41.42	-13.90
1862	*sulcatus*	-42.39	-13.82
1931	*sulcatus*	-41.31	-13.84
1951	*sulcatus*	-42.47	-14.07
1984	*sulcatus*	-42.39	-13.82
2285	*sulcatus*	-42.38	-13.80
2328	*sulcatus*	-42.28	-11.68
2393	*sulcatus*	-42.49	-13.55
2533	*sulcatus*	-42.21	-13.99
2612	*sulcatus*	-42.39	-13.82
2672	*sulcatus*	-42.87	-15.16
2754	*sulcatus*	-41.31	-13.84
2782	*sulcatus*	-40.51	-11.16
2863	*sulcatus*	-42.38	-13.76
2958	*sulcatus*	-42.39	-13.82
3088	*sulcatus*	-42.68	-14.77
3151	*sulcatus*	-53.09	-10.77
3231	*sulcatus*	-42.00	-12.18
40	*trachyllophus*	-41.85	-13.23
67	*trachyllophus*	-44.10	-19.69

2332	*pennatifolius*	-56.64	-21.26
2372	*pennatifolius*	-50.83	-17.17
2379	*pennatifolius*	-50.08	-17.73
2397	*pennatifolius*	-55.58	-27.75
2399	*pennatifolius*	-54.95	-27.08
2400	*pennatifolius*	-49.23	-28.35
2401	*pennatifolius*	-49.32	-28.35
2403	*pennatifolius*	-51.87	-23.35
2406	*pennatifolius*	-54.82	-27.05
2407	*pennatifolius*	-53.87	-26.33
2408	*pennatifolius*	-53.88	-27.13
2409	*pennatifolius*	-53.88	-27.07
2413	*pennatifolius*	-51.87	-23.35
2417	*pennatifolius*	-55.58	-27.75
2419	*pennatifolius*	-55.48	-27.65
2420	*pennatifolius*	-54.45	-25.68
2421	*pennatifolius*	-55.57	-27.28
2422	*pennatifolius*	-54.45	-25.67
2425	*pennatifolius*	-51.87	-23.34
2430	*pennatifolius*	-55.62	-27.75
2445	*pennatifolius*	-51.11	-19.31
2447	*pennatifolius*	-55.53	-27.32
2448	*pennatifolius*	-54.95	-27.08
2451	*pennatifolius*	-54.12	-25.97
2453	*pennatifolius*	-54.53	-27.27
2454	*pennatifolius*	-49.77	-26.91
2462	*pennatifolius*	-49.68	-27.54
2463	*pennatifolius*	-49.95	-29.05
2469	*pennatifolius*	-51.94	-23.39
2470	*pennatifolius*	-50.48	-23.54
2472	*pennatifolius*	-49.99	-26.83
2482	*pennatifolius*	-55.28	-26.17
2486	*pennatifolius*	-55.33	-26.17
2545	*pennatifolius*	-49.68	-26.82
2549	*pennatifolius*	-58.17	-25.99
2582	*pennatifolius*	-54.92	-26.88
2583	*pennatifolius*	-55.47	-27.91
2585	*pennatifolius*	-54.27	-26.98
2590	*pennatifolius*	-49.90	-26.90
2591	*pennatifolius*	-49.92	-26.87
2637	*pennatifolius*	-54.59	-25.55
2638	*pennatifolius*	-50.58	-29.45
2640	*pennatifolius*	-53.81	-25.68
95	*trachyllophus*	-42.50	-11.43
256	*trachyllophus*	-42.38	-13.80
257	*trachyllophus*	-42.60	-11.70
258	*trachyllophus*	-39.73	-10.65
259	*trachyllophus*	-43.23	-13.14
260	*trachyllophus*	-42.47	-14.06
261	*trachyllophus*	-42.50	-11.40
262	*trachyllophus*	-43.18	-13.12
263	*trachyllophus*	-41.66	-13.25
264	*trachyllophus*	-44.53	-14.18
265	*trachyllophus*	-42.50	-11.42
267	*trachyllophus*	-41.85	-13.23
269	*trachyllophus*	-41.82	-13.27
270	*trachyllophus*	-41.82	-13.28
271	*trachyllophus*	-41.80	-13.30
272	*trachyllophus*	-44.83	-15.45
273	*trachyllophus*	-42.70	-9.02
319	*trachyllophus*	-42.48	-14.07
320	*trachyllophus*	-44.36	-15.49
434	*trachyllophus*	-43.18	-13.12
435	*trachyllophus*	-42.42	-13.90
445	*trachyllophus*	-44.75	-15.51
580	*trachyllophus*	-44.53	-14.18
588	*trachyllophus*	-41.66	-13.25
600	*trachyllophus*	-42.49	-14.06
601	*trachyllophus*	-42.50	-11.42
710	*trachyllophus*	-42.70	-13.02
711	*trachyllophus*	-40.99	-3.73
712	*trachyllophus*	-45.16	-16.06
713	*trachyllophus*	-41.77	-12.42
714	*trachyllophus*	-40.93	-4.74
718	*trachyllophus*	-44.86	-15.95
723	*trachyllophus*	-45.24	-5.51
726	*trachyllophus*	-43.95	-10.14
727	*trachyllophus*	-43.33	-9.28
822	*trachyllophus*	-52.87	-10.83
901	*trachyllophus*	-44.96	-17.35
903	*trachyllophus*	-44.17	-13.45
904	*trachyllophus*	-42.61	-11.60
905	*trachyllophus*	-45.33	-5.67
906	*trachyllophus*	-45.10	-10.80
908	*trachyllophus*	-42.70	-13.02
909	*trachyllophus*	-41.84	-9.02

