SPATIAL, TEMPORAL, AND GENOME-WIDE PATTERNS OF DIFFERENTIATION FOR THE LOUISIANA IRIS SPECIES COMPLEX

by

JENNAFER ASHLEY PARKER HAMLIN

(Under the Direction of Michael L. Arnold)

ABSTRACT

The classic mode of allopatric speciation emphasizes geographical isolation and assumes ecological factors as insignificant, whereas now there is renewed interest in highlighting the importance of ecology during speciation. Ecological speciation, as a by-product of local adaption, results when barriers to gene flow evolve between populations as an outcome of ecologically-based divergent natural selection. However, gene flow tends to counteract evolutionary forces, such as natural selection, because gene flow constrains population differentiation via genetic exchange. Thus, work on young incipient species allows one to disentangle the influence of gene flow, ecological differentiation, and local adaptation during the early stages of divergence. The Louisiana irises are a species complex of relatively recent origin and here I focus on the three species that are morphologically distinct, widespread, and interfertile: *Iris brevicaulis*, *Iris fulva*, and *Iris hexagona*. I found that there are species-level differences in connectivity among populations of Louisiana irises; in that bumblebee pollinated species (*I. brevicaulis* and *I. hexagona*) are more restricted via intraspecific gene flow than *I. fulva*, which is pollinated by hummingbirds. Moreover, gene flow was detected in populations found in the southern parts of the range and occurred from *I. fulva* into *I. brevicaulis*. Furthermore, it appears that both selection and neutral processes are at play in generating population divergence for *I. hexagona* in terms of floral trait differences. This implies that the potential for divergent selection and local adaptation, in terms of pollinators, can promote diversification within a single lineage. Differences observed between species, with regard to associated environmental factors, suggest an effect from these components on the distributions and habitats occupied. Furthermore, niche divergence was implicated in all pairwise comparisons regardless of range overlap and niches remained differentiated after lineage diversification. Thus it appears that even in the face of gene flow, niche divergence can be maintained and that ecological speciation is the likely speciation mechanism for the Louisiana irises.

INDEX WORDS: Louisiana irises, ecological speciation, hybridization, population genetics, local adaptation, niche divergence

SPATIAL, TEMPORAL, AND GENOME-WIDE PATTERNS OF DIFFERENTIATION FOR THE LOUISIANA IRIS SPECIES COMPLEX

by

JENNAFER ASHLEY PARKER HAMLIN

BS, University of North Carolina – Asheville, 2010

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

SPATIAL, TEMPORAL, AND GENOME-WIDE PATTERNS OF

DIFFERENTIATION FOR THE LOUISIANA IRIS SPECIES COMPLEX

by

JENNAFER ASHLEY PARKER HAMLIN

Major Professor: Michael L. Arnold

Committee: Bryan Carstens
Travis Glenn
Rodney Mauricio
John Wares

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

DEDICATION

I would like to dedicate this dissertation to my Nana and Papa, who imparted a love of both puzzles and plants.

And

I would like to dedicate this dissertation to my mother, for her unwavering support.

ACKNOWLEDGEMENTS

First, I would like to acknowledge and thank my advisor Michael Arnold for his support, advice, and ever-quick response via email. The members of my committee (John Wares, Rodney Mauricio, Travis Glenn, and Bryan Carstens) have been very generous with their time and expertise. My advancement as a student of ecology, evolution, and genetics is a direct product of their commitment to my education.

I, particularly, want to acknowledge the members of the Arnold, Sweigart, and Wares labs. First, I want to thank both Katie Bockrath and Christine Ewers-Saucedo for supporting me professional and for our friendships over the years. I would, also, like to acknowledge all of the members of the Wares lab, both past and present, for providing a space to connect and communicate about science and non-science. I would like to acknowledge all the members of the Sweigart lab, both past and present, who have provided invaluable feedback regarding many aspects of this work and helped sharpen my critical thinking skills. Most especially to Amanda Kenney during our walks. I would be remiss to not mention three of my biggest influences during my graduate career: Anji Ballerini, Mandy Brothers, and Ron Eytan. I am indebted for all their thoughtful guidance and support especially as a new student.

And lastly, I would like to thank Brendan Cornett, Kerry Steinberg, and Mar Costa-Hurtado for encouraging me through the final stages of this dissertation.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Driven by rapid improvements in computational resources and accumulation of fine-scale individual genetic data, speciation research has expanded beyond traditional boundaries (Wolf et al. 2010; Supple et al. 2014). This surge of data allows one to work at the interface of genomes and biological complexity with the goal of understanding the origin of biodiversity (Supple et al. 2014). The question of how do new species arise is still at the heart of what evolutionary biologists seek to address but, now, with new levels of complexity. However, to understand the generality and action of evolutionary processes, speciation research must continue to encompass a greater diversity of organisms and at different stages along the speciation continuum. For example, a single species with highly differentiated populations may represent the early stages of the speciation process (Nosil 2012). And in such cases, we may we may not know if speciation will go to completion, but nonetheless still learn about the process.

Population structure via geographical subdivision is an inescapable fact of biology as many landscape features (e. g. deserts, mountains, and rivers) separate potentially interbreeding populations (Manel et al. 2003). Ultimately, this could lead to complete reproductive isolation and is historically considered the most common mode of speciation. The strict definition of allopatric speciation is one in which reproductive isolation evolves as a byproduct of divergent evolution in geographically isolated populations (Mayr 1942). Even without major barriers to gene flow, our landscape is not homogenous and organisms do not disperse randomly across their range, resulting in potentially low levels of gene flow between populations (Slatkin 1987; Manel et al. 2003). With low levels of genetic exchange, divergence can happen even in the absence of selection, as local allele frequencies change through genetic drift. Wright (1951) described this population genetic pattern as 'isolation by distance', with expectations that differ for the history and future of populations isolated by distance and locally adapted populations.

Most often local adaptation is a result of phenotypic variation that aligns with environmental variation (e.g. flowering time differences; Stinchcombe et al. 2004), where natural selection favors certain traits in different environments (Lenormand 2002). There is a large body of work, which supports the ubiquitous nature of local adaptation in natural populations via reciprocal transplant studies and common-garden experiments (Hereford 2009), where individuals demonstrate a pattern of higher fitness in their local environment relative to foreign individuals (Kawecki and Ebert 2004). Furthermore, evolutionary models predict that genetic and phenotypic differences can accumulate between these populations (Slatkin 1987) and this differentiation between local populations can represent the early stages of speciation (Papadopulos et al. 2014).

We know that speciation is a complex process, which reflects a culmination of various factors such as: assortment of ancestral variation, ecological interactions, dispersal limitations, and in some cases introgression (Brandvain et al. 2014). In the last two decades, the process of speciation is increasingly recognized to be associated with environmental and ecological differences between populations, thus constraining reproduction and migration. But the extent to which species formation is driven by ecology and the possible role of gene flow during this process is still debated (Schluter 2000; Coyne and Orr 2004; Mallet 2009). Darwin (1859) was a clear proponent that environmental differences can generate divergent selection and eventually drive two populations apart (Wolf et al. 2010) and now, many consider ecology to be an important component during speciation (Schluter 2001, 2009; Nosil 2012).

Ecological speciation is broad in its definition (Rundle and Nosil 2005) and encompasses all instances whereby reproductive isolation can evolve as a by-product of adaptation to different environments (Schluter 2001, 2009; Rundle & Nosil 2005). Studies of young sympatric species, generally implicate a role of ecological speciation (Rundell and Price 2009), as seen in work on adaptive radiations (Schluter 2000). Establishing the crucial link between adaptation and reproductive isolation is challenging and two key conditions must be verified before concluding that any given pair of species originated through ecological speciation (Faria et al. 2014). Differentiation must be driven by divergent natural selection and reproductive isolation must evolve as a consequence of this selection (Nosil 2012; Faria et al. 2014). For selection to be considered ecological, it must arise as a consequence of individual interactions with their external environment (Nosil 2012) and selection is considered divergent when it acts in contrasting directions in two populations and could be induced via environmental differences, sexual selection, or ecological interactions (Rundle and Nosil 2005). Under this process, when divergent selection acts, individuals exploit ecological or environmental differences with the associated changes in certain key traits. Such traits could be those involved in defense, behavioral differences, resource use, or abiotic interactions; however, this is not an exhaustive list (Nosil 2012). Moreover, divergent selection and local adaptation has been implicated in a number of systems (Via et al. 2000; Hoekstra et al. 2005; Nosil et al. 2005; Roesti et al. 2012; Anderson et al. 2014) and some consider local adaptation as the first step in the process of ecological speciation (Lenormand 2012). The evolution of reproductive isolation via divergent selection could be caused by a variety of pre- or post-zygotic forms of isolation such as: habitat and temporal isolation, selection against immigrants, sexual isolation, intrinsic post-zygotic isolation, and ecologically-dependent isolation (Rundle and Nosil 2005 for review). When divergent selection acts and results in a reproductive isolating barrier, one must determine the genetic mechanism linking the two to conclude ecological speciation and numerous studies have dissected the genetic basis of ecological speciation (Rogers and Bernatchez 2007; Via and West 2008; Chamberlain et al. 2009; Nosil and Schluter 2011). For example, empirical work has shown that traits evolving via divergent selection and that confer reproductive isolation can be affected by many or few genes, which may be of small or large effect, and that vary in dominance and epistatic interactions (Rundle and Nosil 2005). Furthermore, there are two genetic mechanisms that link divergent selection with reproductive isolation: pleiotropy or linkage disequilibrium, which act either directly or indirectly, respectively, and with different consequences. And if linkage disequilibrium is implicated, via ecological speciation, than one could ask: *How do appropriate trait combinations, which are not in linkage disequilibrium, remain associated in the face of gene flow?*

One factor that can affect the trajectory of speciation is gene flow between species. Traditionally, hybridization or gene flow was viewed as a local phenomenon with little evolutionary importance (Mayr 1963). Furthermore, gene flow is expected to act as a homogenizing force between populations. At the other extreme, hybridization is considered a major source of adaptive genetic variation (Arnold 2006), which can shape a species. A general assumption is that when species are sympatric and reproductive isolation is incomplete, hybridization and the potential for introgression can occur (Fontaine et al. 2015). There are a number of evolutionary outcomes when incompletely isolated taxa hybridize. For example, hybridization may result in the production of a novel species (Seehausen 2004), extinction of native species via genetic exchange with non-native species (Rhymer and Simberloff 1996), or secondary contact might result in the reinforcement of pre-zygotic reproductive barriers (Howard 1993; Servedio and Noor 2003). Thus, hybridization has many and varied impacts on the process of speciation (Abbott et al. 2013).

Furthermore, species vary widely in the extent to which they experience gene flow (Slatkin 1985) and gene flow can also vary spatially and temporally, among populations and between individuals within the same populations. Thus, introgression can be difficult to detect and establishing the frequency and effect of gene flow is usually done within closely related species where reproductive isolation is incomplete. Conversely, ecological speciation can occur in the face of gene flow as has been documented in a number of systems (Papadopulos et al. 2011; Via 2012; Chapman et al. 2013). When gene flow is implicated for diverging lineages, one must elucidate if and what ecological differences caused divergent or disruptive selection (Feder et al. 2012). Moreover, it is predicted that taxa can be differentiated only at a small number of locally confined areas in the genome and these regions may be under strong divergent selection, which contribute to local adaptation or possibly reproductive isolation (Feder et al. 2012). Thus, making the connection between divergent selection and reproductive isolation can be done in the face of gene flow.

Within the plant literature, pollination is a critical ecological factor and interactions with animal pollinators can be considered a source of divergent selection (Campbell 2004; Van der Niet et al. 2014). However, pollinators can be promiscuous and thus result in gene flow between species where reproductive isolation has not gone to completion for additional pre- and post-zygotic barriers. Moreover, the evolution of a before-mating reproductive barrier, such as pollinator differences, is easy to evolve (Turelli et al. 2001); however, no single evolutionary mechanism plays the primary role in lineage diversification. Rather complete reproductive isolation requires divergence in multiple genes/traits and generally occurs over long periods of time (Mendelson et al. 2007). Thus, the production of a first generation hybrid could change the trajectory of a population or species even in the face of divergent natural selection.

This intersection of species interactions via gene flow, environmental differences, and local adaptation allows one to elucidate the process of differentiation within an ecological and evolutionary framework. My dissertation work aims to understand the factors shaping spatial, temporal, and genome-wide patterns of differentiation within a species complex: the Louisiana irises. This clade is composed of four species: *Iris brevicaulis, I. fulva, I. hexagona*, and *I. nelsonii*, which have been the focus of numerous studies especially in the context of hybridization and adaptive introgression (Arnold et al. 2012). My work focuses on the three species which are morphologically distinct, widespread, and interfertile. Two of the species, *I. brevicaulis* and *I. fulva*, are broadly sympatric throughout the Mississippi River valley but exhibit habitat partitioning as demonstrated when species are found in geographically close proximity (Cruzan and Arnold 1993; Johnston et al. 2001). Both species occur in or near wetland systems; where *I. fulva* typically grows at lower elevations along bayou edges with roots often submerged in water, whereas *I. brevicaulis* generally occurs at slightly higher elevations in surrounding mixed hardwood forest (Viosca 1935; Cruzan and Arnold 1993; Johnston et al. 2001). Furthermore, floral characteristics are strikingly different for the two species. *Iris brevicaulis* floral morphology is typical of a bumblebee pollination syndrome; floral color varies from light to deep blue with marked nectar guides, stiff upright sepals, and strong scent (Cruzan and Arnold 1994; Martin et al. 2007). In contrast, *I. fulva* flowers are a deep crimson color, have protruding anthers, lack nectar guides, and produce large volumes of nectar (Viosca 1935; Wesselingh and Arnold 2000). Hummingbirds and butterflies pollinate these flowers. Furthermore, these two species differ in flowering time: *I. fulva* begin flowering about a month earlier than *I. brevicaulis* in when the species are sympatric in southern Louisiana (Cruzan and Arnold 1994). The third species, *I. hexagona*, is the most southerly occurring iris species in the United States with a coastal distribution and is found mostly in open, freshwater marshes and swamps (Meerow et al. 2011). The flower stalks of *I. hexagona* attain heights of one to two meters and are characterized by large purple/blue flowers and distinctive yellow nectar guides, and shares a primary pollinator, *Bombus* spp. bumblebees, with *I. brevicaulis* (Viosca 1935; Emms and Arnold 2000). Conversely, *I. hexagona* and *I. fulva* have almost synchronous flowering periods during the early spring (Arnold 1993; Brothers et al. 2013). Despite these differences in habitat, floral morphology, and phenology, hybrid zones between these species have been documented in southern Louisiana (Arnold et al. 1992; Arnold 1993; Cruzan and Arnold 1993; Johnston et al. 2001). The formation of first generation hybrid individuals is rare, however, these F_1 individuals are in fact both viable and fertile, resulting in widespread introgression between species within this complex (Arnold et al. 1990; Arnold et al. 1992). To that end, the Louisiana irises have been developed as a model system for understanding the genetic architecture of reproductive isolation in both wild and experimental hybrid zones; however the population dynamics of the system have not been explored outside of Louisiana. Chapters two through four describe multiple approaches utilized to understand the mechanisms driving divergence both at the population and species level in this species complex.

In my second chapter, I assess introgression and population genetic structure for two Louisiana irises, *Iris brevicaulis* and *I. fulva*, across the range of both species. I use estimates of population differentiation, as expressed by Wright's Fixation index, population assignment methods, and a species tree approach to infer the patterns and relationships within and between species. In addition, I address whether gene flow between populations is restricted across the Mississippi River. In my third chapter, I sought to identify factors driving intraspecific variation within *I hexagona* to illuminate how gene flow and selection may be interacting to promote or hinder divergence between populations. In particular, I examined the relationship between molecular, morphological, climate, and geographic differentiation for this species. Furthermore, I ask whether certain loci show greater than expected differentiation, possibly indicating regions under selection or those involved in local adaptation. In my fourth chapter, a species phylogeny was generated from chloroplast markers using *BEAST to infer the phylogenetic relationships within this clade. Ecological Niche Models were constructed for present day distributions, to better understand how environmental factors contribute to species divergence. Comparisons between niche models for each species were used to assess if sister species exhibit niche divergence and if that holds when compared to their most recent common ancestor, *I. hexagona.* Working at both the population and species level and examining various factors that can influence differentiation, this study has assessed a more whole picture of the ecological and evolutionary history of this species complex across their geographic range.

Literature Cited

Abbott, R., D. Albach, S. Ansell, J. Arntzen, S. Baird, N. Bierne, J. Boughman, A. Brelsford, C. Buerkle, and R. Buggs. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26:229-246.

Anderson, J. T., C. R. Lee, and T. Mitchell ‐ Olds. 2014. Strong selection genome-wide enhances fitness trade-offs across environments and episodes of selection. Evolution 68:16-31.

Arnold, M., J. Hamrick, and B. Bennett. 1990. Allozyme variation in Louisiana irises: a test for introgression and hybrid speciation. Heredity 65:297-306.

Arnold, M., J. Robinson, C. Buckner, and B. Bennett. 1992. Pollen dispersal and interspecific gene flow in Louisiana irises. Heredity 68:9.

Arnold, M. L. 1993. Iris nelsonii (Iridaceae): origin and genetic composition of a homoploid hybrid species. American Journal of Botany:577-583.

Arnold, M. L. 2006. Evolution Through Genetic Exchange. Oxford University Press, New York.

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. Heredity 108:159-166.

Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. 2014. Speciation and introgression between Mimulus nasutus and Mimulus guttatus. PLoS genetics 10:e1004410.

Brothers, A. N., J. G. Barb, E. S. Ballerini, D. W. Drury, S. J. Knapp, and M. L. Arnold. 2013. Genetic architecture of floral traits in Iris hexagona and Iris fulva. Journal of Heredity 104:853-861.

Campbell, D. R. 2004. Natural selection in Ipomopsis hybrid zones: implications for ecological speciation. New Phytologist 161:83-90.

Chamberlain, N. L., R. I. Hill, D. D. Kapan, L. E. Gilbert, and M. R. Kronforst. 2009. Polymorphic butterfly reveals the missing link in ecological speciation. Science 326:847-850.

Chapman, M. A., S. J. Hiscock, and D. A. Filatov. 2013. Genomic divergence during speciation driven by adaptation to altitude. Molecular biology and evolution:mst168.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates Sunderland, MA.

Cruzan, M. B. and M. L. Arnold. 1993. Ecological and genetic associations in an Iris hybrid zone. Evolution:1432-1445.

Cruzan, M. B. and M. L. Arnold. 1994. Assortative mating and natural selection in an Iris hybrid zone. Evolution:1946-1958.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection or the Preservaton of Favoured Races in the Struggle for Life. John Murray, London.

Emms, S. K. and M. L. Arnold. 2000. Site‐to‐site differences in pollinator visitation patterns in a Louisiana iris hybrid zone. Oikos 91:568-578.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo‐Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in ecological speciation: an integrative approach. Molecular ecology 23:513-521.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28:342-350.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, and E. Kakani. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347:1258524.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade‐offs. The American Naturalist 173:579-588.

Hoekstra, H. E., J. Krenz, and M. W. Nachman. 2005. Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. Heredity 94:217-228.

Howard, D. J. 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. Hybrid zones and the evolutionary process:46-69.

Johnston, J. A., R. A. Wesselingh, A. C. Bouck, L. A. Donovan, and M. L. Arnold. 2001. Intimately linked or hardly speaking? The relationship between genotype and environmental gradients in a Louisiana Iris hybrid population. Molecular Ecology 10:673-681.

Kawecki, T. J. and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology Letters 7:1225-1241.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17:183-189.

Lenormand, T. 2012. From local adaptation to speciation: specialization and reinforcement. International Journal of Ecology 2012.

Mallet, J. 2009. Rapid speciation, hybridization and adaptive radiation in the Heliconius melpomene group. Speciation and patterns of diversity:177-194.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2007. The genetic architecture of reproductive isolation in Louisiana irises: flowering phenology. Genetics 175:1803-1812.

Mayr, E. 1942. Systematics and the origin of species. Columbia University Press, New York.

Mayr, E. 1963. Animal species and evolution. Animal species and their evolution.

Meerow, A. W., M. Gideon, D. N. Kuhn, S. Mopper, and K. Nakamura. 2011. The genetic mosaic of iris series Hexagonae in Florida: inferences on the holocene history of the Louisiana irises and anthropogenic effects on their distribution. International Journal of Plant Sciences 172:1026-1052.

Mendelson, T. C., V. E. Imhoff, and J. J. Venditti. 2007. The accumulation of reproductive barriers during speciation: postmating barriers in two behaviorally isolated species of darters (Percidae: Etheostoma). Evolution 61:2596-2606.

Nosil, P. 2012. Ecological speciation. Oxford University Press.

Nosil, P. and D. Schluter. 2011. The genes underlying the process of speciation. Trends in Ecology & Evolution 26:160-167.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705-719.

Papadopulos, A. S., W. J. Baker, D. Crayn, R. K. Butlin, R. G. Kynast, I. Hutton, and V. Savolainen. 2011. Speciation with gene flow on Lord Howe Island. Proceedings of the National Academy of Sciences 108:13188-13193.

Papadopulos, A. S., M. Kaye, C. Devaux, H. Hipperson, J. Lighten, L. T. Dunning, I. Hutton, W. J. Baker, R. K. Butlin, and V. Savolainen. 2014. Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130342.

Rhymer, J. M. and D. Simberloff. 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics:83-109.

Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. Molecular Ecology 21:2852-2862.

Rogers, S. and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. Molecular Biology and Evolution 24:1423-1438.

Rundell, R. J. and T. D. Price. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution 24:394-399.

Rundle, H. D. and P. Nosil. 2005. Ecological speciation. Ecology letters 8:336-352.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323:737-741.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends in ecology & evolution 19:198-207.

Servedio, M. R. and M. A. Noor. 2003. The role of reinforcement in speciation: theory and data. Annual Review of Ecology, Evolution, and Systematics:339-364.

Slatkin, M. 1985. Gene flow in natural populations. Annual review of ecology and systematics:393-430.

Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural Populations. Science 236:787-792.

Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences of the United States of America 101:4712-4717.

Supple, M., R. Papa, B. Counterman, and W. O. McMillan. 2014. The Genomics of an Adaptive Radiation: Insights Across the Heliconius Speciation Continuum. Pp. 249-271. Ecological Genomics. Springer.

Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends in Ecology & Evolution 16:330-343.

Van der Niet, T., R. Peakall, and S. D. Johnson. 2014. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. Annals of botany 113:199-212.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philosophical Transactions of the Royal Society B: Biological Sciences 367:451-460.

Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. Evolution 54:1626-1637.

Via, S. and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Molecular Ecology 17:4334-4345.

Viosca, P. J. 1935. The irises of southeastern Louisiana - a taxonomic and ecological interpretation. Bulletin of the American Iris Society 57:3-56.

Wesselingh, R. A. and M. L. Arnold. 2000. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. Journal of Evolutionary Biology 13:171-180.

Wolf, J. B., J. Lindell, and N. Backström. 2010. Speciation genetics: current status and evolving approaches. Philosophical Transactions of the Royal Society B: Biological Sciences 365:1717-1733.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

CHAPTER 2

DETERMINING POPULATION STRUCTURE AND HYBRIDIZATION FOR TWO IRIS SPECIES[1]

[1] Hamlin, J.A.P and M. L. Arnold. 2014. *Ecology and Evolution* 4:743-755
Reprinted here with the permission of the publisher.

Abstract

Identifying processes that promote or limit gene flow can help define the ecological and evolutionary history of a species. Furthermore, defining those factors that make up 'species boundaries' can provide a definition of the independent evolutionary trajectories of related taxa. For many species, the historical processes that account for their distribution of genetic variation remain unresolved. In this study, we examine the geographic distribution of genetic diversity for two species of Louisiana Irises, *Iris brevicaulis* and *Iris fulva*. Specifically, we asked how populations are structured and if population structure coincides with potential barriers to gene flow. We also asked if there is evidence of hybridization between these two species outside Louisiana hybrid zones. We used a genotyping-by-sequencing approach and sampled a large number of single nucleotide polymorphisms across these species' genomes. Two different population assignment methods were used to resolve population structure in *I. brevicaulis*; however, there was considerably less population structure in *I. fulva*. We used a species tree approach to infer phylogenies both within and between populations and species. For *I. brevicaulis,* the geography of the collection locality was reflected in the phylogeny. The *I. fulva* phylogeny defined much less structure than detected for *I. brevicaulis*. Lastly, combining both species into a phylogenetic analysis resolved two out of six populations of *I. brevicaulis* that shared alleles with *I. fulva.* Taken together, our results suggest major differences in the level and pattern of connectivity among populations of these two Louisiana Iris species.

Introduction

Evolutionary model systems abound, each holding their own promises and pitfalls dependent on the research regime, but many of these systems lack information regarding patterns of genetic diversity within natural populations beyond the reporting of estimates based on a handful of molecular markers. With current advances in genomic approaches, examining such geographic patterns of genetic variation has become possible (e.g. see June 2013 *Molecular Ecology* Special Issue: Genotyping By Sequencing in Ecological and Conservation Genomics). Understanding these patterns is important because of the influence past geological events and genetic drift play on the demography and genetics of populations and species (Hewitt 2000). For example, some species may be continuously distributed throughout their range, however; many species are not found uniformly throughout their range and thus include small, isolated populations (Ellstrand and Elan 1993). Furthermore, evolutionary processes, such as natural selection and gene flow, can be reflected in the geographic sorting of genetic variation in populations (Slatkin 1987).

Gene flow can play an influential role in the evolutionary trajectories of populations through its ability to increase genetic diversity and change allele frequencies. In this investigation we are defining populations as a group of individuals of the same species living in close enough proximity that any member of the group can potentially mate with any other member (Waples and Gaggiotti 2006). Our study also allows us to identify two different kinds of genetic diversity: 1) genetic diversity harbored within species due to population structure and 2) that which is shared among species due to hybridization.

Population structure reflects shared alleles between individuals. Factors such as physical barriers, historical processes, or even variation in life histories can shape population differentiation thus resulting in variation in genetic connectivity among populations (Balloux and Lugon-Moulin 2002; Lowe and Allendorf 2010). Moreover, when gene flow occurs among populations, allele frequencies can become homogenized (Slatkin 1985). Reduced levels of gene flow and ecological differences associated with particular habitat patches can lead to local adaptation and may promote speciation (Barton and Hewitt 1985). In contrast, gene flow may generate new polymorphisms within a population and increase local effective population size, thereby opposing genetic drift (Wright 1931; Slatkin 1985).

Hybridization, as a result of gene flow between divergent but closely related taxa, can occur when species are found in sympatry and reproductive isolation is incomplete (Arnold 2006). One consequence of hybridization is the production of novel combinations of parental genotypes in otherwise isolated genomes (Arnold 1992, 2006). This will result in some unfit hybrid offspring, but it may also promote the exchange of genetic material between species (introgression), particularly if these novel gene combinations provide a selective advantage (Arnold and Hodges 1995; Arnold and Martin 2010). There are many well-known systems for studying hybridization and introgression such as *Mus musculus* and the genus *Heliconius.*

Recent work using *Mus musculus* concluded that directional introgression had occurred from *M. m. musculus* into *M. m. domesticus* supporting the hypothesis that alleles from one lineage may be adaptive in a sister lineage as well (Staubach et al. 2012). Another potential outcome of novel adaptations arising from introgressive hybridization is the formation of hybrid species. In this regard, work in the genus *Heliconius* has shown that *Heliconius heurippa* is of hybrid origin. Specifically, adaptive trait introgression produced a novel wing pattern in the homoploid hybrid species and resulted in reproductive isolation between the hybrid species and its parental species, *H. melpomene* and *H. cydno* (Salazar et al. 2010).Given the extensive work within these species, information is still lacking on the population genetics across these species' ranges.

The Louisiana iris species complex has been developed into a model system for the study of evolutionary genetics. Interest in this clade began with the postulation of a large number of species within Louisiana based on morphology (Small and Alexander 1931). Subsequently, Viosca (1935) and Riley (1938) demonstrated that many of these 'species' were actually hybrids between three taxa, *Iris brevicaulis*, *I. fulva,* and *I. hexagona.* Over the subsequent 70+ years, the Louisiana irises have been used to address a variety of evolutionary hypotheses, such as the link between hybrid fitness, introgression and adaptation, and, in the case of *I. nelsonii*, homoploid hybrid speciation (Randolph 1966; Arnold 1993; Taylor et al. 2011; Taylor et al. 2013). However, much of this work has been focused on populations from the state of Louisiana where all of these species occur sympatrically. Though there is some information concerning population genetic variation in *I. hexagona* (Meerow et al. 2011), the geographic distribution of genetic diversity for two of the species, *I. brevicaulis* and *I. fulva*, has not been studied.