2641	*pennatifolius*	-54.09	-25.30
2642	*pennatifolius*	-52.42	-25.41
2643	*pennatifolius*	-49.52	-27.06
2731	*pennatifolius*	-50.71	-24.10
2737	*pennatifolius*	-52.44	-24.73
2769	*pennatifolius*	-51.70	-22.69
2772	*pennatifolius*	-49.07	-24.06
2785	*pennatifolius*	-49.24	-25.44
2819	*pennatifolius*	-50.85	-29.31
2824	*pennatifolius*	-53.28	-26.46
2832	*pennatifolius*	-53.97	-25.31
2834	*pennatifolius*	-53.99	-25.24
2839	*pennatifolius*	-54.11	-25.30
2841	*pennatifolius*	-53.96	-25.31
2848	*pennatifolius*	-49.85	-27.20
2851	*pennatifolius*	-52.30	-24.10
2852	*pennatifolius*	-53.87	-25.39
2861	*pennatifolius*	-49.09	-26.88
2867	*pennatifolius*	-52.68	-29.60
2875	*pennatifolius*	-49.47	-28.75
2877	*pennatifolius*	-47.06	-22.91
2900	*pennatifolius*	-56.64	-21.26
2910	*pennatifolius*	-48.68	-23.94
2933	*pennatifolius*	-50.77	-19.15
2936	*pennatifolius*	-50.81	-27.49
2937	*pennatifolius*	-50.91	-23.07
2965	*pennatifolius*	-50.76	-28.20
2970	*pennatifolius*	-57.15	-25.89
2982	*pennatifolius*	-53.34	-25.21
2984	*pennatifolius*	-53.67	-25.51
2995	*pennatifolius*	-49.42	-28.60
3006	*pennatifolius*	-50.40	-24.56
3007	*pennatifolius*	-46.08	-14.53
3012	*pennatifolius*	-49.37	-28.68
3015	*pennatifolius*	-53.19	-24.53
3025	*pennatifolius*	-53.88	-27.13
3031	*pennatifolius*	-47.04	-22.80
3051	*pennatifolius*	-52.35	-24.03
3053	*pennatifolius*	-53.87	-26.33
3054	*pennatifolius*	-53.88	-27.07
3055	*pennatifolius*	-56.23	-28.21
3089	*pennatifolius*	-54.41	-25.88
3092	*pennatifolius*	-52.35	-24.03
930	*trachyllophus*	-43.36	-9.99
963	*trachyllophus*	-46.55	-12.91
1051	*trachyllophus*	-42.37	-11.76
1052	*trachyllophus*	-42.28	-13.58
1053	*trachyllophus*	-41.85	-13.23
1054	*trachyllophus*	-42.13	-11.33
1055	*trachyllophus*	-42.50	-11.42
1335	*trachyllophus*	-50.83	-17.17
1457	*trachyllophus*	-43.93	-14.76
1711	*trachyllophus*	-43.14	-11.09
1863	*trachyllophus*	-42.61	-11.62
1865	*trachyllophus*	-41.82	-13.28
1866	*trachyllophus*	-42.89	-12.32
1867	*trachyllophus*	-43.56	-10.83
1868	*trachyllophus*	-42.38	-13.80
1869	*trachyllophus*	-43.44	-9.02
1870	*trachyllophus*	-43.17	-8.88
1871	*trachyllophus*	-43.44	-8.90
1876	*trachyllophus*	-43.23	-13.14
1879	*trachyllophus*	-43.22	-9.49
1881	*trachyllophus*	-42.53	-11.47
1883	*trachyllophus*	-42.61	-11.60
1884	*trachyllophus*	-39.73	-10.65
1995	*trachyllophus*	-42.61	-11.62
2044	*trachyllophus*	-41.82	-13.28
2183	*trachyllophus*	-42.37	-11.76
2205	*trachyllophus*	-42.89	-12.32
2281	*trachyllophus*	-43.56	-10.83
2297	*trachyllophus*	-42.38	-13.80
2305	*trachyllophus*	-43.44	-9.02
2319	*trachyllophus*	-43.17	-8.88
2370	*trachyllophus*	-50.83	-17.17
2390	*trachyllophus*	-43.44	-8.90
2446	*trachyllophus*	-42.13	-11.33
2471	*trachyllophus*	-43.18	-13.12
2475	*trachyllophus*	-46.55	-12.91
2615	*trachyllophus*	-42.48	-13.60
2781	*trachyllophus*	-43.23	-13.14
2783	*trachyllophus*	-44.17	-13.45
2784	*trachyllophus*	-43.36	-9.99
2858	*trachyllophus*	-42.61	-11.60
2862	*trachyllophus*	-42.28	-13.58
2864	*trachyllophus*	-43.22	-9.49

3101	*pennatifolius*	-51.39	-26.91
3116	*pennatifolius*	-51.30	-26.91
3119	*pennatifolius*	-55.34	-27.16
3120	*pennatifolius*	-55.92	-27.48
3121	*pennatifolius*	-55.59	-27.44
3131	*pennatifolius*	-51.61	-27.35
3148	*pennatifolius*	-58.39	-23.24
3149	*pennatifolius*	-57.29	-25.40
3150	*pennatifolius*	-56.02	-25.47
3158	*pennatifolius*	-54.83	-27.31

2892	*trachyllophus*	-43.36	-10.08
2896	*trachyllophus*	-45.33	-5.67
2912	*trachyllophus*	-42.53	-11.47
2959	*trachyllophus*	-42.38	-13.80
2992	*trachyllophus*	-46.55	-12.76
3003	*trachyllophus*	-45.10	-10.80
3036	*trachyllophus*	-39.73	-10.65
3040	*trachyllophus*	-46.12	-14.53
3047	*trachyllophus*	-42.26	-13.55
3073	*trachyllophus*	-42.70	-9.02
3156	*trachyllophus*	-44.10	-19.69

APPENDIX 5
WorldClim Bioclimatic preferences for each species

Species	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
P. alatus	27.14	10.44	76.40	670	34.24	20.58	13.66	26.72	27.46	28.10
P. carajaensis	25.11	11.34	79.81	388	32.37	18.13	14.23	24.84	25.13	25.56
P. demerarae	26.87	9.40	78.00	625	33.20	21.17	12.03	26.53	27.40	27.73
P. giganteus	22.32	8.56	58.42	1892	29.32	14.77	14.55	24.26	20.05	24.68
P. grandiflorus	24.04	7.68	63.36	1294	30.00	17.97	12.03	24.51	23.09	25.60
P. jaborandi	24.07	11.34	74.50	977	31.81	16.69	15.12	24.08	23.86	25.17
P. manuensis	25.65	10.83	76.33	591	32.17	18.10	14.07	25.82	24.73	26.18
P. microphyllus	26.00	10.05	76.11	698	32.77	19.51	13.26	25.78	26.18	26.88
P. pauciflorus	20.91	10.61	61.01	2222	29.00	11.70	17.30	23.12	18.35	23.50
P.pennatifolius	19.83	11.63	57.05	2997	29.69	9.47	20.22	21.61	17.26	23.40
P. peruvianus	25.50	12.14	73.63	691	33.31	16.85	16.46	25.78	24.69	26.14
P. racemosus	24.35	10.34	71.36	1113	31.67	17.19	14.48	24.65	24.09	25.62
P. riedelianus	23.63	9.69	68.52	1320	30.52	16.33	14.19	24.38	22.51	25.09
P. spicatus	22.50	10.13	68.17	1391	29.69	14.94	14.74	23.41	21.29	23.97
P. sulcatus	22.26	11.41	71.68	1192	29.81	13.99	15.83	23.01	20.67	23.30
P.trachyllophus	23.40	12.77	72.81	1083	31.78	14.34	17.44	23.92	22.08	24.52