Iris brevicaulis and *I. fulva* occur throughout the Southeastern United States with overlapping ranges along the Mississippi River and *I. fulva* distribution nested within *I. brevicaulis* (Figure 2.1). *Iris brevicaulis* is distributed as far north as Ohio and as far west as Texas. *Iris fulva* is geographically more restricted, being associated with the alluvial valley of the Mississippi River. The Lower Mississippi Alluvial Valley (LMAV), which stretches from Illinois to the Gulf of Mexico, is the historic floodplain and valley of the lower Mississippi (Stanturf et al. 2000). In addition, the entire region of Eastern North America is both geologically and topographically complex, and a number of common phylogeographic patterns have been reported across a wide range of co-distributed taxa (Soltis et al. 2006). From the present study, we address the following questions regarding *I. brevicaulis* and *I. fulva*: 1) How is genetic diversity partitioned and how are populations structured? 2) Does population structure coincide with potential geographic barriers to gene flow, such as the Mississippi River? 3) Is there evidence of hybridization between these species beyond the hybrid zones defined in the state of Louisiana?

Materials and Methods

Study system and sampling

Both species occur in or near wetland systems, but each species differs with regard to microhabitat, floral morphology, and pollination system (Viosca 1935) *Iris fulva* is found at lower elevations in intermittently flooded, forested wetlands. *Iris brevicaulis* occurs at slightly higher elevation in mixed hardwoods (Viosca 1935; Cruzan and Arnold 1993; Johnston et al. 2001). *Iris fulva* flowers are a deep crimson color, have protruding anthers, lack nectar guides, and produce large volumes of nectar (Viosca 1935; Wesselingh and Arnold 2000). Hummingbirds and butterflies pollinate these flowers (Viosca 1935; Wesselingh and Arnold 2000; Martin et al. 2008). Floral characteristics of *I. brevicaulis* are typical of a bumblebee pollination syndrome; floral color varies from light blue to deep blue with marked nectar guides, stiff upright sepals, and strong scent. However, even given these floral differences, *I. brevicaulis* and *I. fulva* are known to form hybrid zones within the state of Louisiana (Viosca 1935; Foster 1937; Riley 1938). Additionally, individuals found in hybrid zones are sometimes characterized by distinct phenotypes and genotypes that are associated with particular ecological traits (Cruzan and Arnold 1994).

During the summer of 2011 and 2012, we performed collections for both species. These collections included populations throughout the species ranges (Figure 2.1). For *I. brevicaulis*, collections occurred from eight states resulting in a total of 15 localities. For *I. fulva*, collections occurred in seven states resulting in a total of ten collection localities. A subset of the populations collected for both species was used in this study (Table 2.1). Sequenced populations included samples from both sides of the Mississippi River. We did not sequence individuals found in previously documented hybrid zones.

DNA extraction, library construction and sequencing

DNA was extracted from eight individuals for six populations of each species from populations throughout the species' range (Figure 2.1). Extractions were performed using the Qiagen DNeasy plant kit (Qiagen, Valencia, CA). Extracted DNA was sent to the Cornell Institute for Genomic Diversity for genotyping-by-sequencing (GBS) (Elshire et al. 2011) is similar to RAD sequencing in that it generates reduced representation libraries by digesting DNA with a restriction enzyme; however, GBS differs from RAD sequencing in that single-well digestion of genomic DNA is included, adaptor ligation results in reduced sample handling and DNA fragments are not size selected (Elshire et al. 2011). Libraries were prepared from 48 *I. brevicaulis* individuals, 47 *I. fulva* individuals, and one blank (i.e. control) for sequencing. DNA from each individual was separately digested using *Eco*T221, a six base cutter, and the fragmented DNA was then ligated to a barcoded adaptor and a common adaptor. Within a 96-well plate each well contained DNA from a different individual and a barcode adaptor unique to that well. Individuals were barcoded to allow discovery of genetic variation both within and between species. The resulting libraries were sent for sequencing using single-end 100-bp reads on the Illumina HiSeq 2000 with 48 samples sequenced per lane.

SNP discovery, genotyping, and summary statistics

The non-reference pipeline, Universal Network-Enables Analysis Kit (UNEAK; http://www.maizegenetics.net/gbs-bioinformatics), was used for SNP discovery and genotyping both between and within species. UNEAK takes raw Illumina sequence files and converts them into individual genotypes. Reads are retained, and trimmed to 64 bp, when they possess a barcode, cut site, and no 'N's in the first 64 bp of sequence after the barcode. Identical reads are clustered into tags and counts of these tags present in each barcoded individual are stored. Pairwise alignment, then, identifies tag pairs having a single base pair mismatch and these single base pair mismatches are considered candidate SNPs. Any tag pair that contains more than one mismatch is discarded to minimize SNPs resulting from alignment of paralogous sequences. Error tolerance rate was set at 0.03, so that only reciprocal pairs of tags are retained for SNP calling according to standard protocols of Cornell Institute for Genomic Diversity.

Subsequent filtering of tags was done, using the program TASSEL 4.0 in order to identify SNPs, both within and between species (Bradbury et al. 2007). A few individuals were not used in identifying SNPs, because those samples failed during sequencing. After removing failed samples and setting a threshold of 500k reads, 20% missing data ('N's), and a minor allele frequency of >1% (or the minimum frequency at which a common allele must occur), a total of 67 individuals were used to generated a filtered set of SNPs between species. The total number of individuals used to call SNPs within a species was 43, which we excluded the failed samples and set a threshold of 20% missing data ('N's), and a minor allele frequency of >1%. Pairwise F_{st} values, observed and expected heterozygosity, inbreeding coefficient (G_{is}), which is analogous to F_{is} by Wright (1951), and isolation by distance (IBD) were estimated for each SNP and their average across loci was computed using Genodive (Meirmans and Van Tienderen 2004).

Population assignments

Individual and population assignments were conducted using the program STRUCTURE version 2.3.4, both within and between species (Pritchard et al. 2000). For all analyses, we used a model of admixture to determine the number of population clusters (K) with a burn-in of 1,000,000 and 10,000,000 iterations. Analyses were repeated 10 times for each k value, ranging from 1 to 7 within species or 1 to 13 between species. The average and standard deviation (SD) of the natural log probability of each model was used to calculate ΔK (Evanno et al. 2005) using Structure Harvester (Earl and vonHoldt 2012); clumpp, set at the default settings, was used to assess the similarity between replicate STRUCTURE results (Jakobsson and Rosenberg 2007).

Discriminant Analysis of Principle Components (DAPC) (Jombart et al. 2010)from the package adegenet (Jombart 2008) version 1.2.8 in R (2009) was also used to detect the number of genetic clusters. This technique uses a non-model based multivariate approach. DAPC first transforms genetic data into uncorrelated components using Principle Component Analysis (PCA), and then performs a Discriminant Analysis on the retained principle components (PCs). We used the find.clusters function of DAPC to infer the most likely number of clusters. To calculate the probability of assignment of individuals to each of theses clusters using DAPC, we determined the optimal number of principle components, here the optimal number of PCs retained was N/3; N = number of samples (14), as advised in the manual. We calculated the Bayesian Information Criterion (BIC) for K = 1 – 7, where K = number of populations. Optimal number of populations was identified as the one in which the BIC was the lowest value and after which the BIC either increased or decreased by the least amount.

Phylogenetic analyses

Maximum likelihood phylogenies were inferred using RAxML(Stamatakis 2006). For each individual, SNPs were concatenated into one sequence. RAxML was run using a GTR+G model of nucleotide substitution. The -D option, which stops the ML searches when they have reached the asymptotic convergence phase, was used. The criterion for stopping the searches is based on computing the Robinson-Foulds (RF) distance (Robinson and Foulds 1981) between two consecutive intermediate trees. If the RF distance between two consecutive trees is smaller than 1%, the ML search is stopped. Support for nodes in the RAxML inferred tree was assessed using a bootstrap analysis with 100 replicates. However, treating SNPs as a contiguous sequence violates assumptions about recombination and because of this we also inferred a species tree.

Species trees from SNPs were inferred using SNAPP (SNP and AFLP Package for Phylogenetic analysis) with the default settings. SNAPP, part of BEAST 2.0, is a package for inferring species trees and species demographics from independent (unlinked) biallelic markers (Bryant et al. 2012). This package implements a full coalescent model to integrate over all possible gene trees rather than sampling them explicitly. SNAPP was run for at least 10,000,000 generations with sampling every 1000 generations. Convergence of parameters onto posterior distribution was assessed using Tracer version 1.5 (http://tree.bio.ed.ac.uk/software/tracer/). All parameters had effective sample size (ESS) values > 250. Convergence onto posterior topology was assessed using AWTY (Nylander et al. 2008). All samples before convergence were removed as burn-in, and maximum clade credibility trees were generated using TreeAnnotator within the BEAST 2.0 package. Finally, due to lack of an outgroup all trees are midpoint rooted.

Results

Genotyping and population statistics

A total of 151,189 SNPs were identified from individuals from all populations of *I. brevicaulis*. After filtering out failed samples, missing data ≤ 20% and a minimum allele frequency of >1%, 387 SNPs were retained for 43 individuals. For *I. fulva*, 97,432 SNPs were called from individuals from all populations. Applying the same filtering as above, a total of 560 SNPs were retained for analyses from 43 individuals. There were 196,133 SNPs in total generated when *I. brevicaulis* and *I. fulva* were run through the UNEAK pipeline. For 67 individuals combined, and the same filtering protocol except with a minimum of 500k reads, a total of 1140 SNPs were retained. To determine whether the tag pairs were derived from nuclear or chloroplast DNA, a subset of the tag pairs (n = 200) were blasted against the non-redundant, nucleotide database collection using BLASTN. Almost all tag pairs examined resulted in E-value scores which were not lower than 10^{-4}; however, four tag pairs had an E-value of below 10^{-4} that matched sequences from chloroplast DNA. One tag pair with a significant E-value was associated with an *I. brevicaulis* IRRE transposon marker.

For *I. brevicaulis,* within-population observed heterozygosity (H_o) ranged from 0.101 to 0.259 and expected frequency of heterozygotes over all populations (H_e) ranged from 0.144 to 0.217. The inbreeding coefficient (G_{is}) ranged from 0 to 0.3, with three population G_{is} values significantly different than zero (IB_AL, IB_LA, and IB_OH) (Appendix A, Table S1a). For *I. fulva*, within-population observed heterozygosity (H_o) ranged from 0.314 to 0.561 and expected frequency of heterozygotes over all populations (H_e) ranged from 0.357 to 0.403. The inbreeding coefficient (G_{is}), for *I. fulva*, was approximately zero for all populations (Appendix A, Table S1b). Genetic differentiation (pairwise population F_{st}) for *I. brevicaulis* populations was moderately high (2.2a). In contrast for *I. fulva,* the degree to which populations were different from one another was much lower (Table 2.2b).

No isolation-by-distance was detected (data not shown) in either species. The slope of the regression line in the Mantel test (correlation of geographical and genetic distance) was not significantly different than zero (*I. brevicaulis*: $R^2 = 0.051$, $P = 0.192$; and *I. fulva*; $R^2 = 0.118$, $P = 0.130$).

Population assignments

We analyzed models (using STRUCTURE) for within a species with 1 – 7 clusters and between species with 1 – 13 clusters. ΔK, the statistic used to detect the most likely number of clusters, K, peaks at K = 2 for *I. brevicaulis* (Figure 2.2a). *Iris brevicaulis* individuals from Alabama and Louisiana were placed in one cluster along with Texas, to a lesser extent. While all other individuals formed a second cluster. For *I. fulva,* ΔK peaks at both K = 2 (Figure 2.2b) and K = 4 (Appendix A, Figure S1), although with only a slight difference between the two. At K = 2, we do not see two distinct clusters, but a proportion of each individual's genome is shared with both clusters. Because the proportion of the individuals assigned to each of the two populations is roughly symmetric, we conclude that there is basically no population structure. When individuals from both species were combined and analyzed using STRUCTURE, the optimal number of clusters is K = 2 (Figure 2.2c). Individuals within the two distinct clusters were associated with previously assumed species but with some *I. brevicaulis* individuals from Alabama, Louisiana, and Texas, though to a lesser extent, clustering with *I. fulva.*

For *I. brevicaulis* DAPC analysis, the BIC gave the most support to three genetic clusters (Appendix A, Figure S2a), which differs from STRUCTURE (Figure 2.2a). The first principle component separated populations 2 and 5 (Arkansas and Ohio) from 1, 3, 4, and 6. The second component separated these last four populations into two groups, southern populations that are closest to the hybrid zone (clusters 1, 4, and 6; Alabama, Louisiana, and Texas, respectively) and a northern population (cluster 3; Illinois) (Figure 2.3a). In the *I. fulva* DAPC analysis, the BIC gave the most support for one genetic cluster (Appendix A, Figure S2b). However, there is subtle genetic structure where the first principle component separated populations 1, 4, and 6 from populations 2, 3, and 5. The second component separated the first group into populations 1 and 4 (Arkansas and Louisiana) from population 6 (Mississippi). For the second group (2, 3, and 5), the second function separated 2 and 3 (Illinois and Kentucky) from 5 (Missouri) (Figure 2.3b).

Phylogenetic analysis

Nodes with less than 50% support were collapsed for RAxML trees. Most nodes in the *I. fulva* tree were collapsed resulting in a polytomy. The *I. brevicaulis* phylogeny generated by RAxML showed some structuring with Alabama and Louisiana genotypes forming one clade (bootstrap support = 98 %), which was sister to all other samples (Appendix A, Figure S3). The remainder of the genotypes demonstrated a geographic partitioning with populations becoming more nested with increasing latitude. When species were combined, individuals from Alabama and Louisiana formed one clade (bootstrap support = 98 %) that was more closely related to *I. fulva* genotypes than to other *I. brevicaulis* individuals or populations (Appendix A, Figure S4).

SNAPP species trees were constructed by defining populations as different species. We detected high posterior support for each of the nodes ranging from 0.982 – 1 for *I. brevicaulis* (Figure 2.4a). Population level clades were similar to the phylogeny estimated using RAxML (Appendix A, Figure S4). There was also high support for Alabama and Louisiana being sister to each other. The *I. fulva* species tree reflected population differentiation with high posterior support for 1) the Louisiana sample being sister to all other populations and 2) the Arkansas sample being sister to the four other populations (Figure 2.4b). For the remainder of the populations (Illinois, Kentucky, Missouri, Mississippi) the posterior support was much lower (i.e. 0.513). When samples of both species were combined for the species tree analysis, the *I. brevicaulis* Alabama and Louisiana populations were again more closely related to *I. fulva*, than to other *I. brevicaulis* samples (Figure 2.4c).

Discussion

Understanding how individuals, gametes, and genes move between populations and species is fundamental to studies of ecology and evolution. For example, habitat fragmentation can lead to reduced population size and gene flow with resulting decreases in individual heterozygosity and local genetic diversity over time (Ellstrand and Elan 1993). This can result in inbreeding depression and/or differentiation at a regional or range-wide scale (Ellstrand and Elan 1993). Specific biotic regions, such as the Southeastern United States, provide an ideal backdrop to test hypotheses regarding the factors that may have affected genetic partitioning within species (Soltis et al. 2006).

The present study used a genotyping-by-sequencing (GBS) approach to genotype individuals. Advantages of using a GBS approach include genotyping non-model, large genome, organisms using a large panel of SNPs at a cost that is comparable to traditional Sanger sequencing for a handful of molecular markers. Library preparation is much simpler than traditional RAD approaches and can be performed relatively quickly and is amenable to a high level of multiplexing. Essentially, GBS does marker discovery and genotyping simultaneously compared to independent marker discovery, assay designs, and genotyping following the classical approach. Lastly, by generating fragments of DNA spread throughout the genome, one can potentially sequence regions inaccessible to other sequencing approaches. However, GBS suffers the same bioinformatics hurdles as all next-generation sequence methods such as: generating and analyzing massive amounts of data, dealing with complex genomes, and the presence of missing data.

Population structure and the Mississippi River

While the Mississippi River has been documented to be a barrier to gene flow in a number of species (Soltis et al. 2006), this was not found to be the case for *I. brevicaulis* and *I. fulva.* Our results are in line with those previously reported for two species of frog, *Rana catesbiana* and *Pseudacris crucifer* (Austin et al. 2004). These two species showed geographic overlap of genetic diversity suggesting that this waterway has not been a major barrier to gene flow (Austin et al. 2004). In contrast, one species, the Loblolly pine (*Pinus taeda*) does show differentiation between populations located on either side of the Mississippi River, consistent with the river being a barrier to gene flow (Al-Rabab'ah and Williams 2002; Eckert et al. 2010). It was hypothesized that differentiation among the pine populations may have resulted from separate Pleistocene refugia on either side of the river (Al-Rabab'ah and Williams 2002).

In contrast to the Loblolly pine, *I. brevicaulis* does not appear to have been influenced by the Mississippi River acting as a barrier to gene flow. Thus, we did not find populations on either side of this waterway to be more similar to one another than those located on the opposite side. Instead, it appears that populations closest to the Louisiana hybrid zones form one separate genetic unit relative to other samples. This was evident from the Bayesian clustering method, STRUCTURE, as well as the DAPC analysis. *Iris brevicaulis* also demonstrated high levels of genetic differentiation or pairwise F_{st} values. Alabama, Louisiana, and Texas populations have the highest levels of differentiation calculated by F_{st} and lowest levels of genetic diversity (expected heterozygosity) compared to the remaining populations. Additionally, both the RAxML and SNAPP trees failed to detect genetic differentiation due to the Mississippi River. Specifically, we did not resolve two reciprocally monophyletic groups of populations distributed on opposite sides of the Mississippi River. Instead, the phylogeographic pattern for *I. brevicaulis* is one of higher latitude populations being more derived relative to more southerly-distributed samples.

Based on the geographic distribution of *I. brevicaulis*, it is possible to test whether this species conforms to the central – marginal hypothesis (Eckert et al. 2008). The central – marginal hypothesis states that peripheral populations exhibit low genetic diversity and greater genetic differentiation compared to central populations. Additionally, effective population size and the rate of migration should be highest at the range center than at range margins; this results from smaller effective population size and greater geographic isolation of the range margins. Some *I. brevicaulis* populations (i.e. Alabama or Arkansas) would be expected to exhibit characteristics typical of central populations, such as having high levels of genetic diversity as measured by expected heterozygosity (Eckert et al. 2008). The *I. brevicaulis* Arkansas population does indeed reflect characteristics of central populations. In contrast, we detected the opposite pattern for the *I. brevicaulis* sample from Alabama, possibly resulting from high levels of inbreeding within this latter population.

In contrast to *I. brevicaulis, I. fulva* showed no geographic clustering, as reflected by the estimated levels of shared genetic variation among all populations (Figure 2.2a). Furthermore, there were low levels of genetic differentiation among populations as indicated by pairwise F_{st} values. We were unable to resolve the *I. fulva* population phylogeny using a maximum likelihood method, which required us to concatenate all SNPs (n = 560) into one single locus. We were, however, able to resolve the population phylogeny using a species tree approach. We found that the Louisiana population was a sister lineage to all other populations. Future analyses estimating migration rates between collection localities should provide further definition of patterns of gene flow among *I. fulva* populations.

Despite the fact that the two iris species studied here are closely related and have similar life histories, the patterns of genetic diversity suggest two different evolutionary histories. First, it appears that *I. brevicaulis* is experiencing limited gene flow between populations, while *I. fulva* is not. Our results are not consistent with a geographic barrier (i.e. the Mississippi River) to gene flow; there are other processes that could lead to the observed patterns of genetic differentiation. Within plants, selection on various floral morphologies has been well documented and shown to be influenced by both abiotic and biotic factors (Fenster et al. 2004; Strauss and Whittall 2006; Bomblies 2010). In animal pollinated species, patterns of pollen dispersal and receipt depend on the behavior and effectiveness of pollinators (Waser and Price 1983; Galen 1989; Campbell et al. 1996). Floral traits, such as color, fragrance, shape, and size, reflect various 'pollinator syndromes'. *Iris brevicaulis* is primarily bumblebee pollinated (Viosca 1935; Bouck et al.

2007) and *Iris fulva* is predominately pollinated by either hummingbirds or butterflies (Viosca 1935; Bouck et al. 2007).

Fragmentation of habitats can affect bee populations and thereby disrupt plant-pollinator interactions (Steffan-Dewenter and Tscharntke 1999). For example, Steffan-Dewenter and Tscharntke (1999) found that increasing isolation of small habitat islands resulted in both decreased abundance and species richness of flower-visiting bees. From this, they concluded that habitat connectivity is essential to maintain not only abundant and diverse bee communities, but also plant-pollinator interactions in wild plants. In contrast, ruby-throated hummingbirds (*Archilochus colubris*) are well-known for spring migratory flights that cover great distances encompassing most of eastern North America (dePamphilis and Wyatt 1989). This migratory pattern led to the conclusion that ruby-throated hummingbirds contributed to long-distance gene flow in buckeyes (dePamphilis and Wyatt 1989). The patterns of limited gene flow between *I. brevicaulis* populations versus higher rates of gene flow among geographically widely-spaced *I. fulva* populations is thus consistent with the behavior of their major pollinators.

Additional to the possible role of pollinator systems in limiting and promoting gene flow within the two Louisiana iris species, their divergent ecological associations may also be affecting the other avenue of gene flow available for plant species, that of seed dispersal. Because iris seeds float they can be dispersed via waterways. As *I. fulva* is closely associated with the LMAV, which is prone to flooding, seed dispersal may have led to the higher estimated levels of gene flow for this species. Additionally, whole parts of plants could be washed from one place to another via flooding, potentially spreading the plants clonally across the flood plane. For *I. brevicaulis*, it is more likely that there is a lack of connectivity from seed dispersal due to its greater separation from waterways (Viosca 1935; Cruzan and Arnold 1993).

Evidence of hybridization beyond Louisiana hybrid zones

Outcomes from natural hybridization are varied and often have significant evolutionary consequences (Arnold 2006). In the present study, we first tested for introgression between the two iris species using STRUCTURE. This analysis resolved two distinct clusters consisting of either *I. brevicaulis* or *I. fulva* individuals. However, a proportion of the genome of Alabama, Louisiana, and Texas *I. brevicaulis* individuals was associated with the *I. fulva* cluster. A similar finding was apparent from the phylogeny containing both species, in which *I. brevicaulis* individuals from Alabama and Louisiana collection localities formed a monophyletic clade that was more closely related to *I. fulva* than to other *I. brevicaulis* populations. Finally, the species tree approach corroborated our maximum likelihood phylogeny with Alabama and Louisiana *I. brevicaulis* populations forming a sister clade to *I. fulva*, however node support for was fairly low (0.63).

Previous analyses of the Louisiana irises found that, within a hybrid population, pollinator-mediated pollen movement between two species was more likely than seed movement. Specifically, some individuals were exclusively the products of pollen transfer from *I. fulva,* or hybrid plants, into a third species of the Louisiana iris species complex, *I. hexagona* (Arnold et al. 1991). However, it was inferred that both seed and pollen-mediated gene flow occurred and/or that this was a region of historical overlap. Regarding the inferred introgression into the Alabama, Louisiana, and Texas *I.*

brevicaulis populations, *I. fulva* has been previously recorded from these states. Though we failed to locate *I. fulva* in either Alabama or Texas, it is likely that past sympatry between these two species generated the patterns of admixture found in the *I. brevicaulis* populations.

Past analyses within the Louisiana irises have consistently detected asymmetrical introgressive hybridization. In regard to *I. fulva* and *I. brevicaulis*, this asymmetry was reflected in a higher frequency of transfer of alleles from the former into the latter species (Cruzan and Arnold 1994; Martin et al. 2005, 2006; Tang et al. 2010). Here, we demonstrate a pattern that is consistent with the previous findings. Thus, the population genetic variation detected here suggests a further case of asymmetric introgression between these two species – from *I. fulva* into the Alabama, Louisiana, and Texas populations of *I. brevicaulis*.

It has been known for decades that introgression is not homogeneous across the genome (e.g. Key 1968); the frequency of exchange of specific genomic regions between species depends on both demographic and deterministic processes (see Arnold and Martin 2010, for a review). In a number of documented cases of introgression, it has been shown that some loci move freely between genomes while others remain highly differentiated (Nosil et al. 2009). Loci that do not introgress are thought to be associated with genomic regions that contribute to reproductive isolation. In a previous study of *I. fulva* and *I. brevicaulis,* Tang et al. (2010) found that alleles from *I. fulva* were significantly favored compared to regions from *I. brevicaulis*, but concluded that both species genomes were permeable to introgression. Additionally, using both greenhouse studies and natural field experiments, Martin et al. (2005, 2006) documented that adaptive trait introgression had occurred between *I. brevicaulis* and *I. fulva*. These authors detected an increased survival of genotypes of a backcross-1 generation toward *I. brevicaulis*, when these genotypes contained significantly more alleles from *I. fulva*. These patterns reflected signatures of positive selection for certain introgressed alleles (i.e. adaptive trait introgression), which apparently underlie adaptations to ecological settings within these iris species (Martin et al. 2005, 2006). Further examination of the loci used in this study will elucidate whether those loci are adaptive.

Adaptive trait introgression has been inferred in a number of organisms other than iris. For example in *Helianthus debilis* var *cucumerifolious* and *Helianthus annuus* ssp *annuus* (Kim and Rieseberg 1999; Whitney et al. 2010), such transfer was hypothesized to account for the expansion of *Helianthus annuus* into novel environments and in producing its unique phenotype. These hypotheses were tested using a variety of experimental hybrid and parental genotypes/phenotypes, in both greenhouse and field experiments. A pattern of overrepresentation of *H. debilis* alleles suggested that these genomic components might confer a fitness advantage to the introgressed *H. annuus*. Additionally, introgression apparently altered multiple traits of the *H. annuus* phenotype causing adaptive changes in components interacting with the biotic and abiotic environments (Whitney et al. 2010).

The present study demonstrates species-level differences in connectivity among populations of Louisiana irises. In particular, *I. brevicaulis* populations appear much more isolated from intraspecific gene flow than do *I. fulva* populations. However, neither species demonstrates phylogeographic breaks indicative of the Mississippi River acting as a barrier to gene flow. In contrast to isolation among *I. brevicaulis* populations, we detected asymmetric introgressive hybridization of alleles largely from *I. fulva* into *I. brevicaulis*, with some of the allele transfer possibly being adaptive. For example, the introgressed, Alabama population of *I. brevicaulis* was found in an *I. fulva* like habitat, i.e. in standing water. Martin et al. (2006) documented adaptive trait introgression of 'flood-tolerance' alleles from *I. fulva* into *I. brevicaulis* that allowed the latter to survive an extreme flooding event. Performing a transplant study, in order to evaluate if *I. fulva* parents or hybrid individuals survive as well or better in the *I. brevicaulis* collection site in Alabama, should provide a further test of the hypothesis of adaptive trait introgression within this species complex.

Acknowledgments

JAPH would like to acknowledge the many governmental agencies and individuals who provided valuable information regarding natural populations. A special thank you goes to D. White of the Kentucky State Nature Preserve, B. Pendley of the Mingo National Wildlife Refuge, D. Gardner, and E. Dearing. JAPH would also like to thank E. Ballerini, A. Brothers, and R. Eytan. Funds from the American Iris Society supported the collections and genotyping for this project. JAPH and the research reported in this publication were supported by the National Institute of Health under award number T32GM007103. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Current National Science Foundation funding for this work comes from DEB-0949479/0949424 (collaborative grant with NH Martin of Texas State University) and DEB-1049757.

Data Accessibility

The associated SNP genotypes for both within and between species (Hapmap format) along with GPS coordinates of collection localities are deposited in the Dryad data repository: doi: 10.5061/dryad.sd845.

Literature Cited

Al-Rabab'ah, M. and C. Williams. 2002. Population dynamics of Pinus taeda L. based on nuclear microsatellites. Forest Ecology and Management 163:263-271.

Arnold, M. L. 1992. Natural Hybridization as an Evolutionary Process. Annual Review of Ecology and Systematics 23:237-261.

Arnold, M. L. 1993. Iris nelsonii (Iridaceae): origin and genetic composition of a homoploid hybrid species. American Journal of Botany:577-583.

Arnold, M. L. 2006. Evolution Through Genetic Exchange. Oxford University Press, New York.

Arnold, M. L., C. M. Buckner, and J. J. Robinson. 1991. Pollen-mediated intrgression and hybrid speciation in Louisiana irises. Proceedings of the National Academy of Sciences 88:1398-1402.

Arnold, M. L. and S. A. Hodges. 1995. Are natural hybrids fit or unfit relative to their parents? Trends in Ecology & Evolution 10:67-71.