Species	B11	B12	B13	B14	B15	B16	B17	B18	B19
P. alatus	26.40	1517	340.60	11.20	90.60	884	40	73	410
P. carajaensis	24.58	1925	318.92	18.92	67.00	895	85	259	713
P. demerarae	26.13	1854	349.00	45.00	72.67	920	152	157	465
P. giganteus	19.81	1683	237.12	59.73	44.96	669	199	599	211
P. grandiflorus	22.25	1586	198.62	78.81	32.83	550	265	408	354
P. jaborandi	22.75	1084	238.56	6.06	88.17	620	27	165	380
P. manuensis	24.73	1844	262.17	39.17	51.33	721	151	584	191
P. microphyllus	25.16	1663	311.43	24.80	74.82	844	89	166	573
P. pauciflorus	17.88	1448	222.90	45.82	50.99	616	163	541	208
P.pennatifolius	15.87	1578	189.82	81.84	27.15	512	283	477	307
P. peruvianus	24.45	2045	315.75	20.75	62.00	898	88	529	150
P. racemosus	22.81	1823	303.26	29.60	64.13	797	110	332	331
P. riedelianus	21.77	1735	254.49	52.57	49.39	689	184	442	335
P. spicatus	20.48	1059	170.61	29.21	57.63	460	106	333	146
P. sulcatus	20.44	869	175.41	4.59	83.45	468	21	338	26
P.trachyllophus	21.81	905	178.45	3.58	84.79	482	15	254	42

APPENDIX 6
Species Field Collections: Herbarium sheet references and silica records for individuals in the study

Genus	Species	Site #	Location	Daniella Allevato Silica Collection Number	Herbarium	Herbarium Sheet Collection Number
Pilocarpus	*pennatifolius*	A	Foz do Iguaçu 1	5	SP	Coelho, R.L.G.; Allevato, D. 642
Pilocarpus	*pennatifolius*	A	Foz do Iguaçu 2	6	SP	Coelho, R.L.G.; Allevato, D. 643
Pilocarpus	*pennatifolius*	A	Foz do Iguaçu 3	7	SP	Coelho, R.L.G.; Allevato, D. 644
Pilocarpus	*pennatifolius*	B	Campo Mourao	75	BH, SPFR	Allevato 75
Pilocarpus	*pennatifolius*	B	Campo Mourao	76	BH, SPFR	Allevato 76
Pilocarpus	*pennatifolius*	B	Campo Mourao	77	BH, SPFR	Allevato 77
Pilocarpus	*pennatifolius*	B	Campo Mourao	78	BH, SPFR	Allevato 78
Pilocarpus	*pennatifolius*	B	Campo Mourao	79	BH, SPFR	Allevato 79
Pilocarpus	*pennatifolius*	C	Campo Mourao B	80	BH, SPFR	Allevato 80
Pilocarpus	*pennatifolius*	C	Campo Mourao B	81	BH, SPFR	Allevato 81
Pilocarpus	*pennatifolius*	C	Campo Mourao B	82	BH, SPFR	Allevato 82
Pilocarpus	*pennatifolius*	C	Campo Mourao B	83	BH, SPFR	Allevato 83
Pilocarpus	*pennatifolius*	D	Lago Azul	95	BH, SPFR	Allevato 95
Pilocarpus	*pennatifolius*	D	Lago Azul	96	BH, SPFR	Allevato 96
Pilocarpus	*pennatifolius*	D	Lago Azul	97	BH, SPFR	Allevato 97
Pilocarpus	*pennatifolius*	D	Lago Azul	98	BH, SPFR	Allevato 98
Pilocarpus	*pennatifolius*	D	Lago Azul	99	BH, SPFR	Allevato 99
Pilocarpus	*pennatifolius*	D	Lago Azul	100	BH, SPFR	Allevato 100
Pilocarpus	*pennatifolius*	D	Lago Azul	101	BH, SPFR	Allevato 101
Pilocarpus	*pennatifolius*	D	Lago Azul	102	BH, SPFR	Allevato 102
Pilocarpus	*pennatifolius*	D	Lago Azul	103	BH, SPFR	Allevato 103
Pilocarpus	*pennatifolius*	E	Cruz do Pedro	110	SPFR	Groppo 2300-A
Pilocarpus	*pennatifolius*	E	Cruz do Pedro	111	SPFR	Groppo 2300-B

Pilocarpus	*pennatifolius*	E	Cruz do Pedro	112	SPFR	Groppo 2300-C
Pilocarpus	*pennatifolius*	E	Cruz do Pedro	113	SPFR	Groppo 2300-D
Pilocarpus	*pennatifolius*	E	Cruz do Pedro	114	SPFR	Groppo 2300-E
Pilocarpus	*pennatifolius*	F	Estacao Ecologica St. Theresa	115	SPFR	Groppo 2302-A
Pilocarpus	*pennatifolius*	F	Estacao Ecologica St. Theresa	116	SPFR	Groppo 2302-B
Pilocarpus	*pennatifolius*	F	Estacao Ecologica St. Theresa	117	SPFR	Groppo 2302-C
Pilocarpus	*pennatifolius*	F	Estacao Ecologica St. Theresa	118	SPFR	Groppo 2302-D
Pilocarpus	*pennatifolius*	F	Estacao Ecologica St. Theresa	119	SPFR	Groppo 2302-E

APPENDIX 7

Soil analysis; Average of two soil samples for each site analysed by Instituto Agronômico in Campinas, SP, Brazil

	M.O.	pH	P	K	Ca	Mg	Na	Al	H + Al	S.B.	C.E.C.
A	74	6.4	37	5.9	201	27	0.6	0	16	234.5	250.5
B	89	4.35	100.5	2.75	83.5	6.5	0.3	3	139	93.05	232.05
C	98	5.8	66	3.05	362	38.5	0.3	1	30	403.85	433.85
D	96	5.8	57	3.4	328	35	0.3	1	29	366.7	395.7
E	60	5.5	32	2.7	73	18	0.2	0	29	93.9	122.9
F	72	6.8	186	4.6	246	36	0.2	0	11	286.8	297.8

	V%	S	B	Cu	Fe	Mn	Zn	EC	N
A	94	0	0.52	4.9	18	42.6	2.8	1.3	3.6
B	39.5	18.5	0.605	1.95	150.5	22.85	13.55	1.1	16.4
C	92.5	29	0.815	2.45	72.5	29	7.8	0.9	16.75
D	93	31	0.54	1.4	56	43.6	18	1.5	16.7
E	76	6	0.27	5.5	22	27.4	1.3	0.7	3.3
F	96	16	0.43	3.4	24	9.5	8.1	2.2	5.5

APPENDIX 8

Bioclimatic variables extracted from WorldClim and altitude from DIVA GIS

	Alt	Annual Mean Temp	Mean Diurnal Range	Isothermality	Temp Seasonality	Max Temp Warmest Month	Min Temp Coldest Month	Temp Annual Range	Mean Temp Wettest Qtr	Mean Temp Driest Qtr
A	172	21.4958	12.6583	55.5190	359.7882	32.5	9.7	22.8	23.7333	17.0667
B	554	20.2167	11.2500	59.2105	291.0118	29.1	10.1	19	23.5000	16.6500
C	554	20.2167	11.2500	59.2105	291.0118	29.1	10.1	19	23.5000	16.6500
D	585	20.0667	11.2333	59.4356	290.2689	28.9	10	18.9	23.3500	16.5167
E	635	21.2167	12.2333	67.5875	204.7097	28.6	10.5	18.1	23.1333	18.4333
F	611	21.3500	12.3333	67.3953	203.0114	28.7	10.4	18.3	23.2333	18.5833

	Mean Temp Warmest Qtr	Mean Temp Coldest Qtr	Annual Precip	Precip Wettest Month	Precip Driest Month	Precip Seasonality	Precip Wettest Qtr	Precip Driest Qtr	Precip Warmest Qtr	Precip Coldest Qtr
A	25.7000	17.0667	1729	189	97	17.6189	503	331	463	376
B	23.5000	16.5000	1536	188	78	27.6495	516	276	516	317
C	23.5000	16.5000	1536	188	78	27.6495	516	276	516	317
D	23.3500	16.3833	1555	194	79	28.0923	527	279	527	318
E	23.1333	18.3333	1453	257	23	75.6499	732	81	732	101
F	23.2333	18.4667	1486	265	24	74.3092	742	88	742	108