Arnold, M. L. and N. H. Martin. 2010. Hybrid fitness across time and habitats. Trends in Ecology & Evolution 25:530-536.

Austin, J. D., S. C. Lougheed, and P. T. Boag. 2004. Discordant temporal and geographic patterns in maternal lineages of eastern north American frogs, Rana catesbeiana (Ranidae) and Pseudacris crucifer (Hylidae). Molecular Phylogenetics and Evolution 32:799-816.

Balloux, F. and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. Molecular Ecology 11:155 - 165.

Barton, N. H. and G. Hewitt. 1985. Analysis of hybrid zones. Annual review of Ecology and Systematics:113-148.

Bomblies, K. 2010. Doomed lovers: mechanisms of isolation and incompatibility in plants. Annual review of plant biology 61:109-124.

Bouck, A., S. R. Wessler, and M. L. Arnold. 2007. Qtl analysis of floral traits in Louisiana Iris hybrids. Evolution 61:2308-2319.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular biology and evolution 29:1917-1932.

Campbell, D. R., N. M. Waser, and M. V. Price. 1996. Mechanisms of hummingbird-mediated selection for flower width in *Ipomopsis aggregata*. Ecology 77:1463 - 1472.

Cruzan, M. B. and M. L. Arnold. 1993. Ecological and genetic associations in an Iris hybrid zone. Evolution:1432-1445.

Cruzan, M. B. and M. L. Arnold. 1994. Assortative mating and natural selection in an Iris hybrid zone. Evolution:1946-1958.

dePamphilis, C. W. and R. Wyatt. 1989. Hybridization and introgression in buckeyes (Aesculus: Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. Systematic botany:593-611.

Earl, D. A. and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361.

Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. Gonzalez-Martinez, and D. B. Neale. 2010. Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). Genetics 185:969 - 082.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Molecular Ecology 17:1170-1188.

Ellstrand, N. C. and D. R. Elan. 1993. Population Genetic Consequences of Small Population Size: Implication for Plant Conservation. Annual Review of Ecology and Systematics 24:217 - 242.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-sequeing (GBS) Approach for High Diversity Species. PLos One 6.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14:2611-2620.

Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson. 2004. Pollination syndromes and floral specialization. Annual Review of Ecology Evolution and Systematics 35:375-403.

Foster, R. C. 1937. A cyto-taxonomic survey of the North American species of Iris. Contributions form the Gray Herbarium:3 - 80.

Galen, C. 1989. Measuring pollinator-mediated selection on morphometric floral traits: bumblebees and the alpine sky pilot, Polemonium viscosum. Evolution:882-890.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907 - 913.

Jakobsson, M. and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Johnston, J. A., R. A. Wesselingh, A. C. Bouck, L. A. Donovan, and M. L. Arnold. 2001. Intimately linked or hardly speaking? The relationship between genotype and environmental gradients in a Louisiana Iris hybrid population. Molecular Ecology 10:673-681.

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

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

Key, K. H. L. 1968. The concept of stasipatric speciataion. Systematic Zoology 17:14-22.

Kim, S.-C. and L. H. Rieseberg. 1999. Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. Genetics 153:965 - 977.

Lowe, W. H. and F. W. Allendorf. 2010. What can genetics tell us about population connectivity? Molecular Ecology 19:3038-3051.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2005. Loci affecting long-term hybrid survivorship in Louisiana irises: Implications for reproductive isolation and introgression. Evolution 59:2116-2124.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2006. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics 172:2481-2489.

Martin, N. H., Y. Sapir, and M. L. Arnold. 2008. The genetic architecture of reproductive isolation in Louisiana irises: pollination syndromes and pollinator preferences. Evolution 62:740-752.

Meerow, A. W., M. Gideon, D. N. Kuhn, S. Mopper, and K. Nakamura. 2011. The genetic mosaic of iris series Hexagonae in Florida: inferences on the holocene history of the Louisiana irises and anthropogenic effects on their distribution. International Journal of Plant Sciences 172:1026-1052.

Meirmans, P. G. and P. H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4:792-794.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375-402.

Nylander, J. A. A., J. C. Wilgenbusch, D. L. Warren, and D. L. Swofford. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581-583.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Randolph, L. F. 1966. *Iris nelsonii*, a new species of Louisiana Iris of hybrid origin. Baileya 14.

Rieseberg, L. H. E. 2013. Genotyping By Sequencing in Ecological and Conservation Genomics (Special Issue). Molecular Ecology 22.

Riley, H. P. 1938. A character analysis of colonies of *Iris fulva*, *Iris hexagona* var giganticaerulea and natural hybrids. American Journal of Botany 25:727 - 738.

Robinson, D. F. and L. R. Foulds. 1981. Comparison of Phylogenetic Trees. Mathematical Biosciences 53:131-147.

Salazar, C., S. W. Baxter, C. Pardo-Diaz, G. Wu, A. Surridge, M. Linares, E. Bermingham, and C. D. Jiggins. 2010. Genetic Evidence for Hybrid Trait Speciation in Heliconius Butterflies. Plos Genetics 6.

Slatkin, M. 1985. Gene Flow in Natural Populations. Annual Review of Ecology and Systematics 16:393 - 430.

Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural Populations. Science 236:787-792.

Small, J. K. and E. J. Alexander. 1931. Botanical interpretation of the iridaceous plants of the Gulf States. Contributions of the New York Botanical Garden 327.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15:4261-4293.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stanturf, J. A., E. S. Gardiner, P. B. Hamel, M. S. Devall, T. D. Leininger, and M. E. Warren. 2000. Restoring Bottomland Hardwood Ecosystems in the Lower Mississippi Alluvial Valley. Journal of Forestry 98:10 - 16.

Staubach, F., A. Lorenc, P. W. Messar, K. Tang, D. A. Petrov, and D. Tautz. 2012. Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculusI)*. Plos Genetics 8.

Steffan-Dewenter, I. and T. Tscharntke. 1999. Effects of habitat isolation on pollinator communities and seed set. Oecologia 121:432 - 440.

Strauss, S. Y. and J. B. Whittall. 2006. Non-pollinator agents of selection on floral traits *in* L. D. Harder, and S. C. H. Barrett, eds. Ecology and evolution of flowers. Oxford Univerisity Press.

Tang, S., R. A. Okashah, S. J. Knapp, M. L. Arnold, and N. H. Martin. 2010. Transmission ratio distortion results in asymmetric introgression in Louisiana Iris. BMC plant biology 10:48.

Taylor, S., L. Rojas, S. Ho, and N. Martin. 2013. Genomic collinearity and the genetic architecture of floral differences between the homoploid hybrid species Iris nelsonii and one of its progenitors, Iris hexagona. Heredity 110:63-70.

Taylor, S. J., R. W. Willard, J. P. Shaw, M. C. Dobson, and N. H. Martin. 2011. Differential response of the homoploid hybrid species Iris nelsonii (Iridaceae) and its progenitors to abiotic habitat conditions. American journal of botany 98:1309-1316.

Viosca, P. J. 1935. The irses of southeastern Louisiana - a taxonomic and ecological interpretation. Bulletin of the American Iris Society 57:3 - 56.

Waples, R. S. and O. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419-1439.

Waser, N. M. and M. V. Price. 1983. Pollinator behaviour and natrual selection for flower colour in *Delphinium nelsonii*. Nature 302:422 - 424.

Wesselingh, R. A. and M. L. Arnold. 2000. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. Journal of Evolutionary Biology 13:171-180.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. New Phytologist 187:230-239.

Wright, S. 1931. Evolution in Mendelian Populations. Genetics 16.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Table 2.1. Collection information for a) *I. brevicaulis* populations. b) *I. fulva* populations.

a	State	Id	Longitude	Latitude	*n*
	Alabama	AL	-86.239	31.920	8
	Arkansas	AR	-91.499	34.854	8
	Illinois	IL	-89.113	38.926	8
	Louisiana	LA	-92.051	30.519	8
	Ohio	OH	-82.551	41.361	8
	Texas	TX	-96.201	30.568	8

b	State	Id	Longitude	Latitude	*n*
	Arkansas	AR	-91.952	34.235	8
	Illinois	IL	-89.396	37.443	8
	Kentucky	KY	-89.274	36.539	8
	Louisiana	LA	-90.819	29.877	7
	Mississippi	MS	-90.910	33.791	8
	Missouri	MO	-90.196	36.973	8

n = per population individuals genotyped

Table 2.2. Pairwise Fst values for a) *I. brevicaulis* populations. b) *I. fulva* populations. All F_{st} values were significantly different between all populations.

a	IB_AL	IB_AR	IB_IL	IB_LA	IB_OH
IB_AL	-				
IB_AR	0.53	-			
IB_IL	0.56	0.19	-		
IB_LA	0.2	0.51	0.54	-	
IB_OH	0.57	0.25	0.15	0.55	-
IB_TX	0.56	0.27	0.35	0.52	0.38

b	IF_AR	IF_IL	IF_KY	IF_LA	IF_MS
IF_AR	-				
IF_IL	0.04	-			
IF_KY	0.06	0.01	-		
IF_LA	0.04	0.05	0.06	-	
IF_MS	0.05	0.03	0.06	0.04	-
IF_MO	0.03	0.02	0.04	0.03	0.02

Here abbreviations are based on state where populations are collected. AL = Alabama, AR = Arkansas, IL = Illinois, KY = Kentucky, LA = Louisiana, MS = Mississippi, MO = Missouri, OH = Ohio, TX = Texas

Figure 2.1.

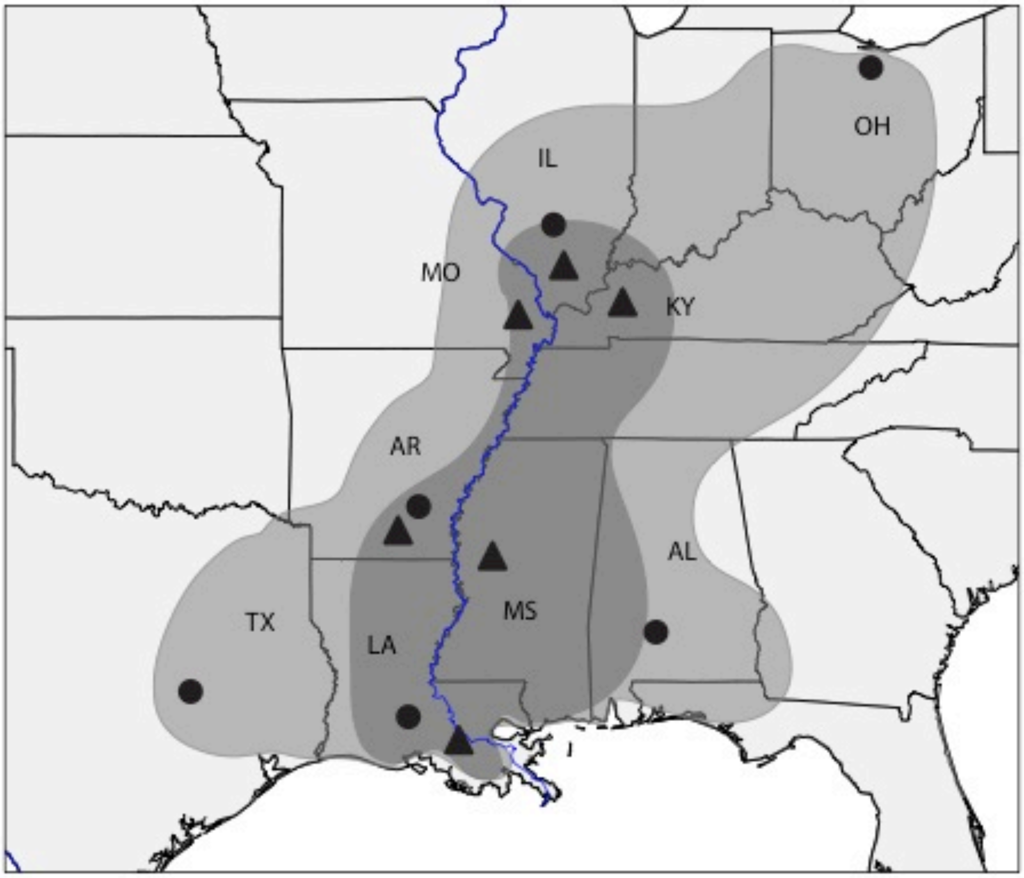


Figure 2.1. Range distribution maps for both *Iris brevicaulis* (light gray) and *Iris fulva* (dark gray). *Iris fulva* distribution is overlaid on *Iris brevicaulis* distribution. Sympatric populations have only been found in Louisiana. Black dots are *Iris brevicaulis* and black triangles are *Iris fulva* collection localities, respectively, from which individuals were sequenced. Abbreviations of states are shown, which corresponds to population names. The Mississippi River is outlined in blue.

Figure 2.2.

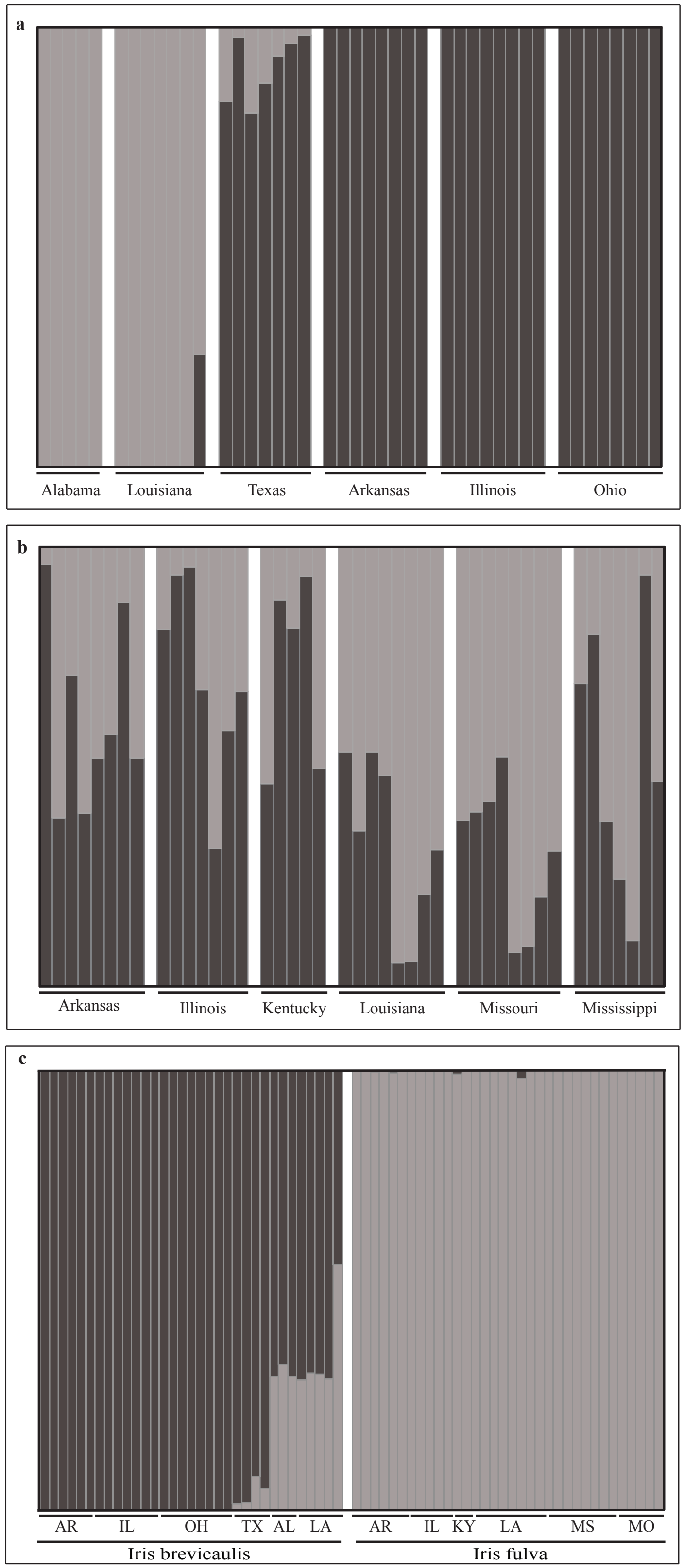

a
Alabama
Louisiana
Texas
Arkansas
Illinois
Ohio
b
Arkansas
Illinois
Kentucky
Louisiana
Missouri
Mississippi
c
AR
IL
OH
TX
AL
LA
AR
IL
KY
LA
MS
MO
Iris brevicaulis
Iris fulva

Figure 2.2. Plots of posterior probabilities of group assignments of each individual into two clusters based on the STRUCTURE analysis. The results are grouped by collection localities for each individual. Each vertical bar represents a different individual from one of six populations. The color proportion for each bar represents the posterior probability of assignment of each individual to one of two clusters. a) *Iris brevicaulis* plot of posterior probabilities of group assignments where K = 2. b) *Iris fulva* plot of posterior probabilities of group assignment where K = 2. c) *Iris brevicaulis* and *I. fulva* species plot of posterior probabilities of group assignments generated where K = 2.

Figure 2.3.

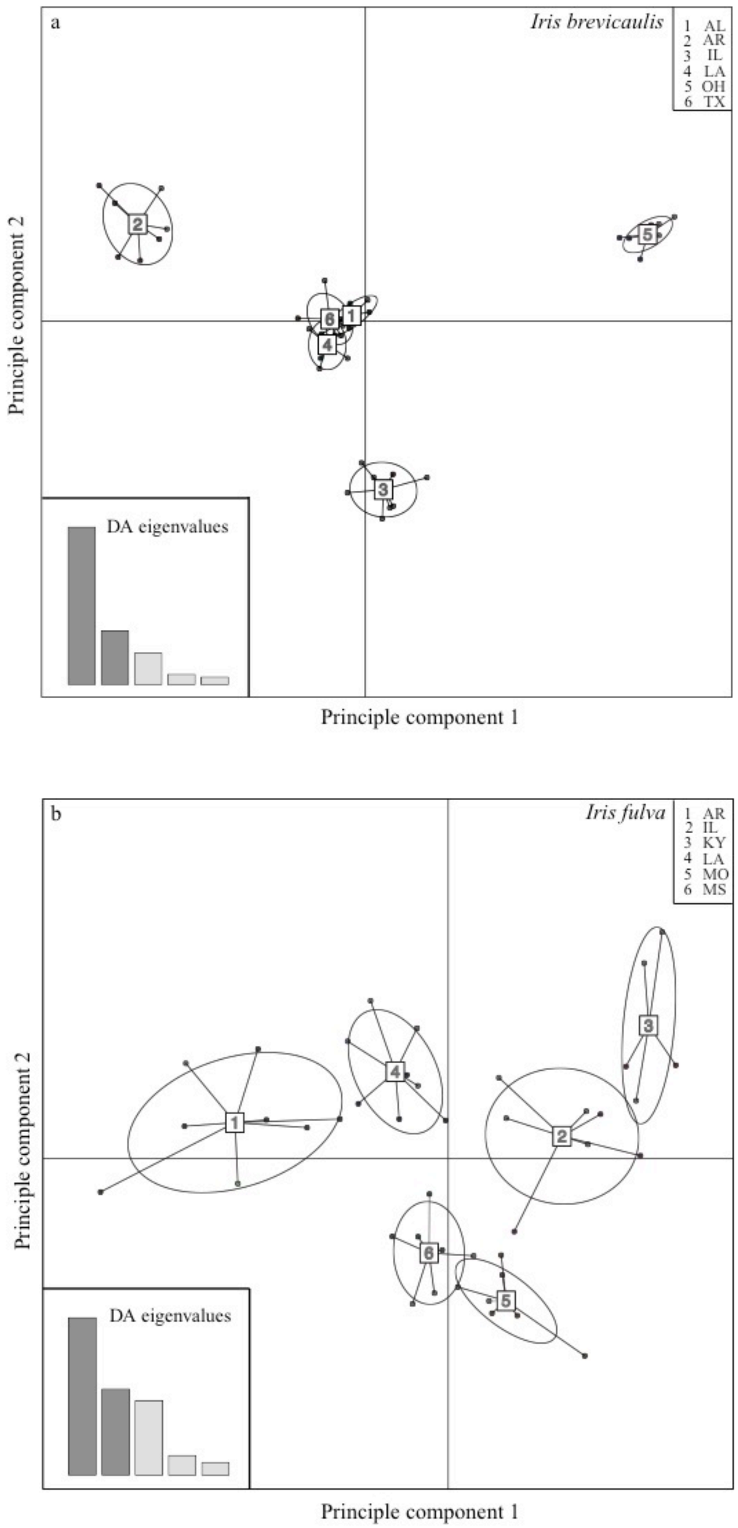

a
Iris brevicaulis
1 AL
2 AR
3 IL
4 LA
5 OH
6 TX
DA eigenvalues
Principle component 2
Principle component 1
b
Iris fulva
1 AR
2 IL
3 KY
4 LA
5 MO
6 MS
DA eigenvalues
Principle component 2
Principle component 1

Figure 2.3. Samples are assigned to their genetic cluster by discriminant analysis of PCs. The bar graph inset displays the eigenvalues of the five principal components in relative magnitude and illustrates the variation explained by the five PCs. The 67% inertia ellipses are drawn for each cluster representing the variance of both PCs. a) Principle component (PC) scatter plot for *I. brevicaulis*. b) Principle component scatter plot for *I. fulva*.

Figure 2.4.

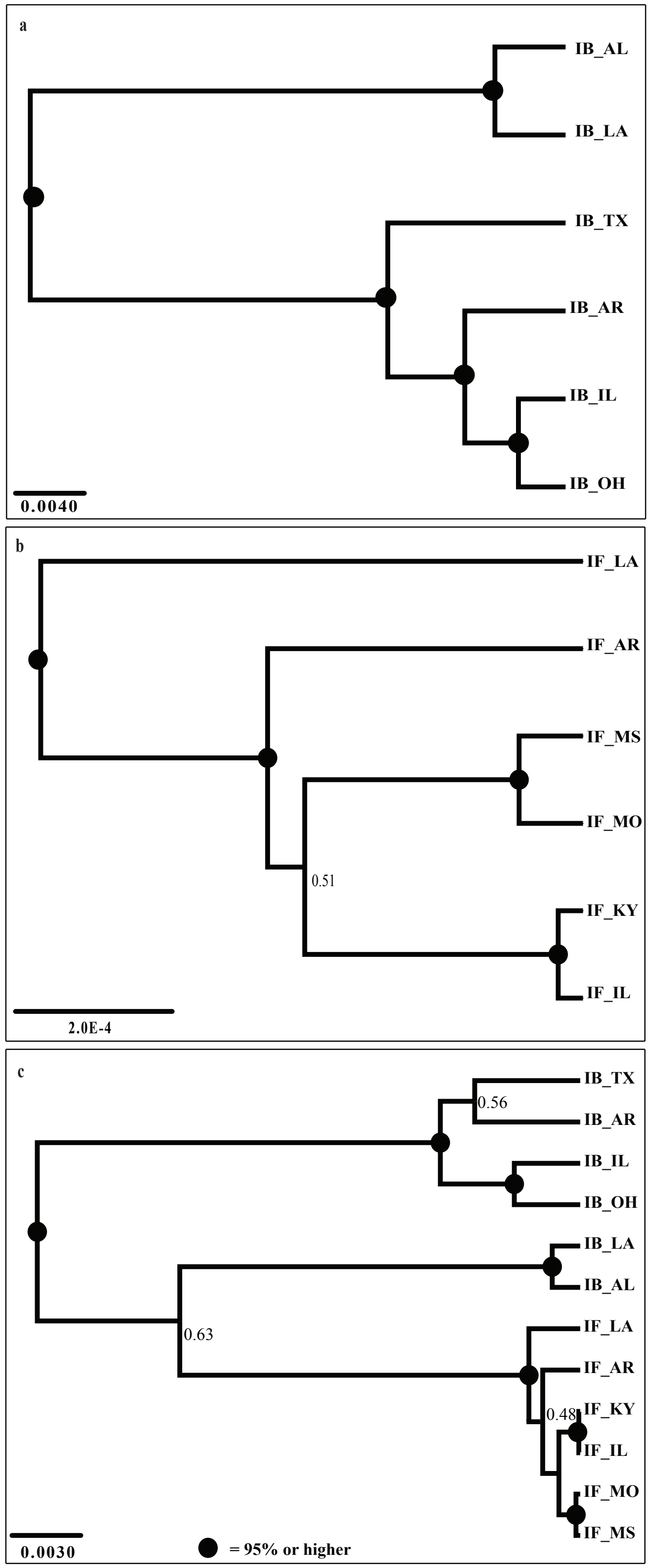

a
IB_AL
IB_LA
IB_TX
IB_AR
IB_IL
IB_OH
0.0040
b
IF_LA
IF_AR
IF_MS
IF_MO
0.51
IF_KY
IF_IL
2.0E-4
c
IB_TX
0.56
IB_AR
IB_IL
IB_OH
IB_LA
IB_AL
0.63
IF_LA
IF_AR
0.48
IF_KY
IF_IL
IF_MO
IF_MS
0.0030
= 95% or higher

Figure 2.4. Species tree inferred using SNAPP with independent biallelic markers. Proportion of posterior support given above each node and nodes with support = 95% or higher have no label. a) *Iris brevicaulis* populations b) *I. fulva* populations c) *I. fulva* and *I. brevicaulis* populations.

CHAPTER 3

NEUTRAL AND SELECTIVE PROCESSES DRIVE POPULATION DIFFERENTIATION FOR *IRIS HEXAGONA* [1]

[1] Hamlin, J.A.P and M. L. Arnold. Submitted to *Journal of Heredity* 03/5/15.

Abstract

Gene flow among widespread populations can be reduced by geographical distance or by divergent selection resulting from local adaptation. In this study, we tested for the divergence of phenotypes and genotypes among eight populations of *Iris hexagona*. Using a genotyping-by-sequencing approach, we generated a panel of 750 single nucleotide polymorphisms and used population genetic analyses to determine what affects patterns of divergence across *I. hexagona* populations. We examined the differences, between populations, for five morphological traits and found that all traits varied significantly. Genetic differentiation was compared between populations at neutral and non-neutral SNPs and we detected significant population differences between the two types of markers. Furthermore, there is an overall relationship between pairwise genetic distance for neutral SNPs, and geographic distance. This suggests that neutral genetic structure is a product of limited migration and drift. However, we show that non-neutral genetic differentiation is correlated with morphological divergence, which may support the adaptive nature of intraspecific floral divergence exhibited by *I. hexagona.* For *I. hexagona*, it appears that both neutral and selective processes are at play in generating population divergence.

Introduction

Natural landscapes are often heterogeneous, which may lead to fragmentation resulting in isolated populations. Genetic divergence and associated morphological differentiation can occur between these populations (Kawecki and Ebert 2004). Understanding the extent to which populations specialize to their local environment provides insight into the relative roles of evolutionary factors.

The observed pattern of population differentiation could be the product of migration, selection, or genetic drift. For example, migration, and the potential for the homogenizing effect of gene flow, tends to occur more commonly between neighboring populations, resulting in a pattern of isolation-by-distance (Wright 1943). Isolation-by-distance (i.e. IBD) is identified by an increase in genetic differentiation among populations with increasing geographic distance as a result of reduced gene flow. This pattern is considered the most simple landscape genetic pattern and is expected in numerous natural systems (Wright 1943; van Strien et al. 2014).

In contrast, if levels of selection are greater than the homogenizing effects of gene flow, populations may become locally adapted (Kawecki and Ebert 2004). This involves genetic divergence of specific loci between populations resulting from contrasting environments, including both biotic and abiotic factors (Savolainen et al. 2013). Thus, evolutionary models predict that low levels of gene flow between populations and strong divergent selection will favor local adaptation among populations, which could, ultimately, drive phenotypic divergence from one generation to the next (Slatkin 1987). There is a large body of work, which supports the ubiquitous nature of local adaptation in natural populations via reciprocal transplant studies and common-garden experiments

(Hereford 2009). Furthermore, simulation studies have also shown that even with considerable gene flow, environmental heterogeneity might cause disruptive selection and result in local adaptation (Yeaman and Whitlock 2011).

However, genetic drift, which leads to random changes in allele frequencies, can render selection less efficient (Haag and Roze 2007). Furthermore, genetic drift is related to isolation by distance in that the random change in population gene frequencies across both space and time is essentially the accumulation of local genetic differences under geographically restricted dispersal. If there is more differentiation in traits or gene frequencies between two populations than we would expect under IBD, then we have reason to infer that natural selection is at work (Wright 1943; Schemske and Bierzychudek 2001).