APPENDIX 9
Species ID for Population Study

Silica ID	ddRAD ID	Pop ID	Species	Sub-species	Location	State
5	1	A1	*P. pennatifolius*		Foz do Iguaçu 1	PR
6	2	A2	*P. pennatifolius*		Foz do Iguaçu 2	PR
7	3	A3	*P. pennatifolius*		Foz do Iguaçu 3	PR
75	7	B1	*P. pennatifolius*		Campo Mourao	PR
76	8	B2	*P. pennatifolius*		Campo Mourao	PR
77	9	B3	*P. pennatifolius*		Campo Mourao	PR
78	10	B4	*P. pennatifolius*		Campo Mourao	PR
79	11	B5	*P. pennatifolius*		Campo Mourao	PR
80	12	C1	*P. pennatifolius*		Campo Mourao B	PR
81	13	C2	*P. pennatifolius*		Campo Mourao B	PR
82	14	C3	*P. pennatifolius*		Campo Mourao B	PR
83	15	C4	*P. pennatifolius*		Campo Mourao B	PR
95	26	F1	*P. pennatifolius*		Lago Azul	PR
96	27	F2	*P. pennatifolius*		Lago Azul	PR
97	28	F3	*P. pennatifolius*		Lago Azul	PR
98	29	F4	*P. pennatifolius*		Lago Azul	PR
99	30	F5	*P. pennatifolius*		Lago Azul	PR
100	31	F6	*P. pennatifolius*		Lago Azul	PR
101	32	F7	*P. pennatifolius*		Lago Azul	PR
102	33	F8	*P. pennatifolius*		Lago Azul	PR
103	34	F9	*P. pennatifolius*		Lago Azul	PR
110	16	D1	*P. pennatifolius*		Cruz do Pedro	SP
111	17	D2	*P. pennatifolius*		Cruz do Pedro	SP
112	18	D3	*P. pennatifolius*		Cruz do Pedro	SP
113	19	D4	*P. pennatifolius*		Cruz do Pedro	SP
114	20	D5	*P. pennatifolius*		Cruz do Pedro	SP
115	21	E1	*P. pennatifolius*		E. E. St. Theresa	SP
116	22	E2	*P. pennatifolius*		E. E. St. Theresa	SP
117	23	E3	*P. pennatifolius*		E. E. St. Theresa	SP
118	24	E4	*P. pennatifolius*		E. E. St. Theresa	SP
119	25	E5	*P. pennatifolius*		E. E. St. Theresa	SP

Silica ID	ddRAD ID	Pop ID	Species	Sub-species	Location	State
84	101	I1	*P. spicatus*	longeracemosus	Rio de Contas	BA
85	102	I2	*P. spicatus*	longeracemosus	Rio de Contas	BA
86	103	J1	*P. spicatus*	aracatensis	Rio de Contas	BA
87	104	J2	*P. spicatus*	aracatensis	Rio de Contas	BA
88	105	J3	*P. spicatus*	aracatensis	Rio de Contas	BA
89	106	J4	*P. spicatus*	aracatensis	Rio de Contas	BA

90	107	I3	*P. spicatus*	longeracemosus	Rio de Contas	BA
91	108	I4	*P. spicatus*	longeracemosus	Rio de Contas	BA
92	109	J5	*P. spicatus*	aracatensis	Rio de Contas	BA
93	110	I5	*P. spicatus*	longeracemosus	Rio de Contas	BA
94	111	J6	*P. spicatus*	aracatensis	Rio de Contas	BA
105	96	H1	*P. spicatus*	spicatus	Cruz do Pedro	SP
106	97	H2	*P. spicatus*	spicatus	Cruz do Pedro	SP
107	98	H3	*P. spicatus*	spicatus	Cruz do Pedro	SP
108	99	H4	*P. spicatus*	spicatus	Cruz do Pedro	SP
109	100	H5	*P. spicatus*	spicatus	Cruz do Pedro	SP
120	91	G1	*P. spicatus*	spicatus	Santa Rita	SP
121	92	G2	*P. spicatus*	spicatus	Santa Rita	SP
122	93	G3	*P. spicatus*	spicatus	Santa Rita	SP
123	94	G4	*P. spicatus*	spicatus	Santa Rita	SP
124	95	G5	*P. spicatus*	spicatus	Santa Rita	SP