Most population genetic analyses that are designed to test for divergence between populations are performed using neutral markers (Hartl and Clark 2007). It is also possible to assay population samples for numerous molecular markers and thus test for evidence of selection through the identification of outlier loci (Nosil et al. 2009). These loci may be identified if they display higher than expected differentiation between populations, a pattern consistent with divergent selection (Foll and Gaggiotti 2008). For example, genome scans have been used for a wide variety of tree species to identify candidate gene regions associated with numerous environmental variables or phenotypic differences (Eckert et al. 2010; Tsumura et al. 2012; Alberto et al. 2013). However, one can evaluate evolutionary patterns at outlier versus neutral loci without any *a priori* knowledge of loci or traits of interest (Foll and Gaggiotti 2008; Collin and Fumagalli 2011; Tiffin and Ross-Ibarra 2014). With the large quantity of genomic data, which is currently being collected for many non-model organisms, genome scans of local adaptation are increasingly more common (Tiffin and Ross-Ibarra 2014). Performing a genome scan does not explicitly elucidate the genetic basis of phenotypic trait differences. However, once outlier loci are identified, further investigation could determine whether these loci are likely involved in adaptive divergence (Savolainen et al. 2013).

Many phenotypic differences found across populations have arisen in response to differential selection regimes (Kawecki and Ebert 2004; Collin and Fumagalli 2011). Within plants, selection has been well documented and shown to be influenced by both abiotic and biotic factors such as pollinators or environmental gradients (Fenster et al. 2004; Strauss and Whittall 2006). For example, evidence of selection along a precipitation gradient was found in *Medicago truncatula* (Yoder et al. 2014). Additionally, floral traits, such as color, fragrance, shape, and size, contribute to pollinator syndromes, and these traits are potentially under selection due to pollinator preferences (Fenster et al. 2004). Demonstrating that floral divergence among conspecific wild populations is the outcome of variable selection can be difficult (Herrera and Bazaga 2008). However, combining population genetic approaches with morphological and climatological analyses for the same individuals can afford a test of such hypotheses.
In the present analysis, we ask whether loci with the strongest degree of population genetic differentiation were also the loci with strongest association to morphology or climate, allowing us to test if pollinators drive population differentiation, climate, or some combination of both. We used this approach for *Iris hexagona*, a member of the Louisiana *Iris* species complex and a model system for speciation research (Lexer and

Widmer 2008). *Iris hexagona* is an ideal candidate in order to focus on the role of intraspecific variation due to its large geographic distribution (Figure 1). It is the most southerly occurring iris species in the United States with a coastal distribution and is found mostly in open, wet habitats. *Iris hexagona* is most common in the state of Florida, but is found as far west as Louisiana and historically as far north as South Carolina in the United States (Figure 3.1). The flower stalks of *I. hexagona* attain heights of one to two meters and are characterized by large purple/blue flowers and distinctive yellow nectar guides, that are attractive to their primary pollinator, *Bombus* spp. bumblebees (Viosca 1935; Emms and Arnold 2000; Van Zandt and Mopper 2004). Previous work has suggested that *I. hexagona* could be considered two different species: *I. giganticaerulea* located in Louisiana and coastal Alabama and *I. savannarum* located throughout Florida. This differentiation was based on variation in capsule (dried fruit) morphology (Henderson 2002; Meerow et al. 2011). Because of this, we predict that populations will be broadly morphologically structured, which will mirror the genetic structure identified. Furthermore, we seek a first approximation towards identifying if there is evidence of local adaptation for *I. hexagona* by performing an outlier scan on a large population genetic data set. If loci with the strongest degree of genetic differentiation (i.e. outliers) associate with either morphology or climate then this may suggest that pollinators, climate differences, or both are driving population differences in *I. hexagona*.

Materials and Methods

Study system sampling and measurement of phenotypic traits

During the spring of 2013, we collected samples throughout the range of *I. hexagona* (Figure 3.1), specifically, from populations in Florida and Louisiana. Number of individuals collected varied dependent on the size of the population; however, number of individuals collected from each population ranged from 10 – 20. Individuals sequenced included only a subset of those that were collected from each population (Table 3.1). By using populations located at the range extremes, we hope to better elucidate the morphological and genetic differences for this species.

Floral and vegetative measurements were recorded from each flowering individual and included: leaf height, flower stalk height (from the base of the rhizome to base of calyx), sepal blade length, sepal blade width, and area of the nectar guide (i.e., roughly triangular area calculated using the formula 1/2 length × width) (Bouck et al. 2007). Two of the morphological traits were measured in the field, which were length of the tallest leaf and flower stalk height. For quantifying other floral traits, each flower was placed on a standardized white background and photographed with a camera positioned 50 cm above. Trait values were measured using ImageJ (Schneider et al. 2012). For three floral traits (petal length, petal width, and nectar guide), all three floral units were measured and averaged for each genotype (*Iris* flowers are tripartite).

DNA extraction, SNP discovery, and genotyping

Leaf material was collected from individuals sampled for phenotypic traits and additional individuals to increase total population sampling (Table 1) and stored in silica gel for DNA extraction. Extractions were performed using the Qiagen DNeasy plant kit and sent to the Cornell Institute for Genomic Diversity for genotyping-by-sequencing (GBS) (Elshire et al. 2011). Libraries were prepared from 95 *I. hexagona* individuals, and one blank was included (i.e. control) for sequencing. DNA from each individual was separately digested using *Eco*T221 and the fragmented DNA was then ligated to a unique barcoded adaptor and a common adaptor. The resulting libraries were sequenced using single-end 100-bp reads on the Illumina HiSeq 2000 with 48 samples sequenced per lane. The GBS UNEAK analysis pipeline (http://www.maizegenetics.net/gbs-bioinformatics) (Lu et al. 2013), an extension to the Java program TASSEL (Bradbury et al. 2007), was used to identify SNPs from the sequenced GBS library. The following options varied from the default settings: minimum number of times a tag must be present set to 5, maximum good reads per lane set to 500,000,000, call heterozygotes, minimum site coverage at 0.8, minimum taxa coverage at 0.1, and minimum minor allele frequency at 0.01. UNEAK takes raw Illumina sequence files and converts them into individual genotypes. Briefly, reads were retained, and trimmed to 64 base pairs (bp), when they possessed a barcode, cut site, and no 'N's in the first 64 bp of sequence after the barcode. Identical reads were clustered into tags and counts of these tags present in each barcoded individual stored. Pairwise alignment identified tag pairs having a single base pair mismatch and these single base pair mismatches were considered candidate SNPs. Any tag pair that contained more than one mismatch was discarded to minimize SNPs resulting from alignment of paralogous sequences.

Subsequent filtering of tags was done using the program TASSEL 4.0 in order to identify concordant and potentially polymorphic SNPs within the species (Bradbury et al.

2007). Three individuals were not used in identifying SNPs, because those samples failed during sequencing. After removing failed samples and setting a threshold of 20% missing data ('N's), and a minor allele frequency of $>$ 2% (or the minimum frequency at which a common allele must occur), a total of 92 individuals generated a filtered set of 750 SNPs.

Outlier detection

To detect loci that depart from the neutral expectation and are therefore potentially influenced by selection, we used the program BAYESCAN v2.0 (Foll and Gaggiotti 2008). For the analysis, the default settings were used along with prior odds of 10:1 for the neutral model relative to the selective model at each SNP. We set a threshold of $\log_{10}$ PO of > 1.5 (very strong evidence; posterior probability of $>$ 0.97) for a marker to be considered to be under selection. An advantage of the posterior probability approach is that it directly allows for control of false discovery rate (FDR); here the FDR was set at 0.01. We, then, performed all subsequent analyses both including and excluding these outlier loci.

Population structure and differentiation

We computed F_{st} matrices and confidence intervals of pairwise genetic distance between populations with these datasets: 750 SNPs (the total dataset), 736 neutral SNPs, and the 14 outlier SNPs (see results below for identification of outliers) using the program *StAMPP* (Pembleton et al. 2013). *StAMPP* uses the method proposed by Wright (1951) and updated by Weir and Cockerham (1984) in order to calculate F_{st}. We compared the overlap of confidence intervals for two datasets (the total dataset of 750

SNPs with the outlier dataset with 14 SNPs) to determine if pairwise F_{st} values were different.

Discriminant analysis of principle components (DAPC) (Jombart et al. 2010) from the package *adegenet* (Jombart 2008) version 1.2.8 in R (R Development Core Team 2009) was used to investigate population genetic structure. DAPC does not assume any underlying population genetic model and can analyze genetic data from large datasets quickly. Specifically, DAPC is a multivariate method that relies on data transformation using Principal Component Analysis (PCA) prior to Discriminant Analysis on the retained Principal Components. DAPC consists of a two-step procedure. First, prior groups must be defined; however groups are often unknown. This can be achieved by using K-means, a clustering algorithm that finds a given number of groups maximizing the variation between groups. To find the optimal number of clusters in our data, we used sequential K-means clustering with increasing values of K (in our case up to 10) using the function find.clusters in *adegenet*. Different clustering solutions were compared using the Bayesian Information Criterion (BIC) with the optimal clustering solution corresponding to the lowest BIC. Second, in order to describe the relationships between the clusters, the retained PCs were submitted to a Discriminant Analysis (DA), based on the groups identified during the preliminary K-means clustering step. During this step, retaining too many PCs could lead to over-fitting the discriminant functions. The optimal number of PCs retained was N/3 where N = number of samples, as advised in the manual.

Morphological and environmental divergence

First, we tested whether population mean values for each trait differed significantly using an ANOVA. To visualize the level of morphological variation across populations, we conducted a PCA of the five morphological measurements and plotted the results using the function *ggbiplot* in R. Next, we identified allelic associations with morphology by running a general linear model using the first two principal components and the population allele frequencies for the 14 outlier SNPs. Pearson's correlations and pairwise comparisons were also calculated for all five traits. For traits that were not correlated, we ran a linear model to assess if the outlier loci associate with the uncorrelated traits, specifically.

For understanding population-level environmental differences, we downloaded raster-formatted climate variables (19 Bioclimatic variables) at 30-arc seconds from the Worldclim database in R and data for four land cover layers derived from satellite-borne remote sensors (NASA-MODIS/Terra data set). Land cover layers included the normalize difference vegetation index (NDVI; measure of vegetative greenness), the yearly standard deviation of NDVI (NDVISTD), QSCAT that provides a measure of vegetation canopy structure and moisture and, in areas of sparse vegetation, soil roughness and wetness, and the percentage of tree cover (TREE) (Wooten and Gibbs 2012). Within R, we extracted values for all 23 ecological variables for each collection site used in this study. To identify the subset of variables that best summarize the range of environments occupied by *I. hexagona*, we performed a correlation analysis in which we kept environmental variables with a Pearson correlation coefficient of $r \leq 0.90$ and than performed a principal component analysis (Supplemental Table 3.1). We, then, determined the variables with the highest contributions on the first three principal components. The first three principal components captured 91.02% of the variation in the climate data. We utilized a general linear model to assess if any of the outlier loci were also associated with climate variables principal components 1 and 2.

Finally, we asked if the number of loci with the strongest degree of population differentiation and the loci with strongest association to morphology or climate was greater than expected by chance using a Fisher's exact test and a subset of randomly sampled neutral SNPs.

Distance matrices and mantel tests

We calculated Euclidean distance matrices for geographic locations between all pairs of populations and for morphology between all pairs of populations. Euclidean distance matrix of morphology was generated using standardized values per populations of the phenotypic traits thus obtaining a single phenotypic distance matrix. We used Mantel tests, as implemented in the R package *Ecodist* (Goslee and Urban 2007) to test if neutral and outlier SNPs were related to geographic or phenotypic distance.

Results

SNP discovery and genotyping

A total of 65,442 unfiltered SNPs were identified for individuals from all populations of *I. hexagona*. After filtering out failed samples and setting a threshold of missing data $\leq 20\%$ with a minimum allele frequency of $>2\%$, 750 SNPs were retained for 92 individuals. To determine whether the tag pairs were derived from nuclear or chloroplast DNA, a subset of the tag pairs (n = 200) were blasted against the NCBI nucleotide database using BLASTN. Almost all tag pairs examined resulted in E-value scores which were not lower than 10^{-2}; however one tag pair matched genomic DNA from *Oryza sativa* with an E-value of 4 x 10^{-5} (data not shown). However, zero tag pairs matched chloroplast DNA. From this, we assume that the tag pairs are derived from nuclear DNA, as we did not identify tag pairs that match sequences from highly conserved chloroplast DNA.

Outlier detection

From the BAYESCAN analyses, a total of 14 loci out of 750 (1.86%) were identified to have an F_{st} value higher than expected when the False Discovery Rate was set at 0.01. These loci were inferred to be under diversifying selection (Table 3.2). We found a significant effect from these 14 outlier loci on the range of estimates of F_{st} (Table 3.3a and 3.3c).

Population structure and differentiation

Using a panel of 750 SNPs and 92 individuals, we estimated pairwise F_{st} to range from 0.085 to 0.392 when we used all 750 SNPs; this estimate is similar for 736 neutral SNPs (Table 3.3a and 3.3b). We estimated F_{st} for the 14 outlier SNPs and examined the overlap of confidence intervals with the total dataset of 750 SNPs. For a number of comparisons, there is no overlap in confidence intervals, suggesting a significant difference for the range of values of F_{st} using the total dataset (750 SNPs) compared to the 14 outlier SNPs (F_{st} values bolded; Table 3.3a and 3.3c).

From the DAPC analyses, the BIC reached its minimum value at K = 5 suggesting five genetic clusters (Appendix B, Figure S1). The two Louisiana populations (LA_02 and LA_04) were assigned as a single population. The two Florida populations, which were geographically very close together (FL_10 and FL_11), were grouped as a single population. Finally, FL_04 and FL_13 were grouped as a single population. Distinct separate genetic clusters are identified and population genetic structure coincided with the population's geographic origin (Figure 3.2a). The only exception, to population clusters coinciding with geographic origin, is the genetic cluster that consists of FL_04 and FL_13, which span a large region in Florida (Figure 3.1). The first principal component clearly separates FL_01 and FL_08 from FL_04, FL_10, FL_11, and FL_13 and the Louisiana genetic cluster (LA) (Figure 2a). The second principal component clearly separates the Louisiana genetic cluster (LA) and FL_10 & FL_11 from the other Florida populations (Figure 3.2a).

Morphological and environmental divergence

From an ANOVA, testing for the effect of population-level trait variation, we found that all traits were significantly different among populations (Leaf height: p = 0.0069, stem height: p = 5.2e-10, petal length: p = 0.0044, petal width: p = 2.22e-05, and nectar guide: p = 3.97e -06). Compared to genetic differentiation, morphological variation was not as highly structured. However, we do see evidence of the Louisiana genetic cluster (red circles) separated from the other Florida populations on principal component 1 (Figure 3.2b). The first two principal components explained 52.2% and

28% of the total variation, respectively (Figure 3.2b). Petal length was the major loading of PC1 (-0.523). Leaf height was the major loading of PC2 (-0.710).

When we tested for an association with outliers and morphology principal components 1 and 2, we found a number of outliers to be associated. Specifically, outliers 71, 117, 349, and 480 were significantly associated with principal component 1 (Appendix B, Table S2a). For principal component 2, outlier 184 was significantly associated. To determine if the number of outliers associated with morphology is greater than expected by chance, we determined population allele frequencies for 15 randomly sampled SNPs and performed a two-tailed Fisher's exact test. We found that our association of four SNPs is greater than expected by chance **(**Fisher's exact test = 0.0365). This was only performed on morphology principal component 1, as there was only one association with principal component 2 (Appendix B, Table S2b).

There was no correlation of leaf height with any other trait (Table 3.4). All other trait comparisons were significantly correlated with each other (Table 3.4). We, then, tested for an association with population allele frequencies of outlier SNPs with leaf height and petal width, as petal width is correlated with all other floral traits (Table 3.4). Thus, if a SNP is significantly associated with one floral trait (petal width), then it should be associated with all other floral traits. We found that SNPs 71, 117, 349, and 480 were significantly correlated with petal width (SNP 71 $p = 0.0042$, SNP 117 $p = 0.0042$, SNP 349 $p = 0.0036$, SNP 480 $p = 0.0042$ with Bonferroni correction). We, then, plotted the frequency of the outlier SNPs to determine the pattern of allele frequencies across populations (Appendix B, Figure S2). For SNP 71, 117, and 480, these SNPs are fixed for one allele in the Louisiana populations and completely fixed for the alternate in the

Florida populations. For SNP 575, the Louisiana populations are fixed for one allele, while the frequency of this SNP is variable in the Florida populations.

We identified population-level differences for environmental variation. The environmental variables that contributed the most to principal component 1 were bio 2 or mean diurnal temperature range (0.499). The variable that contributed the most to principal component 2 was bio 9 or mean temperature of the driest quarter (-0.588). Finally, the variable that contributed the most to principal component 3 was QSCAT (-0.403), which provides a measure of vegetation canopy structure and moisture and, in areas of sparse vegetation, soil roughness and wetness. In terms of environmental space, populations, which are in close geographic proximity, are also close in principal component space (e.g. FL 10 and FL 11). When we tested for an association between outlier SNPs and climate principal components, we did not find any SNPs that significantly associated with principal component 1. Nor where outliers were significantly associated with principal component 2.

Distance matrices and mantel tests

There is an overall relationship between pairwise neutral genetic distance and geographic distance (Mantel test, R = 0.612, p = 0.011), suggesting that isolation-by-distance is causing the neutral, genetic structure (Appendix B, Figure S3a). This is further confirmed when we test for isolation-by-distance only in the Florida populations with neutral genetic distance (Mantel test, R = 0.772, p = 0.004). However, we did find a positive association between non-neutral genetic distance and geographic distance (Mantel test, R = 0.534, p = 0.016) (Appendix B, Figure S3b). The regression of mean floral traits on genetic distance between populations based on neutral loci is not greater than expected by chance (Mantel test, $R = 0.406$, $p = 0.056$) (Appendix B, Figure S3c). In contrast, we detected a moderate and nearly significant, correlation between mean floral traits and the genetic distance between populations based on non-neutral SNPs (Mantel test, $R = 0.299$, $p = 0.05$) (Appendix B, Figure S3d).

Discussion

In this study, we sought to identify factors affecting intraspecific variation within *I. hexagona*, one of the species within the Louisiana iris species complex. Using 750 SNPs, we detected a small proportion of loci (1.86%) that demonstrated significant frequency shifts between populations (i.e. "outlier loci"). These loci exhibited higher F_{st} values than expected and were considered to be under diversifying selection. When the 14 outlier SNPs were removed prior to analysis, we saw a decrease across all population pair-wise values of F_{st}; however, not all values were significantly different from the estimate of F_{st} using all 750 SNPs. This result indicates that these outlier SNPs are the largest contributors to the genetic differentiation observed between these populations. Thus, when examining only the 736 neutral SNPs, we estimate low to moderate levels of genetic differentiation across all pairwise populations.

Floral morphology was variable both among and within populations. Specifically, Louisiana populations are distinctly separate from Florida populations in morphological space. Population allele frequencies, for a few of the outlier SNPs (71, 117, 349, and 480) were significantly correlated with principal component 1 of morphology. Furthermore, these same SNPs correlated with petal width. Populations exhibited distinct environmental differences when examining the 23 environmental variables; however, no outlier SNPs correlated with environmental principal components.

The SNPs putatively under selection were positively correlated with both geographic distance and phenotypic distance, while neutral SNPs were only correlated with geographic distance. Isolation-by-distance was observed for neutral SNPs suggesting that patterns of neutral genetic structure across the *I. hexagona* populations are the product of limited migration and gradual genetic drift. However, isolation-by-distance is not acting on the SNPs putatively under selection and thus, selection may be acting on the genomic regions surrounding these SNPs. Our data suggest that both deterministic and neutral processes have contributed to the evolutionary trajectory of *I. hexagona* populations.

Geographic population assignment

Barriers to gene flow have been inferred for a number species distributed along the southeastern U.S. as well as those with a coastal distribution (Soltis et al. 2006). Genetic breaks along the gulf have been attributed to an east – west division, which could be due to either the Apalachicola or Tombigbee Rivers (Soltis et al. 2006). For *I. hexagona*, the general location of this regional break occurs potentially somewhere along the panhandle of Florida, where these rivers co-occur. That the Apalachicola River acts as a phylogeographic break has been inferred for in species such as pitted striped-seed (*Piriqueta caroliniana*) (Maskas and Cruzan 2000). Additionally, support for the occurrence of a coastal east-west barrier to gene flow also emerged from an analysis of contact zones, which could be interpreted as contact areas between closely related species or populations (Swenson and Howard 2005). Future studies of *I. hexagona* will include additional populations along the panhandle of Florida, thereby testing whether this region demonstrates a transition zone (either genetically or morphologically) between the Florida and Louisiana populations examined here. The genetic differences we see for *I. hexagona* could represent early stages of populations diverging due to long-term isolation. In addition, local adaptation to environmental differences not included in this study (e.g. salinity tolerance) could increase selection against potential migrants.

Neutral or selective loci

Despite the spatial genetic structuring of *I. hexagona* populations revealed by the analyses based on all loci, 14 loci exhibited strong non-neutral signatures and were presumably selected or linked to selected regions of the genome. Additionally, when we examined patterns of allele frequencies for four SNPs, which were significantly associated with a morphological trait, we detected extreme allele frequency differences between different geographic regions (Florida versus Louisiana). It has been suggested that extreme allele frequency differences or a correlation between allele frequencies and important ecological variables may be involved in local adaptation (Coop et al. 2010). Previous analyses of genome scans adopted the idea that differentiation can be maintained in a small portion of the genome, even while extensive gene exchange continues, preventing divergence across most of the genome (Nosil et al. 2009). However, Cruickshank and Hahn (2014) found that a lack of divergence at neutral loci could be easily produced by the maintenance of ancestral polymorphism rather than gene exchange. Here, it appears for *I. hexagona* there is a potential for gene flow between populations that are in geographic proximity, however the lack of divergence could be at least partially the result of shared ancestral polymorphism.

Adaptive floral divergence

In terms of understanding the adaptive divergence associated with outlier loci, a number of studies have found an association between such loci and climatic factors or phenotypic traits. For example, Eckert et al. (2010) found loblolly pine outlier loci to be associated with aridity or temperature. For *I. hexagona*, the finding of a significant relationship between variation at outlier loci and variation in floral traits provides evidence that selection may be acting on phenotypic divergence within this species. If similar correlation patterns between phenotypic and genetic distances had been obtained for both neutral and selected markers, factors other than selection would have been inferred. Instead, our findings lead us to conclude that the differences among *I. hexagona* populations in floral traits have a genetic basis and reflect local adaptation that has likely arisen via divergent selection. It is not possible at present to ascertain the selective agents and mechanisms of selection ultimately responsible for the potential adaptive floral divergence in *I. hexagona.* However, pollinator-mediated selection on components of flower size and shape has been suggested in a number of other species (Schemske and Bradshaw 1999; Herrera and Bazaga 2008). In addition, other biotic and abiotic factors can exert direct or indirect selection on floral features (e.g. water stress or salt tolerance) (Strauss and Whittall 2006; Zhang et al. 2011).

Though we were not able to determine whether specific outlier loci are associated with a particular floral trait (i.e. because petal width, petal length, and nectar guide are correlated), four SNPs do show a positive association with petal width suggesting their contribution to the evolution of this trait. Finally, the pattern of variation in this large population genetic dataset leads to the inference that selection does not impact the majority of loci. It, thus, appears that both neutral and selective processes have been important in the evolution of *I. hexagona*, resulting in both isolation-by-distance and adaptive divergence. Furthermore, future reciprocal transplant experiments involving populations found in Florida and Louisiana should allow a direct test of local adaptation and infer explicit targets of selection.

Acknowledgements

JAPH would like to acknowledge the many governmental agencies that provided valuable information regarding natural populations. JAPH would also like to thank members of the Wares and Sweigart lab for fruitful discussions regarding this manuscript. Funds from the American Iris Society supported the collections for this project. This work was supported by the National Institute of Health under award number T32GM007103 and National Science Foundation grants DEB-0949479/0949424 (collaborative grant with NH Martin of Texas State University) and DEB-1049757.

Literature Cited

Alberto, F. J., J. Derory, C. Boury, J.-M. Frigerio, N. E. Zimmermann, and A. Kremer. 2013. Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of Quercus petraea. Genetics 195:495-+.

Bouck, A., S. R. Wessler, and M. L. Arnold. 2007. Qtl analysis of floral traits in Louisiana Iris hybrids. Evolution 61:2308-2319.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Collin, H. and L. Fumagalli. 2011. Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (Phoxinus phoxinus, Cyprinidae). Molecular Ecology 20:4490-4502.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics 185:1411-1423.

Cruickshank, T. E. and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133-3157.

Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. Gonzalez-Martinez, and D. B. Neale. 2010. Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). Genetics 185:969-982.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Plos One 6.

Emms, S. K. and M. L. Arnold. 2000. Site-to-site differences in pollinator visitation patterns in a Louisiana iris hybrid zone. Oikos 91:568-578.

Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson. 2004. Pollination syndromes and floral specialization. Annual Review of Ecology Evolution and Systematics 35:375-403.

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

Goslee, S. C. and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22:1-19.

Haag, C. R. and D. Roze. 2007. Genetic load in sexual and asexual diploids: Segregation, dominance and genetic drift. Genetics 176:1663-1678.

Hartl, D. L. and A. G. Clark. 2007. Principles of Populations Genetics. Sinauer Associates Sunderland.

Henderson, N. C. 2002. *Iris*. In: Flora of North America Editorial Committee, eds. Oxford University Press, New York.

Hereford, J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. American Naturalist 173:579-588.

Herrera, C. M. and P. Bazaga. 2008. Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. Molecular Ecology 17:5378-5390.

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

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

Kawecki, T. J. and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology Letters 7:1225-1241.

Lexer, C. and A. Widmer. 2008. The genic view of plant speciation: recent progress and emerging questions. Philosophical Transactions of the Royal Society B-Biological Sciences 363:3023-3036.

Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney, M. D. Casler, E. S. Buckler, and D. E. Costich. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. Plos Genetics 9.

Maskas, S. D. and M. B. Cruzan. 2000. Patterns of intraspecific diversification in the Piriqueta caroliniana complex in southeastern North America and the Bahamas. Evolution 54:815-827.

Meerow, A. W., M. Gideon, D. N. Kuhn, S. Mopper, and K. Nakamura. 2011. The genetic mosaic of iris series Hexagonae in Florida: inferences on the holocene history of the Louisiana irises and anthropogenic effects on their distribution. International Journal of Plant Sciences 172:1026-1052.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375-402.

Pembleton, L. W., N. O. I. Cogan, and J. W. Forster. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources 13:946-952.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Savolainen, O., M. Lascoux, and J. Merila. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics 14:807-820.

Schemske, D. W. and P. Bierzychudek. 2001. Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. Evolution 55:1269 - 1282.

Schemske, D. W. and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proceedings of the National Academy of Sciences of the United States of America 96:11910-11915.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9:671 - 675.

Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural Populations. Science 236:787-792.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15:4261-4293.

Strauss, S. Y. and J. B. Whittall. 2006. Non-pollinator agents of selection on floral traits *in* L. D. Harder, and S. C. H. Barrett, eds. Ecology and Evolution of Flowers. Oxford University Press.

Swenson, N. G. and D. J. Howard. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. American Naturalist 166:581-591.

Tiffin, P. and J. Ross-Ibarra. 2014. Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution 29:673-680.

Tsumura, Y., K. Uchiyama, Y. Moriguchi, S. Ueno, and T. Ihara-Ujino. 2012. Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, Cryptomeria japonica. Heredity 109:349-360.

van Strien, M. J., R. Holderegger, and H. J. Van Heck. 2014. Isolation-by-distance in landscapes: considerations for landscape genetics. Heredity.

Van Zandt, P. A. and S. Mopper. 2004. The effects of maternal salinity and seed environment on germination and growth in Iris hexagona. Evolutionary Ecology Research 6:813-832.

Viosca, P. J. 1935. The irises of southeastern Louisiana - a taxonomic and ecological interpretation. Bulletin of the American Iris Society 57:3-56.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of poulation-structure. Evolution 38:1358-1370.

Wooten, J. A. and H. L. Gibbs. 2012. Niche divergence and lineage diversification among closely related Sistrurus rattlesnakes. Journal of Evolutionary Biology 25:317-328.

Wright, S. 1943. Isolation by distance. Genetics 28:114 - 138.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Yeaman, S. and M. C. Whitlock. 2011. The genetic architecture of adaptation under migration-selection balance. Evolution 65:1897-1911.

Yoder, J. B., J. Stanton-Geddes, P. Zhou, R. Briskine, N. D. Young, and P. Tiffin. 2014. Genomic Signature of Adaptation to Climate in Medicago truncatula. Genetics 196:1263.

Zhang, Z., S. Zhang, Y. Zhang, X. Wang, D. Li, Q. Li, M. Yue, Q. Li, Y.-e. Zhang, Y. Xu, Y. Xue, K. Chong, and S. Bao. 2011. Arabidopsis Floral Initiator SKB1 Confers High Salt Tolerance by Regulating Transcription and Pre-mRNA Splicing through Altering Histone H4R3 and Small Nuclear Ribonucleoprotein LSM4 Methylation. Plant Cell 23:396-411.

Table 3.1. Collection information for *I. hexagona* populations

Population ID	Latitude	Longitude	n
FL_01	25.790	-81. 100	12
FL_04	27.267	-82.121	12
FL_08	26.944	-81.449	12
FL_10	28.463	-82.054	12
FL_11	28.507	-82.124	13
FL_13	29.983	-81.675	11
LA_02	29.880	-91.784	11
LA_04	30.084	-90.449	12

n = number of individuals

Table 3.2. 14 outliers identified from 750 SNPs

SNPID	Prob*	Alpha	F_{st}
16	0.983	1.3333	0.62981
71	0.9974	1.8353	0.72171
117	0.9962	1.8104	0.71708
172	0.9958	1.4317	0.65004
184	0.9964	1.5854	0.67854
294	0.9986	1.6781	0.69651
330	0.9788	1.4377	0.64931
349	0.9798	1.4638	0.65427
381	0.9842	1.5478	0.67065
480	0.998	1.8041	0.71655
575	0.9862	1.5046	0.66269
605	0.99	1.2707	0.61786
647	0.9994	1.7354	0.7074
733	0.9838	1.323	0.62773

*Prob is the posterior Bayes probability and a positive Alpha value indicates diversifying selection.

Table 3.3a. Pairwise F_{st} values for all populations using all 750 SNPs

	FL_01	FL_04	FL_08	FL_10	FL_11	FL_13	LA_02
FL_01	--						
FL_04	**0.224**	--					
FL_08	0.234	**0.232**	--				
FL_10	0.298	**0.186**	0.302	--			
FL_11	0.295	**0.172**	0.283	0.082	--		
FL_13	**0.199**	**0.135**	**0.219**	**0.156**	**0.138**	--	
LA_02	**0.328**	**0.227**	**0.356**	**0.304**	**0.276**	**0.182**	--
LA_04	**0.370**	**0.259**	**0.392**	**0.335**	**0.313**	**0.212**	0.0854

Table 3.3b. Pairwise F_{st} values for all populations using 736 putatively neutral SNPs

	FL_01	FL_04	FL_08	FL_10	FL_11	FL_13	LA_02
FL_01	--						
FL_04	0.217	--					
FL_08	0.232	0.226	--				
FL_10	0.294	0.176	0.298	--			
FL_11	0.291	0.163	0.279	0.081	--		
FL_13	0.196	0.132	0.216	0.152	0.135	--	
LA_02	0.317	0.221	0.346	0.292	0.267	0.180	--
LA_04	0.359	0.253	0.382	0.322	0.303	0.209	0.085

Table 3.3c. Pairwise F_{st} values for all populations using 14 putatively non-neutral SNPs

	FL_01	FL_04	FL_08	FL_10	FL_11	FL_13	LA_02
FL_01	--						
FL_04	**0.658**	--					
FL_08	0.510	**0.639**	--				
FL_10	0.592	**0.691**	0.622	--			
FL_11	0.574	**0.627**	0.543	0.137	--		
FL_13	**0.399**	**0.310**	**0.397**	**0.356**	**0.301**	--	
LA_02	**0.789**	**0.615**	**0.781**	**0.766**	**0.678**	**0.320**	--
LA_04	**0.829**	**0.641**	**0.829**	**0.809**	**0.735**	**0.373**	0.098

BOLD are those, which have non-overlapping confidence intervals between the total dataset of 750 SNPs (3a) and the 14 putatively non-neutral SNPs (3c) and thus are significantly different.

Table 3.4. Correlations among the five morphological traits measured

	Petal.length	**Petal.width**	**Nectar.guide**	**Leaf.height**	**Stem.height**
Petal.length	--	**0.006**	**0.02**	0.213	**0.023**
Petal.width	**0.858**	--	**0.015**	0.373	**0.0012**
Nectar.guide	**0.766**	**0.806**	--	0.856	**0.058**
Leaf.height	0.494	0.365	0.077	--	0.209
Stem.height	**0.777**	**0.919**	**0.688**	0.497	--

Below diagonal is the Pearson's product correlation coefficient. Above the diagonal is associated p-value. Significant correlations and the associated p-value are in **bold**

Figure 3.1.

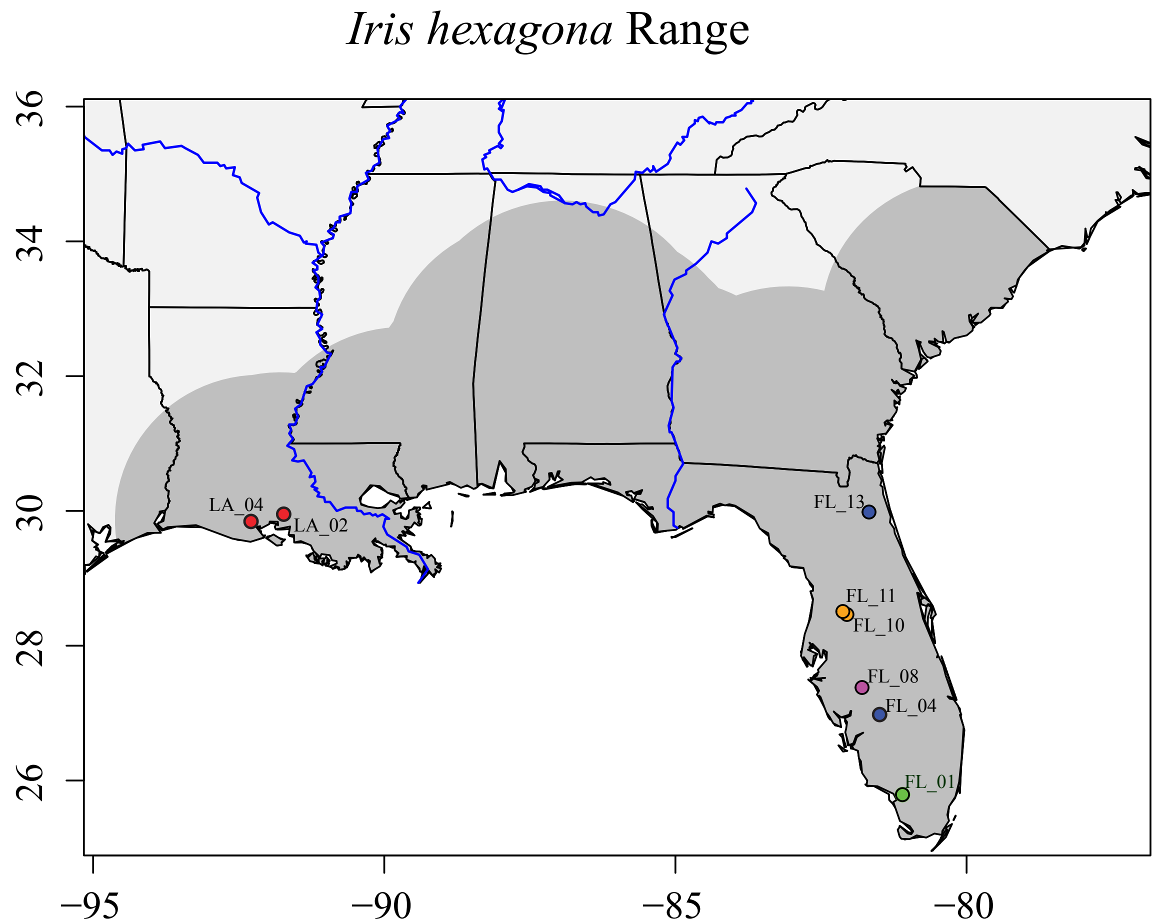
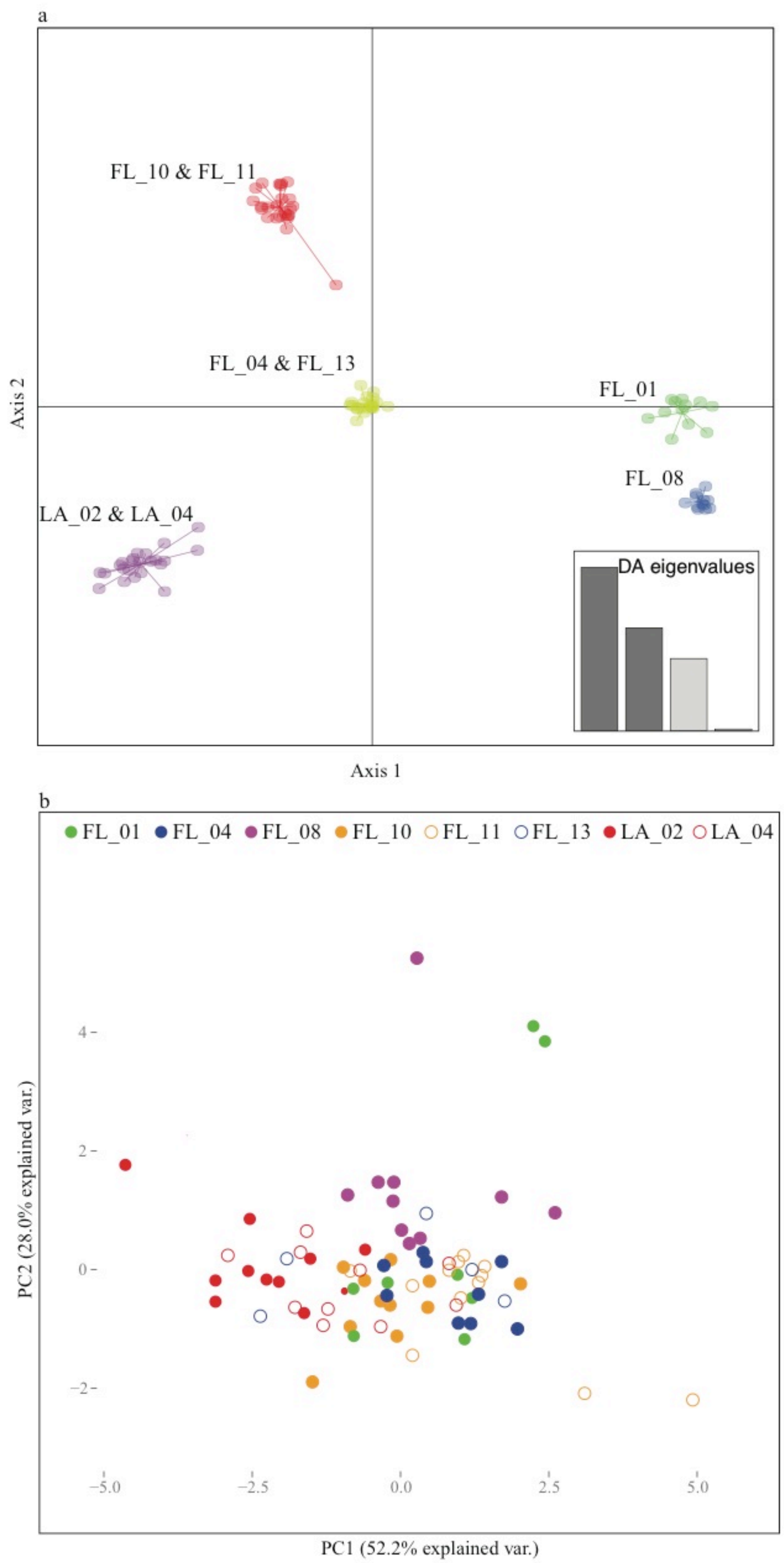


Figure 3.1. Range distribution map of *Iris hexagona.* Colored dots indicate collection localities used in this study. Major rivers for the region are outlined. Population codes and associated colors are used throughout other figures.

Figure 3.2.



a
FL_10 & FL_11
FL_04 & FL_13
FL_01
FL_08
LA_02 & LA_04
DA eigenvalues
Axis 2
Axis 1
b
FL_01
FL_04
FL_08
FL_10
FL_11
FL_13
LA_02
LA_04
4
2
0
−2
−5.0
−2.5
0.0
2.5
5.0
PC2 (28.0% explained var.)
PC1 (52.2% explained var.)

Figure 3.2. a) Samples are assigned to their genetic cluster by discriminant analysis of PCs. The bar graph inset displays the eigenvalues of the four principal components in relative magnitude and illustrates the variation explained by the four PCs. b) Principal Component Analysis of five traits recorded for *I. hexagona* samples. The total amount of variance explained is 80.2% and within and among population differences for floral traits is variable. Individuals are color-coded the same as their determined population genetic cluster in 2a.

Figure 3.3.

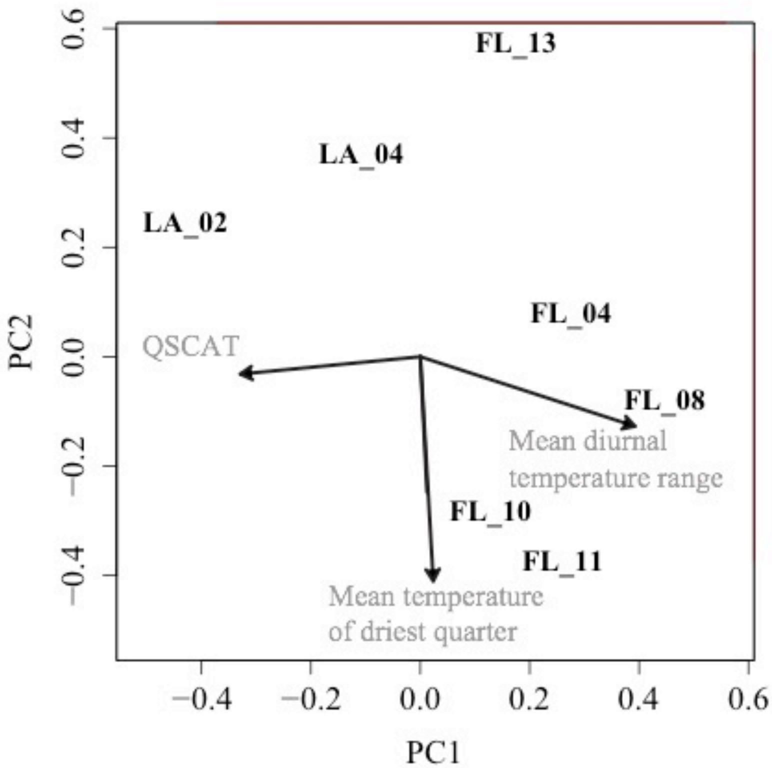


Figure 3.3. Principal component analysis of 23 environmental variables for *Iris hexagona* populations. Population-level differences in environmental preferences are shown in that no populations overlap. One exception is FL_10 and FL_11, which are relatively close together both in principal component space and geographic proximity.

CHAPTER 4

NICHE DIVERGENCE IMPLICATES ECOLOGICAL SPECIATION FOR THE RECENTLY DIVERGED LOUISIANA IRISES[1]

[1] Hamlin, J.A.P, T.J. Simmonds, and M. L. Arnold. Submitted to *Journal of Biogeography* 04/7/15.

Abstract

Spatial and temporal environmental variation influences evolutionary processes such as divergence among populations and species. Studies that address processes affecting ecological niches over evolutionary time are crucial for understanding how and when environmental factors may influence lineage diversification. Here, we seek to address the role ecology has played in the speciation of the Louisiana irises as well as understanding the timing of divergence between these species. This study incorporated phylogenetic methods with environmental information to clarify relationships and ecological niches of the Louisiana iris species complex. A species phylogeny and divergence time estimates were generated for the Louisiana irises using *BEAST. Ecological Niche Models (ENMs) were constructed for present day distributions, to test for the environmental factors that contributed to species divergence. Using the ENMs, we then tested for overlap in ecological niches. We find that the Louisiana irises are of recent origin. Given this recent origin, niche divergence was implicated for all species comparisons and thus ecology likely played an important role in the diversification process in these plants.

Introduction

It has been suggested that ecology plays an important role in the origin of new species (Darwin 1859) and thus there is a renewed interest in understanding how ecological setting may contribute to speciation (Schluter 2001, 2009). Wiens (2005) concluded that the occurrence of incipient species in different environments or utilizing different resources (e.g. micro-habitats) reflected the importance of ecology during speciation. In this regard, one can imagine how reproductive isolation could evolve as a result of divergent selection on resource or habitat use.

Studies that address how ecological niches change over evolutionary time are crucial for understanding how and when environmental factors influence natural selection (Wiens 2004; Hickerson et al. 2010; Blair et al. 2013). Analyzing niche relationships of closely related taxa can provide insight into the ecological distinctiveness and potential mechanisms responsible for reproductive isolation and diversification (Wooten and Gibbs 2012). Specifically, such analyses provide a test of whether niche conservatism or niche divergence is associated with lineage diversification. If niche divergence is implicated then this would support a role of ecological speciation in which divergent natural selection promotes diversification through adaptations to new environments (Schluter 2001, 2009; Nosil 2012). Alternatively, evidence that niche characteristics are conserved (i.e. niche conservatism; Harvey and Pagel 1991) would suggest that ecological differences accrue only after speciation as reflected by models in which species evolve in allopatry (Mayr 1942; Wiens and Graham 2005). Thus, combining phylogenetic information with ecological models allows one to assess if closely related taxa show evidence of ecological niche divergence or conservatism.

Evolutionary biologists are now able to utilize long-term environmental data via Ecological Niche Models (ENMs) and tests of similarity (Warren et al. 2008) to address niche evolution. The ENM can be projected onto geographic space to depict the relative occurrence rate for each cell in a geographic area, thereby depicting the potential area for a species' distribution and ecological niche. Analyzing the map generated by projecting the ENM onto a geographic space can give a basic understanding of the abiotic factors that affect the distribution of a given species and can inform further analyses into the specific factors that contribute to the distribution of a species (Alvarado-Serrano and Knowles 2014) and evolution of ecological niches (Kozak and Wiens 2006). Abiotic variables are not the only factors that contribute to the distribution of a species, but they do provide an adequate first approximation (Pearson and Dawson 2003). Thus, caution is warranted as ENMs may not capture all dimensions relevant to defining a species' ecological niche (Pearson and Dawson 2003). However, this spatially explicit environmental data has increasingly been used as an independent source to evaluate or develop hypotheses to broadly understand the role ecology plays in speciation (Alvarado-Serrano and Knowles 2014). More specifically, the increased use of ENMs has addressed topics as varied as: whether lineage divergence accompanies species niche divergence, biogeographic hypotheses about community responses to climate change, and identification of species ranges during the Last Glacial Maximum (Alvarado-Serrano and Knowles 2014).

Furthermore, Warren et al. (2008) developed a statistical framework that takes advantage of niche model output and permits quantitative comparisons of ENMs with comparative similarity measures. Niche similarity is assessed via two tests: identity and background. The niche identity test asks whether two ENMs generated from two populations or species are identical, with the assumption that no two populations niches will be exactly identical given local environmental heterogeneity (Warren et al. 2008; Warren et al. 2010). When the hypothesis of niche identity is rejected, there is assumed to be some level of niche differentiation; however, this could be a result of the spatial heterogeneity of environmental variables found across a landscape. The background similarity test then asks whether ENMs drawn from populations with partially or entirely non-overlapping distributions are any more different from one another than expected by chance (Warren et al. 2008; Warren et al. 2010). Using these two quantitative tests of niche overlap allows one to determine if there is evidence of niche conservatism or niche divergence for the species of interest. However, if species distributions don't overlap spatially then differences in the niches may not have an adaptive explanation, but instead reflect that the taxa of interest simply occupy different geographic regions. As such, there are several considerations that are essential in order to correctly interpret niche conservatism or divergence in sibling species 1) compared taxa should be the most likely sister species, 2) the sister species should be located in a region of probable historical accessibility for both species, and 3) niche predictor variables must be selected based on biological significance and absence of statistical redundancy (Ibarra-Cerdeña et al. 2014).

Ecological Niche Models have been applied to a multitude of plant taxa at various population sizes, ploidy levels, and conservation statuses (Wilson et al. 2005; Gallien et al. 2010; Jakob et al. 2010). One reason plant taxa have so ubiquitously been used in studies of species distributions and ecological niches is due to well catalogued collection records, limited dispersal which may allow for strong responses to local environmental change, and generally well-understood species relationships (Phillips et al. 2004). The Louisiana irises are a promising group for applying ENMs to assess the impact of environmental factors on lineage diversification. Viosca (1935) and Riley (1938) delineated four species within the Louisiana species complex: *Iris brevicaulis* Raf, *I. fulva* Ker-Gawler*, I. hexagona* Walt., and *I. nelsonii* Rand. Three of the species (*Iris brevicaulis*, *I. fulva,* and *I. hexagona*) are morphologically distinct, interfertile, and widespread (Arnold 2006) (Figure 4.1). Hybrid zones between these three species occur where their ranges overlap in southern Louisiana. Repeated observations have noted asymmetric introgression of *I. fulva* alleles into both *I. brevicaulis* and *I. hexagona* (Arnold et al. 2010). The fourth species, *I. nelsonii*, is a local endemic existing in a single parish in southern Louisiana and is likely ecologically isolated from its three congeners (Taylor et al. 2011; Taylor et al. 2012). Traits of interest which vary within the species complex include micro-habitat differences (e.g. flood tolerant vs. dry-adapted), floral traits associated with pollinator syndromes (e.g. bumble vs. hummingbird), and flowering phenology (overlapping vs. non-overlapping) (Arnold 2006).

Our aim was to test the role of niche divergence within the Louisiana iris species complex. To do so, we first clarified the phylogenetic relationships and timing of divergence for the Louisiana irises. To evaluate the role of ecological differentiation in the divergence of three distinct species of irises, we modeled the current distribution of these three species and tested if niche divergence accompanies lineage diversification.

Materials and Methods

Species phylogeny and divergence time estimates

DNA was extracted from three individuals for each of the three Louisiana iris species found from populations throughout their range (Appendix C, Table S1). Extractions were performed using the Qiagen DNeasy plant kit (Qiagen, Valencia, CA). Two chloroplast gene regions were sequenced: *matK* with flanking *trnK* intron and *ndhF* in order to reconstruct the phylogenetic relationships for the three iris species. The protein-coding *matK* gene and flanking *trnK* region (~ 1700 bp) was amplified in two reactions using primers 3914F/1235R and 1176F/trnk2R (Wilson 2004) and following the protocols from Wilson (2003). The *ndhF* gene (~ 2200 bp) was generated via two PCR reactions with the primer sets from Wilson (2011) and following the associated PCR protocol. PCR products were sent to Macrogen USA (https://www.macrogenusa.com/) for purification and bidirectional sequencing. Sequence chromatograms of forward and reverse sequencing reactions obtained from Macrogen USA were edited using Geneious, version 7.1.7 (Kearse et al. 2012). Raw chloroplast sequences for *I. koreana* (the outgroup) were downloaded from GenBank and all associated sequences were aligned in Geneious version 7.1.7.

For the Louisiana iris species clade, we simultaneously estimated the phylogenetic relationships and the absolute divergence times among species in a Bayesian framework using *BEAST Version 1.7.5 (Drummond et al. 2012). Here, we use *BEAST as we treat the two different chloroplast genes as separate loci. Fossils are not known, therefore, to estimate absolute divergence times we used an average substitution rate for chloroplast genes (Wolfe et al. 1987). The substitution rate was set to a normally distributed prior of 2.0 x 10^{-9} substitutions/site/year and the trees were linked. We used the HKY substitution model, which was selected as the best fit to the data with PartitionFinder (Lanfear et al. 2012). We fit a strict clock, which has been shown to be superior for phylogenies with shallow roots because of low levels of rate variation between branches (Brown and Yang 2011). The prior model was set to a Yule process of speciation. *BEAST was run for 10 million generations sampling every 1000 steps and discarding the first 2,500 as burnin. Posterior samples of parameter values were summarized and assessed for convergence and mixing using Tracer v. 1.5 making sure that effective sample sizes of parameter estimates were greater than 250. Trees were generated using TreeAnnotator v1.5.3 (Rambaut and Drummond 2007).

Presence Points

Species occurrence records were obtained from recently collected field specimens of all three Louisiana irises (Hamlin and Arnold 2014; Hamlin and Arnold *submitted*). These records are complemented with herbarium specimens dating back to 1980, where we could include records that could unambiguously be assigned to a precise location via GPS coordinates. All presence points were reported to at least two decimal places (Appendix C, Table S2). Our final data set of unique georeferenced records included 75 data points for *I. brevicaulis,* 68 data points for *I. fulva*, and 76 data points for *I. hexagona*.

Environmental Variables

We compiled a set of climate, radar scatterometer, and satellite remote sensing variables to characterize present day conditions for the Louisiana irises. These included the 19 bioclimatic variables from the WorldClim database (http://www.worldclim.org/bioclim), which are derived from 50 years (1950 – 2000) of climate data. This database consists of estimates of annual means, seasonal extremes, and degrees of seasonality in both temperature and precipitation (Wellenreuther et al. 2012). An additional four land cover layers were also used. These land cover layers included the normalize difference vegetation index (NDVI; measure of vegetative greenness), the yearly standard deviation of NDVI (NDVISTD), and the percentage of tree cover (TREE) from satellite remote sensors (NASA-MODIS/Terra data set). The radar scatterometer variable QSCAT provides a measure of vegetation canopy structure and moisture and, in areas of sparse vegetation, soil roughness and wetness. All environmental variables used have a spatial resolution of 30 arc-seconds (1 km^2).

The selection of environmental layers is a critical step in preparation for ecological niche modeling because many modeling techniques, which incorporate too many variables, can cause over-fitting of the model (Peterson et al. 2007). To prevent over fitting, we performed a correlation analysis for each species to remove variables that were highly correlated ($R > 0.90$). The variables that were not significantly correlated were used for model generation and evaluation for each species (Table 4.1). These variables represent species level ecological differences, which may significantly impact the geographic distribution of the species within this complex.

Ecological Niche Models

We used the Maxent 3.3.k (Phillips et al. 2006; Phillips and Dudík 2008) algorithm as implemented with the *dismo* package in R (R Core Team 2013) to compute the ecological niche model for each species. As several methods are available for constructing ENMs, we chose the Maxent model because it performs well even when the number of occurrence records for modeling is small (Wisz et al. 2008) and has demonstrated accuracy in estimating ecological niche models (Elith et al. 2011). Maxent allows for model performance evaluation by calculating the area under the ROC (receiver operating characteristic) curve (AUC), based on a calibration and evaluation dataset (Phillips et al. 2006). The AUC ranges from 0 to 1; where an AUC value of 0.5 represents a model that performs no better than random. A value in the range of 0.75 – 0.9 are considered good and models with an AUC > 0.9 are considered excellent (Elith 2000).

To correct for sampling bias in the dataset during the modeling process, niche models were restricted to one occurrence record within a $1km^2$ grid over the observation area. Specifically, presence and background points (or pseudoabsence points) were randomly assigned into five groups using the *kfold* function available in the dismo package. Four groups of presence and pseudoabscence points were designated as the calibration group, and were used to generate the models. The fifth group of presence and pseudoabsence points were used as data to evaluate the predictive ability of the models generated. This resulted in 80% of the presence points being used for model creation and 20% of the presence points being used for model evaluation. The process of randomly dividing points into calibration and evaluation groups, creating an ENM, and evaluating the ENM's predictive power was repeated ten times for each species, and an average AUC score and standard deviation was calculated for each species.

Niche comparisons

We tested for ecological niche divergence of all three members of the Louisiana irises species complex using several methods. First, we examined the level of divergence in environmental principal component space. We conducted a Principal Component Analysis (PCA) on all species within the complex to assess the overall level of ecological divergence and to quantify if and how each species overlapped in principal component space. Data for all occurrences localities was the same as those used in building ENMs. We determined the uncorrelated variables for all species (Appendix C, Table S3), performed a principal component analysis, and identified the variables with the highest loadings on the first four principal components. Statistical significance of possible niche differences between species was evaluated using multivariate analyses of variance (MANOVA) of the first three principal components. Statistical analyses were conducted using R (R Core Team 2013).

As a complement to the PCA analysis, niche overlap between species was estimated using two alternative measures of similarity: Schoener's *D* and Hellinger's-based *I* (Warren et al. 2008). Values of these measures range from 0 to 1, with values close to 1 indicating high similarity between niches and values close to 0 indicate niche dissimilarity. The significance of both the similarity indices (*D* and *I*) was evaluated with two tests of niche evolution. First, we used the identity test to determine whether the ENMs produced by two species are equivalent to one another (Warren et al. 2008). The null hypothesis (niches are equivalent) is rejected when the measured overlap is significantly lower than the null distribution. However, niche divergence could result either from actual niche differences or simply through spatial autocorrelation in environmental variables (McCormack 2009). To overcome the potential confounding effect, a background similarity test was performed. Essentially, this test determines whether niche divergence cannot be explained by regional differences in the habitat available to each species (Blair et al. 2013). This test compares the niche of a focal species with a set of pseudoniches modeled from a random sampling of the geographic range of the other species (Warren et al. 2008). The hypothesis of niche divergence is rejected when the empirically observed value for D and/or I is greater than values observed for the null distribution. Tests of niche evolution were performed using the *phyloclim* package in R with 100 permutations (Heibl and Calenge 2011).