Silica ID	**ddRAD ID**	**Pop ID**	**Species**	**Sub-species**	**Location**	**State**
11	67	K1	*P. riedelianus*		Entre Rios	BA
12	68	K2	*P. riedelianus*		Entre Rios	BA
13	69	K3	*P. riedelianus*		Entre Rios	BA
14	70	K4	*P. riedelianus*		Entre Rios	BA
15	71	K5	*P. riedelianus*		Entre Rios	BA
26	72	L1	*P. riedelianus*		Linhares	ES
27	73	L2	*P. riedelianus*		Linhares	ES
28	74	L3	*P. riedelianus*		Linhares	ES
29	75	L4	*P. riedelianus*		Linhares	ES
30	76	L5	*P. riedelianus*		Linhares	ES
31	77	L6	*P. riedelianus*		Linhares	ES
32	78	L7	*P. riedelianus*		Linhares	ES
33	79	L8	*P. riedelianus*		Linhares	ES
34	80	L9	*P. riedelianus*		Linhares	ES
35	81	L10	*P. riedelianus*		Linhares	ES
66	82	M1	*P. riedelianus*		Morro do Chapeu	BA
67	83	M2	*P. riedelianus*		Morro do Chapeu	BA
68	84	M3	*P. riedelianus*		Morro do Chapeu	BA
69	85	M4	*P. riedelianus*		Morro do Chapeu	BA
70	86	M5	*P. riedelianus*		Morro do Chapeu	BA
71	87	M6	*P. riedelianus*		Morro do Chapeu	BA
72	88	M7	*P. riedelianus*		Morro do Chapeu	BA
73	89	M8	*P. riedelianus*		Morro do Chapeu	BA
74	90	M9	*P. riedelianus*		Morro do Chapeu	BA

*** Silica ID is the field collection ID number.*

*** ddRAD ID is the ID given for sequencing of the samples (numbers displayed on MSN diagrams)*

APPENDIX 10
Flow Cytometry to determine *Pilocarpus* genome size for sequencing

Flow Cytometry was conducted using an Accuri C6 flow cytometer in Adrienne Roeder's Laboratory at Cornell University. Leaf tissue [*Pilocarpus jaborandi* seedlings from USBG and *Pilocarpus pennatifolius* mature plants from NYBG] was chopped in LB01 buffer (Doležel *et al.*, 1989). Next, samples were filtered through a 42um nylon mesh, and nuclei were stained with a Propidium Iodide-RNase solution (50ug/ml solution) (Doležel *et al.*, 2007). Internal DNA reference standards were provided by Jaroslav Doležel at the Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Czech Republic. Samples were run with three references: *Glycine max* Merr. 'Polanka' (2C=2.50pg), *Zea mays* L. 'CE-777' (2C= 5.43pg), and *Pisum sativum* L. 'Citrad' (2C=9.09pg) (Doležel and Bartoš, 2005). Estimation of nuclear DNA content (genome size), was calculated by comparing the mean peak position of *Pilocarpus* with that of the three references using the following equation: Sample 2C (pg or Mbp) = Reference 2C value x (Sample 2C mean peak position / Reference 2C mean peak position) (Doležel *et al.*, 2007). The mean of the genome size calculations for these diploid *Pilocarpus* samples was 2C=5.055pg. The genome size for *Pilocarpus* is above the mean for genera in Rutaceae (2C: mean=2.61, median=1.18, min=0.4, max=17.4), according to data retrieved from the Kew DNA C-values database (Bennett and Leitch).

LITERATURE CITED

Bennett M, Leitch I. Angiosperm DNA C-values database (release 8.0, Dec.2012).

Doležel J, Bartoš J. 2005. Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. Annals of Botany **95**, 99–110.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols **2**, 2233–2244.

Doležel J, Binarová P, Lcretti S. 1989. Analysis of Nuclear DNA content in plant cells by Flow cytometry. Biologia Plantarum **31**, 113–120.

***Pilocarpus spicatus* inflorescence with insect visitor, Rio de Contas, Bahia, Brazil**
Photo credit: Daniella M. Allevato