Results

Species phylogeny and divergence time estimates

The chloroplast phylogeny recovered a highly supported clade for the Louisiana irises with a posterior probability of > 95% for all nodes. *Iris brevicaulis* and *I. fulva* are considered to be sister to each other (Figure 4.2) and share a most recent common ancestor with *I. hexagona.* Divergence time estimates between the Louisiana iris species group and the outgroup *I. koreana* is 6.2 mya (HPD 4.2 – 8.1 mya). While divergence within the Louisiana iris species group began around 0.53 mya (HPD 0.19 - 1.02 mya) with *I. hexagona* branching off first. The split between *I. brevicaulis* and *I. fulva* occurred 0.20 mya (HPD 0.027 – 0.50), indicating that the irises are of relatively recent origin.

However, because this inference was derived solely from chloroplast sequence data some caution in the interpretation of divergence time estimates is warranted.

Ecological Niche Models

After sub-setting the presence points, 61 points for *I. brevicaulis*, 48 points for *I. fulva*, and 64 points for *I. hexagona* were available for model creation. Ecological Niche models for *I. brevicaulis* had an average AUC score of 0.744 and a standard deviation of 0.086. For *I. fulva*, the AUC score average was 0.937 (S.D. = 0.021). For *I. hexagona*, the AUC score average was 0.85 (S.D. = 0.046). Values of AUC for all species indicate reliable model performance (Phillips et al. 2006). The variation in the predictive ability of the models for the iris species is likely a result of the disparity in the ranges of these species. The least accurate model is for *I. brevicaulis,* which has a range that is far larger than the other iris species (Figure 4.3a), so the model must account for greater variation in environmental variables between presence points. The most accurate model was for *I. fulva*, despite this species having the lowest number of points. This effect could be attributed to the extremely restrictive range of the species (Figure 4.3b), which results in presence points that have very similar values for each of the environmental variables. Finally, the predicted niche model for *I. hexagona* is in good agreement with the current known geographic distribution of this species as well (Figure 4.3c).

Niche comparisons

The PCA conducted using all species explained 72.55% of the variation. Principal component 1 explained 31.43% of the variation and mean temperature of warmest quarter had the highest loading. Principal component 2 explained 17.48% of the variation and precipitation seasonality had the highest loading. Principal component 3 explained 13.06% of the variation and NDVI (i.e. measure of vegetative greenness) had the highest loading. Finally, principal component 4 explained 10.57% of the variation and mean temperature diurnal range variable had the highest loading. All species overlap to some degree in principal component space (Figure 4.4). However, the *I. hexagona* distribution was distinct from those of *I. fulva* and *I. brevicaulis*. Furthermore, *I. brevicaulis* showed two separate distributions, with the southern *I. brevicaulis* populations clustering with the *I. fulva* and some *I. hexagona* points, and the northern *I. brevicaulis* populations clustering distinctly from all other iris populations (Figure 4.4). Moreover, there are species level differences in environmental space based on the principal components (MANOVA: $p = 3.736e-06$).

For niche identity, all species comparisons indicated that the different taxa were not ecologically equivalent, regardless of the measure of similarity used (either Schoener's *D* or *I*) (Appendix C, Figure S1). These differences could reflect either ecological divergence between these species or different environmental backgrounds as a result of their partial allopatric distributions (Figure 4.1). Results of the background test were more complex but also support niche divergence for a number of comparisons. For the similarity metric Schoener's *D*, niche divergence is supported when we compare species 1 on the background of species 2 (Table 4.2a). For two comparisons, of the opposite direction (species 2 versus species 1), the background test indicated that the species were more similar (niche conservatism) than expected. Specifically, for either *I. fulva or I. hexagona* compared to the background of the widely distributed *I. brevicaulis*

(Table 4.2a). For example, background tests exhibited opposite patterns for *I. brevicaulis* and *I. fulva* (Table 4.2a). However, this is not surprising as *I. fulva* is more narrowly distributed and thus when compared to the local environmental background of *I. brevicaulis*, evidence of niche conservatism is expected. For the similarity metric *I*, we see the same patterns as Schoener's *D*, where niche divergence is implicated for species 1 on the background of species 2 for all three comparisons (Table 4.2b). As *I. brevicaulis* and *I. fulva* are considered to be sister to each other, evidence of niche divergence between these two species is implicated. Furthermore, tests of niche divergence for either sister species (*I. brevicaulis* or *I. fulva*) indicate niche divergence when compared to *I. hexagona*.

Discussion

According to our results, the origin of the Louisiana iris species clade is dated to the Middle – Late Pleistocene (0.53 mya; HPD 0.19 - 1.02 mya). There is evidence of variation in the predictability for our ENMs and the Louisiana irises. However, our ENMs do predict, with confidence, the potential distributions of the Louisiana irises suggesting that these species' geographic distributions are substantially influenced by the environmental conditions used in this study. Here, we show that precipitation and temperature are more important in explaining ecological differences for the irises than the vegetation-based variables (Figure 4.4). A significant result of our study is that multiple analyses show strong evidence for niche divergence for the Louisiana irises.

Divergence time estimates

This is the first estimate of phylogenetic relationships using the three widely distributed irises within the Louisiana iris species complex. Previous work has only included *I. fulva* and *I. brevicaulis* in which they were also classified as being sister to each other (Wilson 2004; Wilson 2011). By elucidating the relationships for the Louisiana irises, we were able to ask if niche divergence was seen for both sister species when compared to the more distantly related *I. hexagona*. Furthermore, the estimated origin of the Louisiana irises (Middle – Late Pleistocene) is roughly comparable to other plant species found in this region (Morris et al. 2008; Ellison et al. 2012).

Ecological differences for the Louisiana irises

Based on collection information, the observed patterns of the geographic distribution indicate that two of the Louisiana iris species have unique ranges that include a large allopatric distribution (*I. brevicaulis* and *I. hexagona*). The predictive accuracy of the ENMs varied, due to the differences in the distributions and the relative importance of the uncorrelated environmental variable, for each species. The small number of presence points for *I. brevicaulis*, compared to the breadth of its range and environmental variation, could explain the poor model performance for this species (Hernandez et al. 2006). In contrast, model performance for both *I. fulva* and *I. hexagona* included strong predictive power relative to previous descriptions of the species' ranges.

Because our analyses included a subset of environmental variables, better estimates of ecological setting might be resolved from the addition of additional niche dimensions. For example, the bioclimatic variables chosen for this study are thought to describe abiotic conditions rather than biotic interactions (Thuiller et al. 2005; Kissling et al. 2012). However, the data from land cover layers should have provided higher resolution of the ecological resources and thus potentially reflect important environmental variables (McCormack et al. 2010). Prior knowledge of the species' habitat preferences indicates that the use of an environmental layer depicting fine scale differences in soil measures could improve the predictive quality of our models (Cruzan and Arnold 1993). Furthermore, *I. hexagona* has been shown to be sensitive to various levels of saline tolerance (Van Zandt and Mopper 2004) and thus including a measure of salinity change over the coastal landscape could also provide fine scale resolution for population ecological differences within this species. One method which may over come the difficulties associated with elucidating biotic interactions in terms of species distributions is a newly described method: joint species distribution models (JSDM) (Pollock et al. 2014). Here, JSDM will simultaneously estimate distributions for a number of species and describe shared environmental responses and residual patterns of co-occurrence. This may allow one to better understand biotic interactions here in terms of the distribution of pollinators and associated distribution of the Louisiana irises.

The strong differences observed between the iris species in regard to associated environmental factors suggest an effect from these components on the differences in distributions and habitats occupied. The environmental variables that contributed most to the distributions of all three species suggest a sensitivity to changes in the seasonality of precipitation and the mean temperature of the warmest quarter. This could have implications for the maintenance of populations during fluctuating climatic conditions. Furthermore, the environmental variables used in this study have successfully generated accurate ecological niche models across a wide range of taxa found in the Southeast United States (Kalkvik et al. 2012; Wooten and Gibbs 2012; Zellmer et al. 2012). This observation suggests that these variables are able to adequately describe the environmental conditions that are important for species found in this region. However, one of the challenges of ENMs is that we cannot evaluate easily how appropriate the data are for a particular species.

Tests of niche divergence

One of the strengths of the methodologies applied in the current study is the accounting for spatial autocorrelation among environmental variables (i.e. by the similarity tests). This is important, as abiotic environmental variables are highly correlated with latitude and longitude (Wooten 2012, McCormack 2010). From these tests, niche divergence is implicated as being important for all sets of comparisons between closely (*I. brevicaulis* and *I. fulva*) and more distantly (either *I. brevicaulis* or *I. fulva* with *I. hexagona*) related species in this complex; this inference is independent of the extent to which geographic ranges overlap. Evidence of niche divergence between *I. hexagona* and the two sister species, indicates that niches remain differentiated after speciation and that geographic isolation has thus been of less importance in the evolution of species differences. This result is similar to that found for the geographically widespread and ecologically distinct tomato clade (Nakazato et al. 2010). Likewise, studies of taxa as varied as dendrobatid frogs and in the genus *Pelargonium* have shown that the niche is not necessarily conserved, especially beyond sister species (Graham et al. 2004; Martínez-Cabrera et al. 2012).

Niche divergence is less apparent when *I. fulva* is compared to the background of *I. brevicaulis*, which is not surprising as the entire range of *I. fulva* is nested within the range of *I. brevicaulis*. This suggests that ecological differentiation between these two species, as reflected by change in temperature, precipitation, and the vegetative measure of change of greenness (NDVI), may play an important role in structuring the distribution and the evolution of their niches. Furthermore, combining species delimitation and ecological niche modeling within *I. brevicaulis*, suggests that the populations of *I. brevicaulis* are on different evolutionary trajectories and could be considered incipient and cryptic species (Rissler and Apodaca 2007).

Quantitative Trait Locus analysis (QTL) within the Louisiana irises has detected loci associated with ecological divergence and thus reproductive isolation between *I. fulva* and *I. brevicaulis* (Martin et al. 2005, 2006; Martin et al. 2008; Arnold et al. 2012). For example, Martin et al. (2005) defined QTL underlying survivorship of parental and hybrid (F_1 and first-generation backcross) genotypes in a greenhouse setting. Thus, conncet loci associated with post-zygotic reproductive isolating mechanisms (i.e. mortality and survivorship) with those associated with divergent selection.

Furthermore, QTL mapping in the same reciprocal backcross mapping populations defined the genetic architecture of loci that affected prezygotic reproductive isolation as mediated by pollinator behavior. Specifically, Martin et al. (2008) were able to 1) estimate the directionality of pollinator mediated selection on *I. brevicaulis* and *I. fulva* floral traits, 2) determine the potential for pollinator preference to act as a prezygotic reproductive barrier, and 3) identify QTL(s) that contribute to differential pollinator preferences. The results led to the inference that pollinator-mediated selection on floral traits might impact the Louisiana irises as different pollinator classes demonstrated alternate preferences (Martin et al. 2008). They concluded that if either pollinator predominated, selection would favor the respective floral traits associated with that pollination syndrome leading to reproductive isolation (Martin et al. 2008).

We conclude that for the Louisiana irises, niche divergence and thus ecological speciation within this clade could be the major component of diversification. Future research will be designed to test for the action of pollinators and/or climatic factors that may have contributed to the initiation of diversification within this iris species complex.

Acknowledgements

JAPH would like to acknowledge the Wares and Sweigart labs at UGA for fruitful discussions regarding this work. JAPH would also like to acknowledge L. Stanton for discussion about ecological niche modeling. JAPH and the research reported in this publication were partially supported by the National Institute of Health under award number T32GM007103. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Current National Science Foundation funding for this work comes from DEB-0949479/0949424 (collaborative grant with NH Marti Martin of Texas State University) and DEB-1049757.

Literature Cited

Alvarado ‐ Serrano, D. F. and L. L. Knowles. 2014. Ecological niche models in phylogeographic studies: applications, advances and precautions. Molecular ecology resources 14:233-248.

Arnold, M. L. 2006. Evolution through genetic exchange. Oxford University Press, New hyork.

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. Heredity 108:159-166.

Arnold, M. L., S. Tang, S. J. Knapp, and N. H. Martin. 2010. Asymmetric introgressive hybridization among Louisiana Iris species. Genes 1:9-22.

Blair, M., E. Sterling, M. Dusch, C. Raxworthy, and R. Pearson. 2013. Ecological divergence and speciation between lemur (Eulemur) sister species in Madagascar. Journal of evolutionary biology 26:1790-1801.

Brown, R. P. and Z. Yang. 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. BMC evolutionary biology 11:271.

Cruzan, M. B. and M. L. Arnold. 1993. Ecological and genetic associations in an Iris hybrid zone. Evolution:1432-1445.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection or the Preservaton of Favoured Races in the Struggle for Life. John Murray, London.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution 29:1969-1973.

Elith, J. 2000. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. Pp. 39-58. Quantitative methods for conservation biology. Springer.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.

Ellison, A. M., E. D. Butler, E. J. Hicks, R. F. Naczi, P. J. Calie, C. D. Bell, and C. C. Davis. 2012. Phylogeny and biogeography of the carnivorous plant family Sarraceniaceae. PloS one 7:e39291.

Gallien, L., T. Münkemüller, C. H. Albert, I. Boulangeat, and W. Thuiller. 2010. Predicting potential distributions of invasive species: where to go from here? Diversity and Distributions 16:331-342.

Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution 58:1781-1793.

Hamlin, J. A. and M. L. Arnold. 2014. Determining population structure and hybridization for two iris species. Ecology and evolution 4:743-755.

Hamlin, J. A. and M. L. Arnold. *submitted*. Neutral and selective processes drive population differentiation for *Iris hexagona*

Harvey, P. H. and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Heibl, C. and C. Calenge. 2011. Phyloclim: integrating phylogenetics and climatic niche modelling. R package accessed: September-20–2010.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773-785.

Hickerson, M., B. Carstens, J. Cavender-Bares, K. Crandall, C. Graham, J. Johnson, L. Rissler, P. Victoriano, and A. Yoder. 2010. Phylogeography's past, present, and future: 10 years after. Molecular Phylogenetics and Evolution 54:291-301.

Ibarra-Cerdeña, C. N., A. Zaldívar-Riverón, A. T. Peterson, V. Sánchez-Cordero, and J. M. Ramsey. 2014. Phylogeny and Niche Conservatism in North and Central American Triatomine Bugs (Hemiptera: Reduviidae: Triatominae), Vectors of Chagas' Disease. PLoS neglected tropical diseases 8:e3266.

Jakob, S. S., C. Heibl, D. Roedder, and F. R. Blattner. 2010. Population demography influences climatic niche evolution: evidence from diploid American Hordeum species (Poaceae). Molecular Ecology 19:1423-1438.

Kalkvik, H. M., I. J. Stout, T. J. Doonan, and C. L. Parkinson. 2012. Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the Peromyscus maniculatus species group. Ecography 35:54-64.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, and C. Duran. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Kissling, W. D., C. F. Dormann, J. Groeneveld, T. Hickler, I. Kühn, G. J. McInerny, J. M. Montoya, C. Römermann, K. Schiffers, and F. M. Schurr. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. Journal of Biogeography 39:2163-2178.

Kozak, K. H. and J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60:2604-2621.

Lanfear, R., B. Calcott, S. Y. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution 29:1695-1701.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2005. Loci affecting long-term hybrid survivorship in Louisiana irises: Implications for reproductive isolation and introgression. Evolution 59:2116-2124.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2006. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics 172:2481-2489.

Martin, N. H., Y. Sapir, and M. L. Arnold. 2008. The genetic architecture of reproductive isolation in Louisiana irises: pollination syndromes and pollinator preferences. Evolution 62:740-752.

Martínez-Cabrera, H. I., C. D. Schlichting, J. A. Silander, and C. S. Jones. 2012. Low levels of climate niche conservatism may explain clade diversity patterns in the South African genus Pelargonium (Geraniaceae). American journal of botany 99:954-960.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

McCormack, J. E., A. J. Zellmer, and L. L. Knowles. 2010. Does niche divergence accompany allopatric divergence in Aphelocoma jays as predicted under ecological speciation?: insights from tests with niche models. Evolution 64:1231-1244.

Morris, A. B., S. M. Ickert-Bond, D. B. Brunson, D. E. Soltis, and P. S. Soltis. 2008. Phylogeographical structure and temporal complexity in American sweetgum (Liquidambar styraciflua; Altingiaceae). Molecular Ecology 17:3889-3900.

Nakazato, T., D. L. Warren, and L. C. Moyle. 2010. Ecological and geographic modes of species divergence in wild tomatoes. American Journal of Botany 97:680-693.

Nosil, P. 2012. Ecological speciation. Oxford University Press.

Pearson, R. G. and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global ecology and biogeography 12:361-371.

Peterson, A. T., M. Papeş, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30:550-560.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190:231-259.

Phillips, S. J. and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Pp. 83. Proceedings of the twenty-first international conference on Machine learning. ACM.

Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesk, and M. A. McCarthy. 2014. Understanding co‐occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). Methods in Ecology and Evolution 5:397-406.

R Core Team. 2013. R: A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.

Rambaut, A. and A. Drummond. 2007. Tracer v1. 4.

Riley, H. P. 1938. A character analysis of colonies of *Iris fulva*, *Iris hexagona* var giganticaerulea and natural hybrids. American Journal of Botany 25:727-738.

Rissler, L. J. and J. J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (Aneides flavipunctatus). Systematic Biology 56:924-942.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Schluter, D. 2009. Evidence for Ecological Speciation and Its Alternative. Science 323:737-741.

Taylor, S. J., K. J. AuBuchon, and N. H. Martin. 2012. Identification of Floral Visitors of Iris nelsonii. Southeastern Naturalist 11:141-144.

Taylor, S. J., R. W. Willard, J. P. Shaw, M. C. Dobson, and N. H. Martin. 2011. Differential response of the homoploid hybrid species Iris nelsonii (Iridaceae) and its progenitors to abiotic habitat conditions. American Journal of Botany 98:1309-1316.

Thuiller, W., S. Lavorel, M. B. Araújo, M. T. Sykes, and I. C. Prentice. 2005. Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences of the united States of America 102:8245-8250.

Van Zandt, P. A. and S. Mopper. 2004. The effects of maternal salinity and seed environment on germination and growth in Iris hexagona. Evolutionary Ecology Research 6:813-832.

Viosca, P. J. 1935. The irses of southeastern Louisiana - a taxonomic and ecological interpretation. Bulletin of the American Iris Society 57:3 - 56.

Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607-611.

Warren, D. L., R. E. Glor, M. Turelli, and D. Funk. 2008. Environmental Niche Equivalency versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution 62:2868-2883.

Wellenreuther, M., K. W. Larson, and E. I. Svensson. 2012. Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. Ecology 93:1353-1366.

Wiens, J. J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution 58:193-197.

Wiens, J. J. and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annual review of ecology, evolution, and systematics:519-539.

Wilson, C. A. 2003. Phylogenetic relationships in Iris series Californicae based on ITS sequences of nuclear ribosomal DNA. Systematic Botany 28:39-46.

Wilson, C. A. 2004. Phylogeny of Iris based on chloroplast matK gene and trnK intron sequence data. Molecular Phylogenetics and Evolution 33:402-412.

Wilson, C. A. 2011. Subgeneric classification in Iris re-examined using chloroplast sequence data. Taxon 60:27-35.

Wilson, K. A., M. I. Westphal, H. P. Possingham, and J. Elith. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122:99-112.

Wisz, M. S., R. Hijmans, J. Li, A. T. Peterson, C. Graham, and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14:763-773.

Wolfe, K. H., W.-H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences 84:9054-9058.

Wooten, J. and H. Gibbs. 2012. Niche divergence and lineage diversification among closely related Sistrurus rattlesnakes. Journal of evolutionary biology 25:317-328.

Zellmer, A. J., M. M. Hanes, S. M. Hird, and B. C. Carstens. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant Sarracenia alata. Systematic Biology 61:763-777.

Table 4.1. Uncorrelated environmental variables used to model ecological niches of the three Louisiana irises.

Species	Uncorrelated Environmental variables[*]
I. brevicaulis	Bio2, Bio5 Bio8, Bio9, Bio10, Bio12, Bio15, NDVI, NDVISTD, QSCAT, TREE
I. fulva	Bio2, Bio5 Bio8, Bio9, Bio14, Bio15, Bio18, NDVI, NDVISTD, QSCAT, TREE
I. hexagona	Bio2, Bio5 Bio8, Bio9, NDVI, NDVISTD, QSCAT, TREE

[*] Bio2 = Mean diurnal range, Bio5 = Max temperature of warmest month, Bio8 = Mean temperature of wettest quarter, Bio9 = Mean temperature of driest quarter, Bio10 = Mean temperature of warmest quarter, Bio12 = Annual precipitation, Bio14 = Precipitation of driest month, Bio15 = Precipitation seasonality, Bio18 = Precipitation of warmest quarter, NDVI = Normalized Difference Vegetation Index (greenness), NDVISTD = Greenness seasonality, TREE = Percent tree cover, QSCAT = Canopy or surface moisture and roughness

Table 4.2a Results of background similarity tests based on Schoener's *D*.

Species pair sp1 vs sp2	**Schoener's *D***	**Schoener's *D* 95% CI sp1 vs sp2**	**Schoener's *D* 95% CI sp2 vs sp1**	**D or C sp1 vs sp2**	**D or C sp2 vs sp1**
I. brevicaulis – I. fulva	0.368	0.535 – 0.629	0.169 – 0.238	D	C
I. brevicaulis – I. hexagona	0.201	0.567 – 0.639	0.136 – 0.210	D	C
Iris fulva – I. hexagona	0.150	0.154 – 0.212	0.159 – 0.214	D	D

Table 4.2b. Results of background similarity tests baed on Hellinger's *I*.

Species pair sp1 vs sp2	**Hellinger's *I***	**Hellingner's *I* 95% CI sp1 vs sp2**	**Hellingner's *I* 95% CI sp2 vs sp1**	**D or C sp1 vs sp2**	**D or C sp2 vs sp1**
I. brevicaulis – I. fulva	0.640	0.810 – 0.872	0.394 – 0.490	D	C
I. brevicaulis – I. hexagona	0.453	0.830 – 0.879	0.363 – 0.472	D	C
Iris fulva – I. hexagona	0.321	0.365 – 0.453	0.407 – 0.496	D	D

D or C either indicate niche divergence (D) or niche conservatism (C), with that value being significant.

Figure 4.1

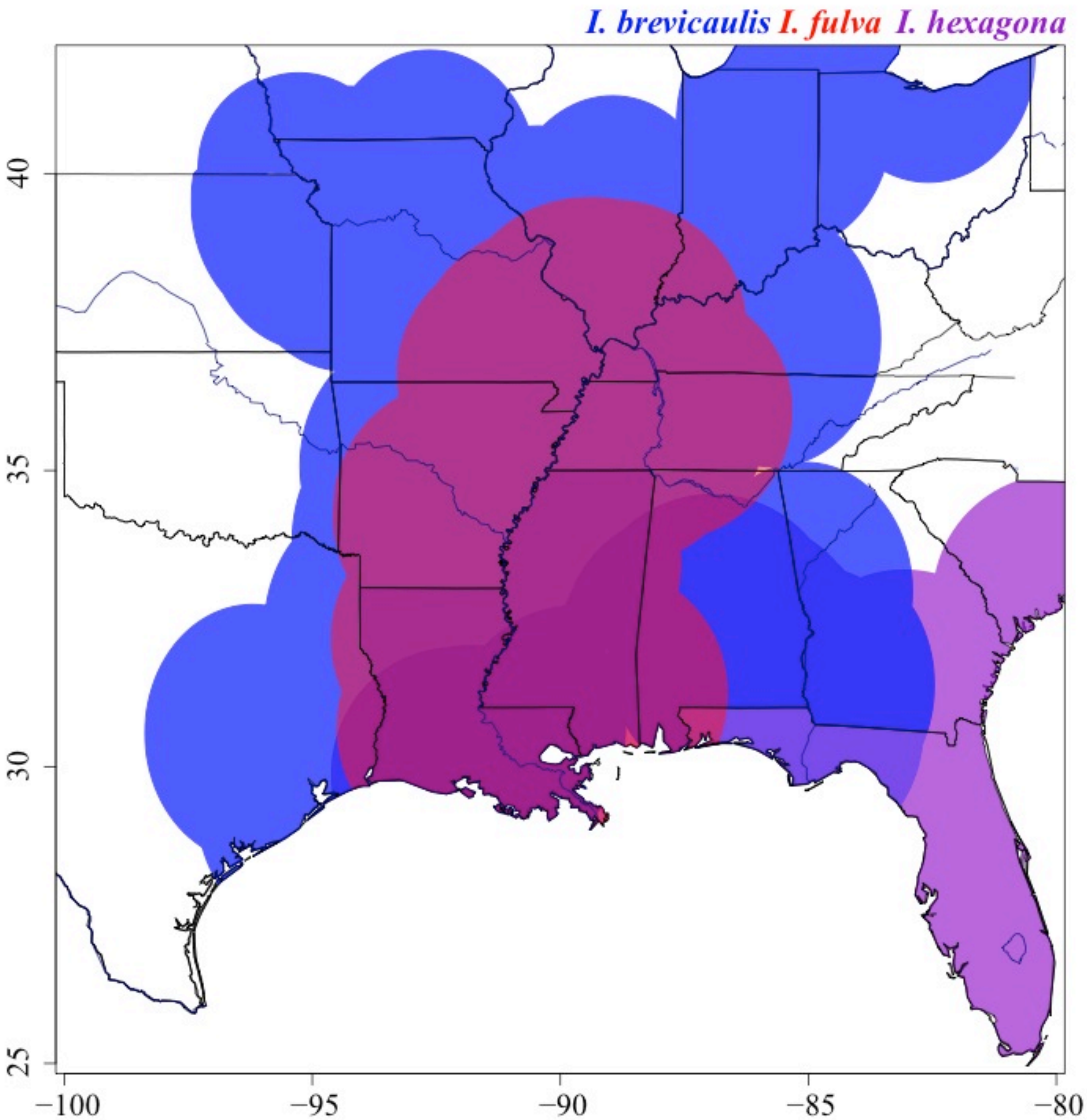


Figure 4.1. Range distribution map for the Louisiana irises. Major rivers for the region are outlined. Species names colored as the same color of their range distribution. *Iris brevicaulis* (blue) has the widest distribution, which is the entire southeast United States. *Iris fulva* (red) is narrowly distributed along the Mississippi River Valley. *Iris hexagona* (purple) is coastally distributed.

Figure 4.2

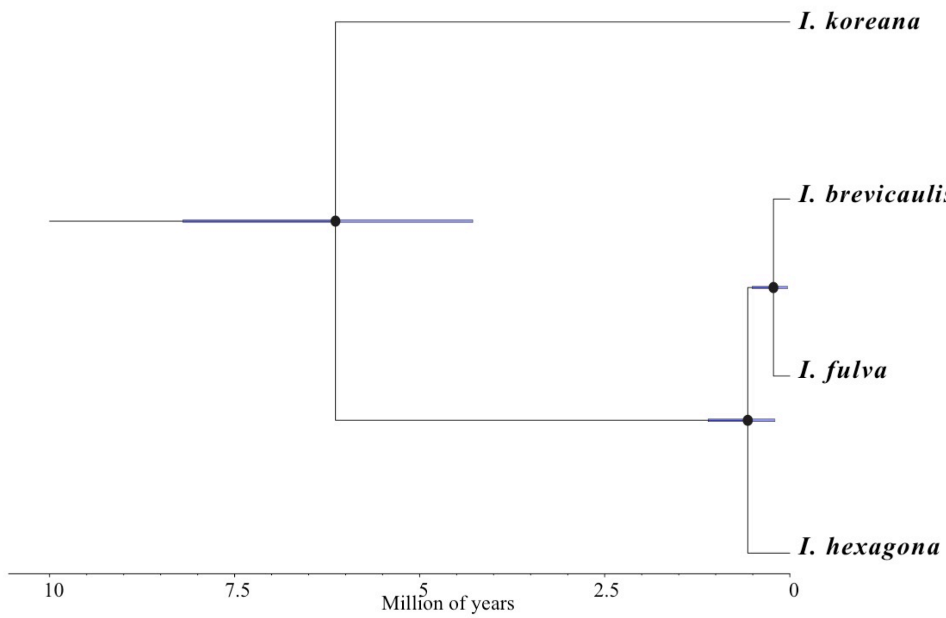


Figure 4.2. Species tree based on two chloroplast markers, which shows the phylogenetic relationships for the Louisiana irises and the estimated divergence times for all three species from *BEAST. Black dots are > 95% posterior probability. *Iris koreana* is the outgroup.

Figure 4.3

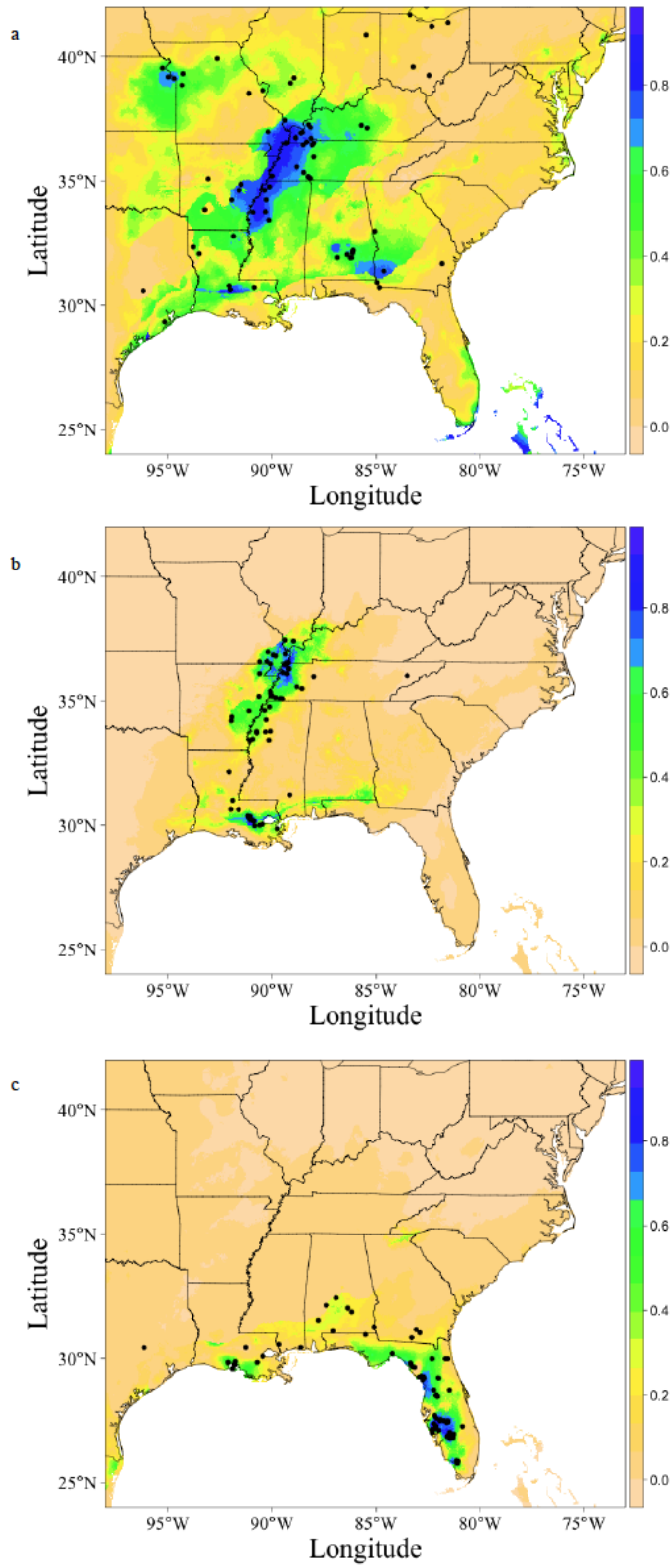

a
b
c
40°N
35°N
30°N
25°N
Latitude
95°W
90°W
85°W
80°W
75°W
Longitude
0.8
0.6
0.4
0.2
0.0

Figure 4.3. Ecological niche models (ENM) for all three Louisiana iris species based on Maxent. Individual dots on each map are the sample locations used to construct the ENMs for each species. Colors of blue represent likely locations based on the projection from the ENM for each species. a) *Iris brevicaulis*; b) *Iris fulva*; and c) *Iris hexagona*.

Figure 4.4

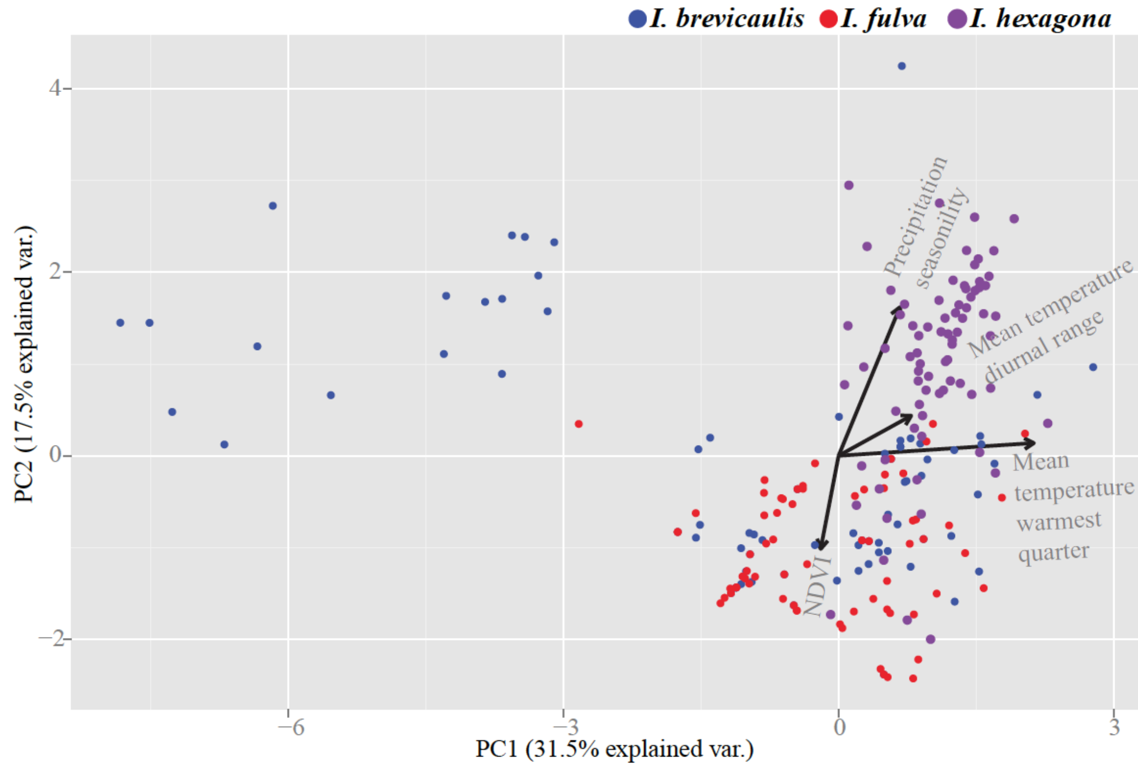


Figure 4.4. Principal component analysis of uncorrelated environmental variables for the Louisiana irises. Species-level differences in environmental preferences are evident. Individual dots are the associated localities, which are color-coded dependent on the species. Environmental variables that contributed to most to the first four principal components are shown on the plot.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

A goal of evolutionary biology is to identify how biodiversity arises and is maintained. Historically, the predominant mode of speciation emphasized geographical isolation and often assumed ecological factors as insignificant, whereas theories now emphasize the importance of ecological differences driving speciation (Nosil 2012). Work on evolutionarily young lineages now recognizes that barriers to gene flow can evolve as a result of ecologically based divergent or disruptive selection (Wolf et al. 2010). This is bolstered with theoretical work evaluating the role of natural selection in a number of empirical systems (Gavrilets and Losos 2009; Barton 2010). However, ecological speciation can occur under any geographic context, as geography influences the ecological sources of divergent selection (Rundle and Nosil 2005). Thus, understanding the extent that environmental differences and geographic distance can limit gene flow sheds light on the relative roles of selection and dispersal limitations (Papadopulos et al. 2014). And work on young incipient species allows one to disentangle the influence of gene flow, ecological differentiation, and local adaptation during the early stages of speciation (Via 2009).

The Louisiana iris species complex is a good system to understand the influence of these factors. Three of the species within the complex are found throughout the southeastern United States with range distributions that overlap predominantly in

Louisiana. Bumblebees or hummingbirds, dependent on the species pollination syndrome, predominantly pollinate the Louisiana irises. Thus, gene flow and local adaptation is likely to be influenced by the dispersal limitations of the associated pollinator as hummingbirds migratory flights occur over a much larger distance (dePamphilis and Wyatt 1989) than bumblebee forging distances (Steffan-Dewenter and Tscharntke 1999). Previous work has also demonstrated that hybrid zones exist when species are sympatric, found often in Louisiana and that hybrid fitness varies among genotypes and across habitats (Arnold et al. 2012). Furthermore, a number of studies have evaluated ecological setting and genomic architecture in order to uncover adaptive trait introgression (Martin et al. 2005, 2006; Martin et al. 2008). Here, I assess the influence of gene flow, environmental differences, and local adaptation to understand if divergence is either stimulated or hindered within this species complex and to what degree across the geographic range.

In this dissertation I set out to examine the role gene flow and selection played in the evolution of the Louisiana iris species complex, both spatially and temporally. Chapter two examined the geographic distribution of genetic diversity in two species of Louisiana iris, *Iris brevicaulis* and *Iris fulva*. Using a genotyping-by-sequencing approach, I sampled a large number of single nucleotide polymorphisms (SNPs) and determined species-level differences in connectivity among populations. In particular, I found strong genetic structure for *I. brevicaulis* populations and thus those populations appear much more isolated from intraspecific gene flow. In contrast, *I. fulva* appears to be acting as a single panmictic population regardless of distance between populations. Thus, *I. brevicaulis* populations could be on a different evolutionary trajectory than *I.*

fulva. Additionally, the Mississippi River does not appear to be acting as a geographic barrier to gene flow for either species. However, the extent of reduced gene flow among allopatric populations of *I. brevicaulis* is apparent from North to South, but at this time we cannot disentangle the migratory capacity of their pollinator (bumblebees) versus other demographic factors acting as a barrier. Furthermore, I detected asymmetric introgressive hybridization of alleles largely from *I. fulva* into *I. brevicaulis*. The pattern of asymmetric introgressive hybridization has been documented previously in crossing and field experiments (Cruzan and Arnold 1994; Martin et al. 2005, 2006; Tang et al. 2010), thus supporting the movement of alleles for these species across a larger geographic scale. Taken together, the results suggest differences in the magnitude and pattern of connectivity among populations and species and over time these species may exhibit different responses to ecological selection.

In chapter three, I found evidence of local adaptation for floral morphology at the longitudinal extremes for *I. hexagona*. First, I demonstrated floral morphology and climate differences are evident for *I. hexagona* populations. In particular, Louisiana populations are distinctly different in terms of morphology from other Florida populations. Populations, also, varied in terms of climate and principal component space, thus suggesting population differentiation via climatic differences. I found moderate population structure for *I. hexagona*, however, population differentiation varied dependent on the pairwise comparison, thus implying some low level(s) gene flow between populations. Furthermore, I detected a small proportion of loci that demonstrated significant frequency shifts between populations (i.e. “outlier loci”) and are considered to be under diversifying selection. There was a relationship between pairwise geographic distance and genetic distance for neutral SNPs, which suggests that neutral genetic structure is at drift-migration equilibrium. I demonstrated allelic association for morphology and for a subset of outliers, which was greater than expected by chance, thus indicating pollinators could drive population genetic structure and variation in morphology may be under divergent selection for this species. The same pattern is not apparent for climate factors and outlier loci, thus not implicating the effect of climate in generating population level differences for floral morphology. It appears that both neutral and selective processes are at play in generating population divergence, in terms of morphology and population differentiation, for *I. hexagona*. At the micro-evolutionary scale, local adaptation often occurs within a species and this can thought of the as the beginning of divergence and potentially ecological speciation.

In chapter four, I demonstrate that the Louisiana iris species complex is of recent origin, dating the initial divergence for the Louisiana iris species clade within the Mid – Late Pleistocene. The strong differences observed among the iris species in regard to associated environmental factors suggest an effect from these components on the distributions and habitats occupied. Furthermore, even given this relatively recent divergence, evidence of niche divergence is apparent for all pairwise comparisons. More specifically, niche divergence occurs with lineage diversification as is seen in the sister species comparison as well as with the most recent common ancestor. Furthermore, niche divergence is not necessarily a product of whether the species are allopatric or sympatric in terms of their geographic ranges as niche divergence holds for all comparisons. Here, ecological speciation via divergent natural selection to habitat differences could be the first component in diversification. Although speciation could be due to abiotic or biotic factors that were not incorporated into niche models, these variables do highlight an initial pre-zygotic factor that could be contributing to divergence, potentially via immigrant inviability.

Moving forward, an analysis incorporating a larger number of populations and individuals found throughout all of the species ranges is necessary. Furthermore, by analyzing the role of ecological factors and geography in a Bayesian framework will allow one to assess the relative importance of each (Bradburd et al. 2013). Here, we can focus on environmental/climate factors or pollinator differences to understand the patterns that have been documented within this species complex. The range of the Louisiana irises, which is the southeastern United States, is an ideal setting to compare the interaction of landscape with the dispersal ability of various pollinators. As seen in three species of *Penstemon*, a similar pattern was found via the dispersal capabilities of their associated pollinators which were also either bumblebees or hummingbirds (Kramer et al. 2011). Thus, given enough time these pollinator dispersal differences could allow for the species to diverge and continue on various evolutionary trajectories.

Furthermore, while I focused on local adaptation within a single species, *I. hexagona*, this work could also be extended to *I. brevicaulis*, as it is also geographically widespread. One could ask the same question: Do we see evidence of local adaptation within *I. brevicaulis*? And if so, can we determine the selection pressures driving those differences? As irises are both long-lived and can be replicated clonally by subdividing rhizomes of a given genotype, performing a reciprocal transplant study for populations found throughout the range would also test for local adaptation within a species. If the average fitness of local individuals is greater than that compared to immigrants then this is suggestive that selection is relatively stronger than the action of other evolutionary forces (Lenormand 2012; Blanquart et al. 2013). Moreover, Quantitative Trait Locus (QTL) mapping project, involving either an F2 or backcrossing design within a species could identify the size and effect of QTL contributing to population differences for various traits. A number of QTL analyses have been performed, albeit, with crosses between species to understand the genetic architecture associated with parental species and the hybrid genotypes (Arnold et al. 2012) and to attempt to explain the genetic basis of variation in complex traits. Crossing experiments within populations have not been performed to date.

Moreover, resolving the evolutionary relationships among recently diverged species can be difficult especially in species with incomplete reproductive isolation as with the Louisiana irises. As hybridization is difficult to disentangle from incomplete lineage sorting as both processes can result in similar phylogenetic signals. Here, the first attempt at understanding species relationships for the Louisiana irises found high support for *I. brevicaulis* and *I. fulva* as sister species. However, only chloroplast genetic sequences were used and thus warrants for additional data to resolve the Louisiana iris species phylogeny. By incorporating a large number of genetic information found both in the chloroplast and nuclear genome, one may infer a different pattern for the relationship of these species. Phylogenetic methods have recently been developed in order to account for discordance across multiple markers and can also determine the extent and directionally of introgression (Durand et al. 2011; Pease and Hahn 2014). Performing these analyses in the Louisiana irises would verify if introgression has occurred and the species relationships in a more robust manner.

Disentangling the effects of environmental differences, local adaptation, and gene flow will require additional work both within irises and across a diversity of systems to understand their contributions and actions during divergence. Nonetheless, some broad patterns emerge from this system with regard to divergence both between and within species. Evidence of ecological speciation is apparent for both sister species and to the more distantly related species. It thus appears that divergent natural selection, potentially in terms of immigrant inviability even when species exhibit microhabitat differences, is enough for ecological speciation. Furthermore, the possibility of on-going gene flow for species found in sympatry does not necessarily inhibit the process of ecological speciation. Lastly, mirroring other studies, it appears that local adaptation is one of the initial steps for ecological speciation.

Literature Cited

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. Heredity 108:159-166.

Barton, N. 2010. What role does natural selection play in speciation? Philosophical Transactions of the Royal Society B: Biological Sciences 365:1825-1840.

Blanquart, F., O. Kaltz, S. L. Nuismer, and S. Gandon. 2013. A practical guide to measuring local adaptation. Ecology letters 16:1195-1205.

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67:3258-3273.

Cruzan, M. B. and M. L. Arnold. 1994. Assortative mating and natural selection in an Iris hybrid zone. Evolution:1946-1958.

dePamphilis, C. W. and R. Wyatt. 1989. Hybridization and introgression in buckeyes (Aesculus: Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. Systematic botany:593-611.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for Ancient Admixture between Closely Related Populations. Molecular Biology and Evolution 28:2239-2252.

Gavrilets, S. and J. B. Losos. 2009. Adaptive radiation: contrasting theory with data. Science 323:732-737.

Kramer, A. T., J. B. Fant, and M. V. Ashley. 2011. Influences of landscape and pollinators on population genetic structure: examples from three Penstemon

(Plantaginaceae) species in the Great Basin. American journal of botany 98:109-121.

Lenormand, T. 2012. From local adaptation to speciation: specialization and reinforcement. International Journal of Ecology 2012.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2005. Loci affecting long-term hybrid survivorship in Louisiana irises: Implications for reproductive isolation and introgression. Evolution 59:2116-2124.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2006. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics 172:2481-2489.

Martin, N. H., Y. Sapir, and M. L. Arnold. 2008. The genetic architecture of reproductive isolation in Louisiana irises: pollination syndromes and pollinator preferences. Evolution 62:740-752.

Nosil, P. 2012. Ecological speciation. Oxford University Press.

Papadopulos, A. S., M. Kaye, C. Devaux, H. Hipperson, J. Lighten, L. T. Dunning, I. Hutton, W. J. Baker, R. K. Butlin, and V. Savolainen. 2014. Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130342.

Pease, J. B. and M. W. Hahn. 2014. Detection and Polarization of Introgression in a Five-taxon Phylogeny. bioRxiv.

Rundle, H. D. and P. Nosil. 2005. Ecological speciation. Ecology letters 8:336-352.

Steffan-Dewenter, I. and T. Tscharntke. 1999. Effects of habitat isolation on pollinator communities and seed set. Oecologia 121:432 - 440.

Tang, S., R. A. Okashah, S. J. Knapp, M. L. Arnold, and N. H. Martin. 2010. Transmission ratio distortion results in asymmetric introgression in Louisiana Iris. BMC plant biology 10:48.

Via, S. 2009. Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106:9939-9946.

Wolf, J. B., J. Lindell, and N. Backström. 2010. Speciation genetics: current status and evolving approaches. Philosophical Transactions of the Royal Society B: Biological Sciences 365:1717-1733.

APPENDIX A

Supplemental Material for Chapter 2, published in *Ecology and Evolution*

Table S1. Observed and total heterozygosity and inbreeding coefficients as calculated by Genodive for all populations. a) *I. brevicaulis* with 389 SNP markers and b) *I. fulva* 561 SNP markers. G_{is} values in bold are significantly different than zero.

a	H_o	H_t	G_{is}
IB_AL	0.101	0.144	**0.3**
IB_AR	0.212	0.217	**0.024**
IB_IL	0.172	0.189	0.09
IB_LA	0.156	0.19	**0.175**
IB_OH	0.259	0.201	-0.291
IB_TX	0.104	0.145	**0.282**

b	H_o	H_t	G_{is}
IF_AR	0.38	0.357	-0.066
IF_IL	0.328	0.361	0.09
IF_KY	0.314	0.347	0.097
IF_LA	0.545	0.403	-0.351
IF_MS	0.561	0.399	-0.405
IF_MO	0.447	0.384	-0.164

H_o = Observed Heterozygosity, H_t = Total Heterozygosity, and G_{is} = Inbreeding coefficient

Figure S1.

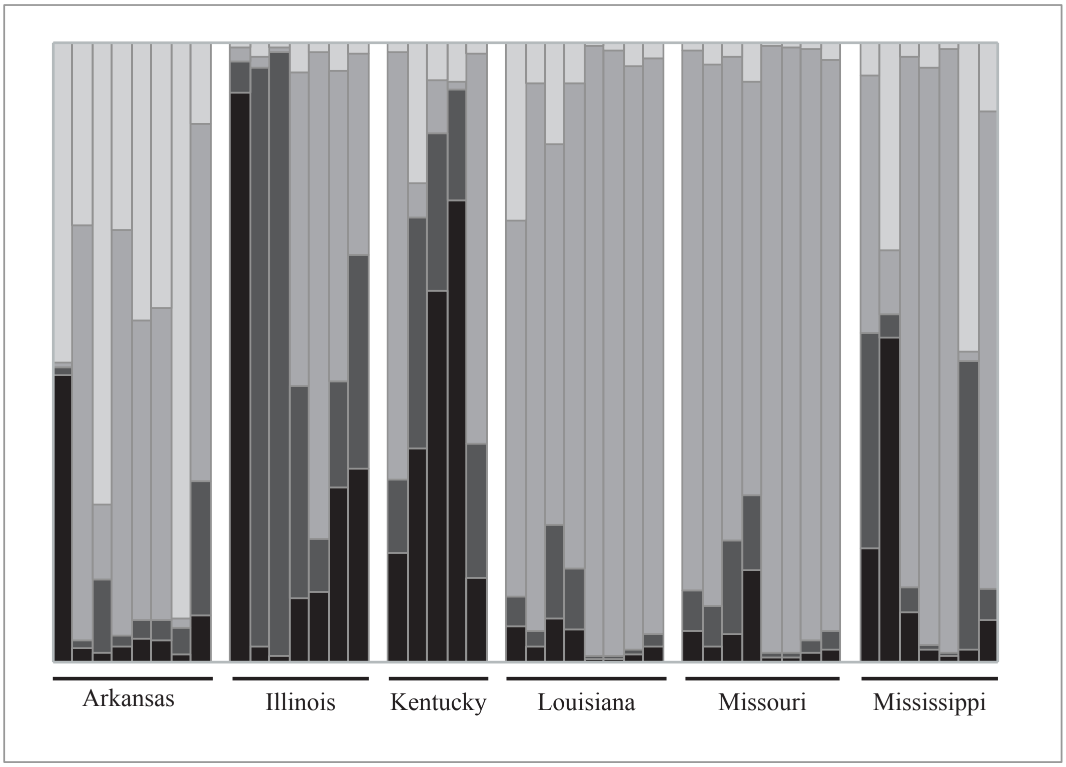


Figure S1. Plots of posterior probabilities of group assignments of each individual into four clusters based on the STRUCTURE analysis for *Iris fulva*. The results are grouped by collection localities for each individual.

Figure S2.

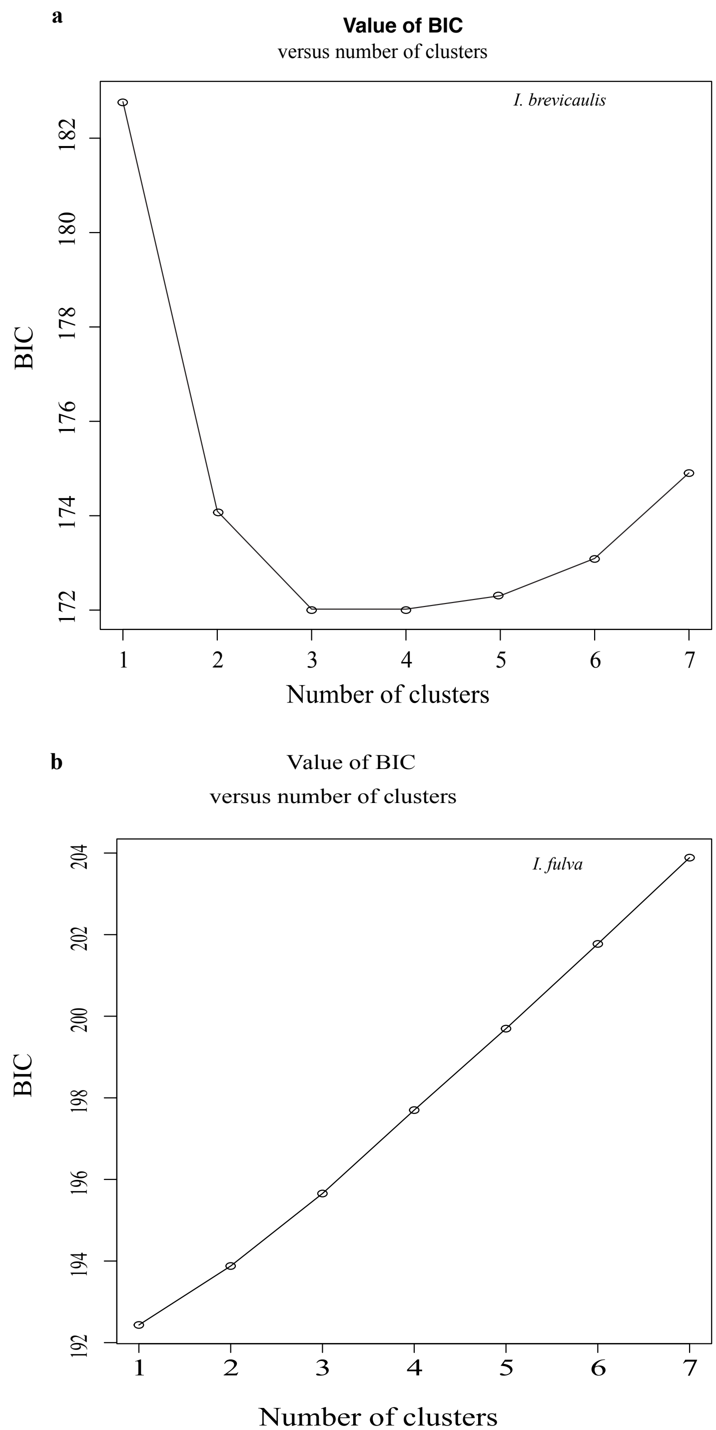


Figure S2. Inference of the number of genetic clusters by discriminant analysis of principle components (DAPC). The lowest Bayesian information criterion (BIC) value are found for a) *I. brevicaulis* to be 3 clusters and b) for *I. fulva* to be one cluster.

Figure S3.

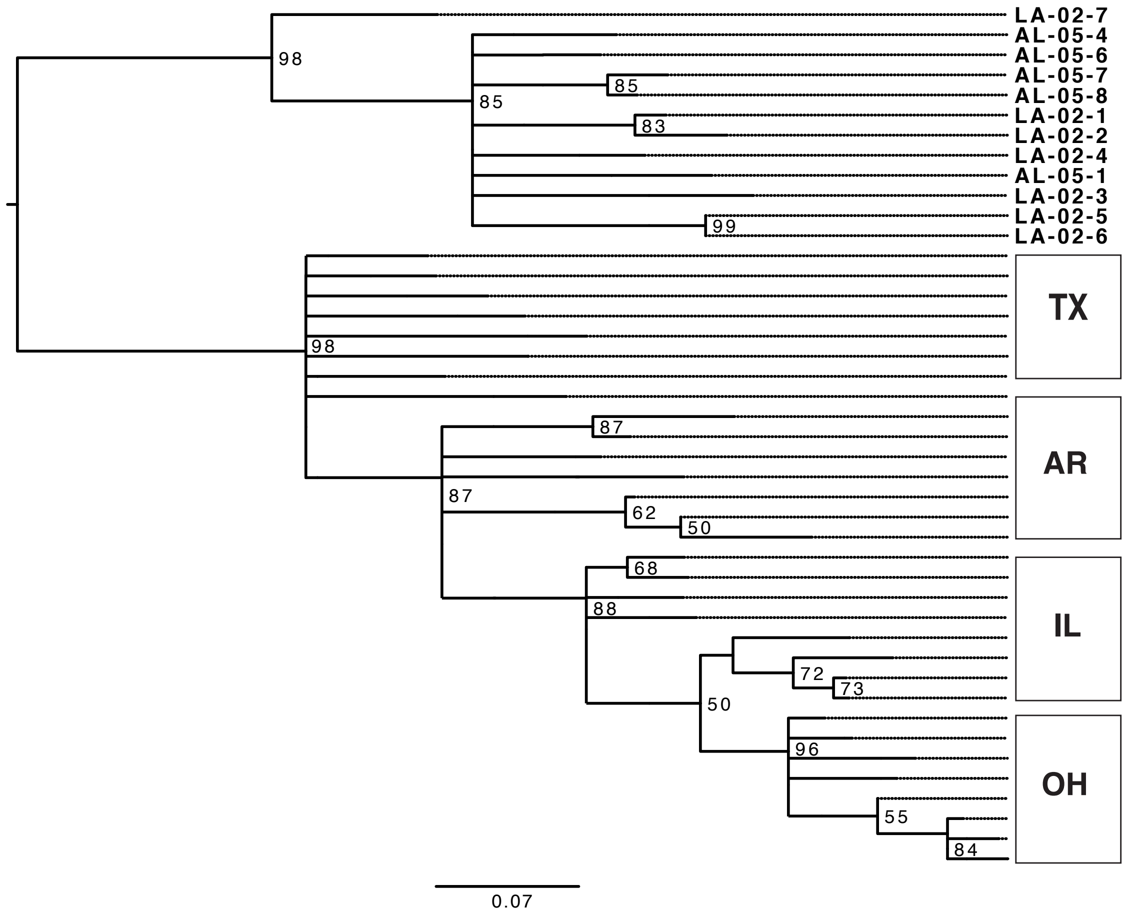


Figure S3. RAxML tree inferred using 387 concatenated SNPs showing the phylogenetic relationship of *I. brevicaulis* populations. Individuals from Louisiana and Alabama form one clade, while all other individuals form a clade based on the respective collection locality.

Figure S4.

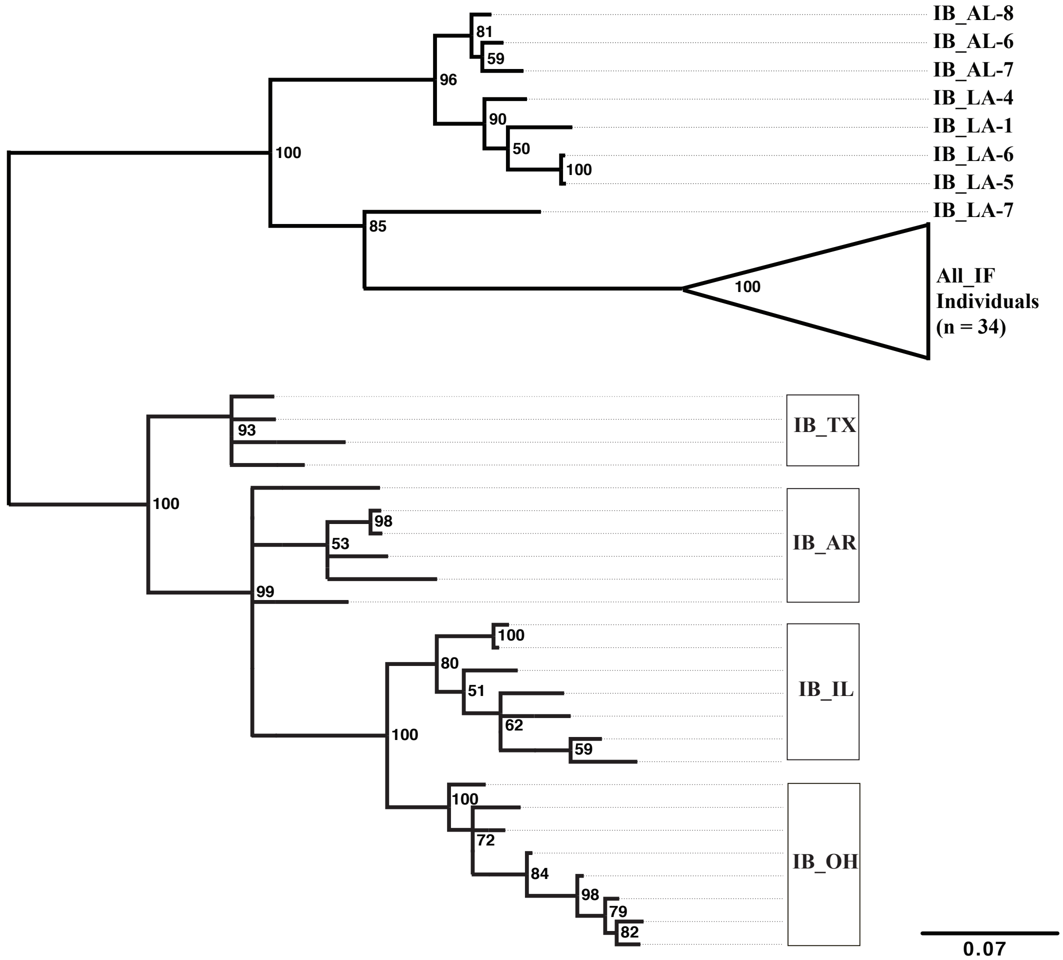


Figure S4. RAxML tree inferred using 468 concatenated SNPs showing the phylogenetic relationship of *I. brevicaulis* and *I. fulva* populations.

APPENDIX B

Supplemental Material for Chapter 3, submitted to *Journal of Heredity* 03/05/15

Table S1. Uncorrelated environmental variables used in PCA.

Environmental Variable	**Description**
Bio2	Mean diurnal range
Bio5	Max temperature of warmest month
Bio8	Mean temperature of wettest quarter,
Bio9	Mean temperature of driest quarter
Bio10	Mean temperature of warmest quarter
NDVI	Normalized Difference Vegetation Index (greenness)
NDVISTD	Greenness seasonality
TREE	Percent tree cover
QSCAT	Canopy or surface moisture and roughness

Table S2a. Outliers significantly correlated with morphology principal component 1.

lm(data$71 ~ hex$PC1)					
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.75	0.08542	8.78	0.000121	***
hex$PC1	-0.21316	0.04803	-4.438	0.004386	**
lm(data$117 ~ hex$PC1)					
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.75	0.08542	8.78	0.000121	***
hex$PC1	-0.21316	0.04803	-4.438	0.004386	**
lm(data$349 ~ hex$PC1)					
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.68269	0.04638	14.719	6.18E-06	***
hex$PC1	-0.23364	0.02608	-8.959	0.000108	***
lm(data$480 ~ hex$PC1)					
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.75	0.08542	8.78	0.000121	***
hex$PC1	-0.21316	0.04803	-4.438	0.004386	**

Table S2b. Outlier that significantly correlated with morphology principal component 2.

lm(data$184 ~ hex$PC2)					
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.3056	0.1039	2.943	0.0259	*
hex$PC2	0.3555	0.1138	3.124	0.0205	*

Figure S1.

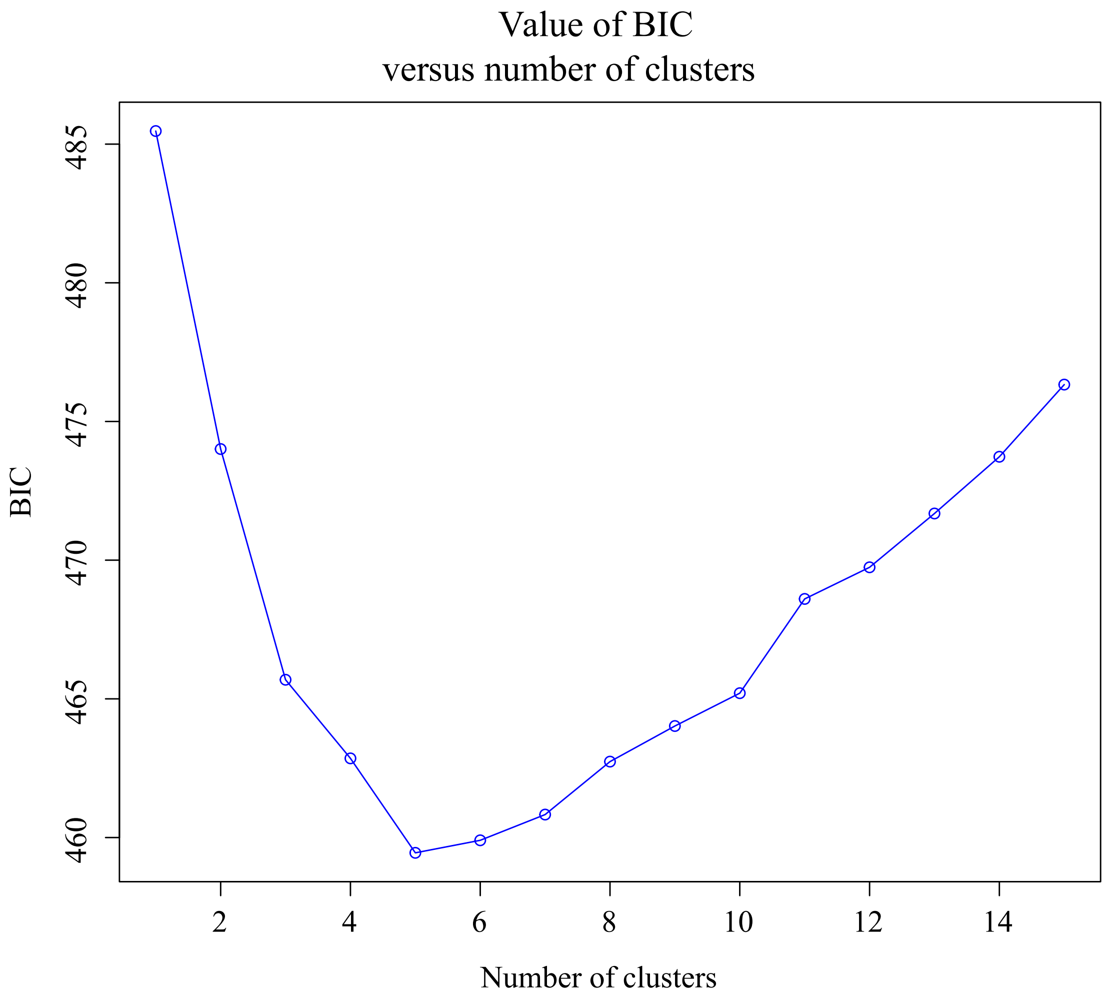


Figure S1. Inference of the number of genetic clusters by discriminant analysis of principle components (DAPC). The lowest Bayesian information criterion (BIC) value was determined to be 5 genetic clusters for *I. hexagona*.

Figure S2.

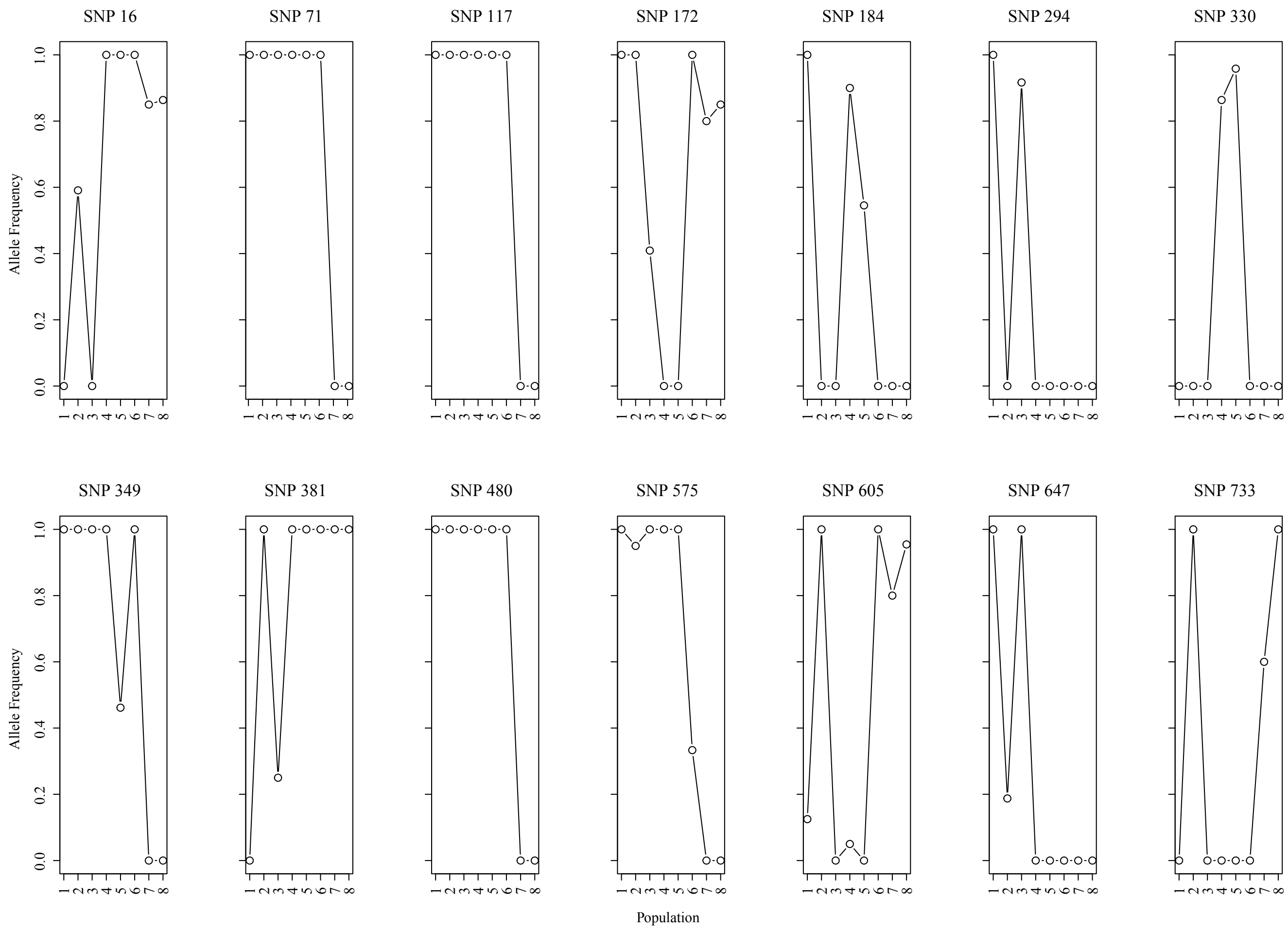


Figure S2. Population allele frequencies for 14 SNPs determined to be outliers. We plotted the associated allele for Florida (either fixed or nearly fixed), which shows the abrupt shift from Florida to Louisiana in terms of the SNP.

Figure S3.

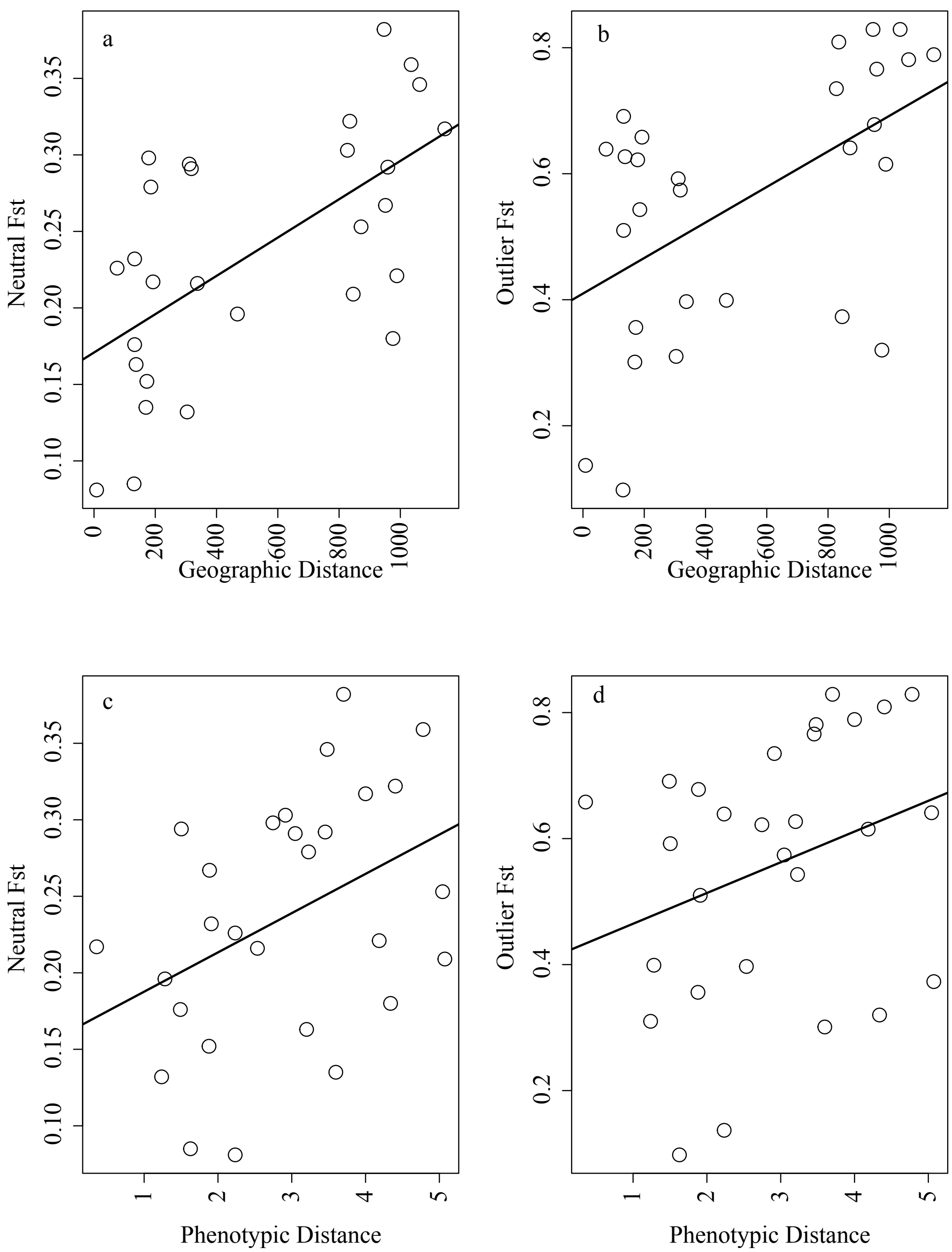


Figure S3. Mantel tests for neutral and outlier SNPs with either geographic or phenotypic distance. a) Neutral genetic distance and geographic distance (R = 0.612, p = 0.011). b) Positive association for non-neutral genetic distance and geographic distance (R = 0.534, p = 0.016). c) No association for neutral genetic distance and phenotypic distance (R = 0.406, p = 0.056). d) Moderate and nearly significant association of non-neutral genetic distance and phenotypic distance (R – 0.299, p = 0.05).

APPENDIX C

Supplemental Material for Chapter 4, to be submitted to *Journal of Biogeography*

Table S1. Collection information for the Louisiana irises used in the sequencing analysis

Species	Latitude	Longitude	State
***I. brevicaulis* 1**	41.361	-82.551	OH
***I. brevicaulis* 2**	35.171	-88.267	TN
***I. brevicaulis* 3**	31.902	-86.239	AL
***I. fulva* 1**	36.518	-89.336	KY
***I. fulva* 2**	35.200	-90.003	TN
***I. fulva* 3**	33.413	-90.155	MS
***I. hexagona* 1**	25.790	-81.100	FL
***I. hexagona* 2**	26.978	-81.493	FL
***I. hexagona* 3**	29.843	-92.128	LA

Table S2. Collection localities used to build each species ecological niche mode

Species	Latitude	Longitude
Iris brevicaulis	42.033	-82.583
Iris brevicaulis	41.683	-83.367
Iris brevicaulis	41.362	-81.551
Iris brevicaulis	41.214	-82.331
Iris brevicaulis	40.88	-85.483
Iris brevicaulis	39.922	-92.635
Iris brevicaulis	39.59	-83.222
Iris brevicaulis	39.536	-95.269
Iris brevicaulis	39.309	-94.288
Iris brevicaulis	39.243	-82.445
Iris brevicaulis	39.188	-94.984
Iris brevicaulis	39.148	-88.944
Iris brevicaulis	39.116	-94.716
Iris brevicaulis	38.926	-89.114
Iris brevicaulis	38.851	-94.343
Iris brevicaulis	38.64	-90.45
Iris brevicaulis	38.521	-91.115
Iris brevicaulis	37.443	-89.396
Iris brevicaulis	37.266	-88.249
Iris brevicaulis	37.233	-85.714
Iris brevicaulis	37.156	-88.146
Iris brevicaulis	37.136	-85.425
Iris brevicaulis	36.973	-90.196
Iris brevicaulis	36.953	-88.55
Iris brevicaulis	36.93	-88.61
Iris brevicaulis	36.743	-88.856
Iris brevicaulis	36.722	-88.123
Iris brevicaulis	36.571	-88.33
Iris brevicaulis	36.555	-88.364
Iris brevicaulis	36.539	-89.274
Iris brevicaulis	36.528	-87.99
Iris brevicaulis	36.518	-89.336
Iris brevicaulis	36.443	-88.512
Iris brevicaulis	36.432	-88.073
Iris brevicaulis	35.569	-88.809
Iris brevicaulis	35.341	-88.483
Iris brevicaulis	35.2	-90.003
Iris brevicaulis	35.172	-88.265
Iris brevicaulis	35.164	-88.265
Iris brevicaulis	35.102	-88.155

Table S2 continued.

Species	Latitude	Longitude
Iris brevicaulis	35.095	-88.156
Iris brevicaulis	35.074	-93.075
Iris brevicaulis	34.859	-91.499
Iris brevicaulis	34.759	-90.129
Iris brevicaulis	34.624	-91.592
Iris brevicaulis	34.24	-90.275
Iris brevicaulis	34.235	-91.952
Iris brevicaulis	33.836	-93.23
Iris brevicaulis	33.744	-90.318
Iris brevicaulis	33.729	-90.91
Iris brevicaulis	33.413	-90.155
Iris brevicaulis	32.966	-85.076
Iris brevicaulis	32.773	-91.854
Iris brevicaulis	32.345	-93.791
Iris brevicaulis	32.205	-86.098
Iris brevicaulis	32.154	-86.121
Iris brevicaulis	32.154	-86.117
Iris brevicaulis	32.102	-86.129
Iris brevicaulis	32.063	-93.502
Iris brevicaulis	32.027	-86.39
Iris brevicaulis	31.918	-86.869
Iris brevicaulis	31.908	-86.178
Iris brevicaulis	31.903	-86.239
Iris brevicaulis	31.667	-81.838
Iris brevicaulis	31.38	-84.634
Iris brevicaulis	30.905	-84.96
Iris brevicaulis	30.776	-92.059
Iris brevicaulis	30.694	-90.859
Iris brevicaulis	30.693	-84.852
Iris brevicaulis	30.624	-92.019
Iris brevicaulis	30.568	-96.201
Iris brevicaulis	29.331	-95.155
Iris fulva	37.443	-89.396
Iris fulva	37.442	-89.395
Iris fulva	37.398	-88.971
Iris fulva	36.973	-90.196
Iris fulva	36.899	-89.365
Iris fulva	36.857	-89.95
Iris fulva	36.817	-89.825
Iris fulva	36.596	-90.238

Table S2 continued.

Species	Latitude	Longitude
Iris fulva	36.585	-90.602
Iris fulva	36.556	-89.226
Iris fulva	36.539	-89.274
Iris fulva	36.528	-89.289
Iris fulva	36.527	-89.263
Iris fulva	36.524	-89.314
Iris fulva	36.523	-89.277
Iris fulva	36.518	-89.336
Iris fulva	36.517	-89.338
Iris fulva	36.513	-89.28
Iris fulva	36.474	-89.315
Iris fulva	36.45	-89.383
Iris fulva	36.293	-89.204
Iris fulva	36.28	-89.185
Iris fulva	36.27	-89.226
Iris fulva	36.139	-89.489
Iris fulva	36.084	-90.602
Iris fulva	36.082	-89.292
Iris fulva	36.00	-83.5
Iris fulva	35.972	-87.994
Iris fulva	35.56	-88.809
Iris fulva	35.49	-88.57
Iris fulva	35.334	-90.065
Iris fulva	35.2	-90.003
Iris fulva	35.2	-90.036
Iris fulva	35.18	-90.637
Iris fulva	35.125	-89.829
Iris fulva	35.084	-89.596
Iris fulva	35.084	-89.596
Iris fulva	35.073	-89.495
Iris fulva	34.76	-90.129
Iris fulva	34.626	-90.364
Iris fulva	34.603	-91.115
Iris fulva	34.35	-91.952
Iris fulva	34.241	-90.275
Iris fulva	34.195	-91.971
Iris fulva	33.773	-90.078
Iris fulva	33.765	-90.734
Iris fulva	33.744	-90.318

Table S2 continued.

Species	Latitude	Longitude
Iris fulva	33.704	-90.729
Iris fulva	33.458	-90.924
Iris fulva	33.413	-90.155
Iris fulva	33.405	-91.06
Iris fulva	32.141	-92.068
Iris fulva	31.216	-89.15
Iris fulva	30.999	-91.8975
Iris fulva	30.643	-91.99
Iris fulva	30.634	-91.609
Iris fulva	30.376	-91.162
Iris fulva	30.34	-91.0477
Iris fulva	30.291	-91.147
Iris fulva	30.219	-91.063
Iris fulva	30.181	-90.927
Iris fulva	30.164	-90.882
Iris fulva	30.04	-90.463
Iris fulva	29.997	-90.589
Iris fulva	29.972	-90.836
Iris fulva	29.85	-89.763
Iris hexagona	30.431	-96.16
Iris hexagona	29.847	-92.117
Iris hexagona	29.573	-91.913
Iris hexagona	29.745	-91.809
Iris hexagona	29.736	-91.804
Iris hexagona	29.880	-91.7842
Iris hexagona	30.439	-91.26
Iris hexagona	29.844	-90.705
Iris hexagona	30.084	-90.4496
Iris hexagona	30.552	-89.682
Iris hexagona	30.438	-88.612
Iris hexagona	31.528	-87.778
Iris hexagona	32.136	-87.402
Iris hexagona	31.105	-87.072
Iris hexagona	32.433	-86.933
Iris hexagona	32.016	-86.359
Iris hexagona	31.858	-86.175
Iris hexagona	30.957	-85.516
Iris hexagona	31.256	-85.111
Iris hexagona	30.191	-84.207
Iris hexagona	29.814	-83.346

Table S2 continued.

Species	Latitude	Longitude
Iris hexagona	29.777	-83.325
Iris hexagona	30.831	-83.283
Iris hexagona	29.675	-83.254
Iris hexagona	29.645	-83.15
Iris hexagona	31.159	-83.066
Iris hexagona	29.239	-82.932
Iris hexagona	31.036	-82.889
Iris hexagona	29.296	-82.826
Iris hexagona	29.117	-82.767
Iris hexagona	29.253	-82.724
Iris hexagona	29.191	-82.683
Iris hexagona	27.106	-82.331
Iris hexagona	26.985	-82.317
Iris hexagona	29.985	-82.317
Iris hexagona	28.717	-82.236
Iris hexagona	27.375	-82.226
Iris hexagona	27.144	-82.184
Iris hexagona	27.7	-82.183
Iris hexagona	27.196	-82.143
Iris hexagona	27.22	-82.125
Iris hexagona	28.507	-82.1249
Iris hexagona	27.245	-82.124
Iris hexagona	27.267	-82.1181
Iris hexagona	27.597	-82.113
Iris hexagona	28.527	-82.097
Iris hexagona	28.463	-82.0548
Iris hexagona	29.207	-82.019
Iris hexagona	29.182	-81.992
Iris hexagona	27.083	-81.985
Iris hexagona	27.482	-81.919
Iris hexagona	27.538	-81.901
Iris hexagona	27.488	-81.743
Iris hexagona	29.983	-81.675
Iris hexagona	27.401	-81.615
Iris hexagona	26.833	-81.6
Iris hexagona	27.442	-81.581
Iris hexagona	29.984	-81.562
Iris hexagona	26.84	-81.56
Iris hexagona	27.433	-81.559
Iris hexagona	27.477	-81.557

Table S2 continued.

Species	Latitude	Longitude
Iris hexagona	27.418	-81.55
Iris hexagona	26.825	-81.492
Iris hexagona	28.712	-81.469
Iris hexagona	26.816	-81.467
Iris hexagona	27.011	-81.455
Iris hexagona	26.8	-81.45
Iris hexagona	26.783	-81.45
Iris hexagona	26.944	-81.449
Iris hexagona	26.8	-81.3
Iris hexagona	26.933	-81.267
Iris hexagona	25.883	-81.15
Iris hexagona	25.790	-81.100
Iris hexagona	25.783	-81.092
Iris hexagona	25.867	-81.058
Iris hexagona	27.244	-80.851

Table S3. Uncorrelated environmental variables used to test for niche divergence with all species.

Environmental Variable	Description
Bio2	Mean diurnal range
Bio5	Max temperature of warmest month
Bio8	Mean temperature of wettest quarter,
Bio9	Mean temperature of driest quarter
Bio10	Mean temperature of warmest quarter
Bio12	Annual precipitation
Bio15	Precipitation seasonality
NDVI	Normalized Difference Vegetation Index (greenness)
NDVISTD	Greenness seasonality
TREE	Percent tree cover
QSCAT	Canopy or surface moisture and roughness

Figure S1.

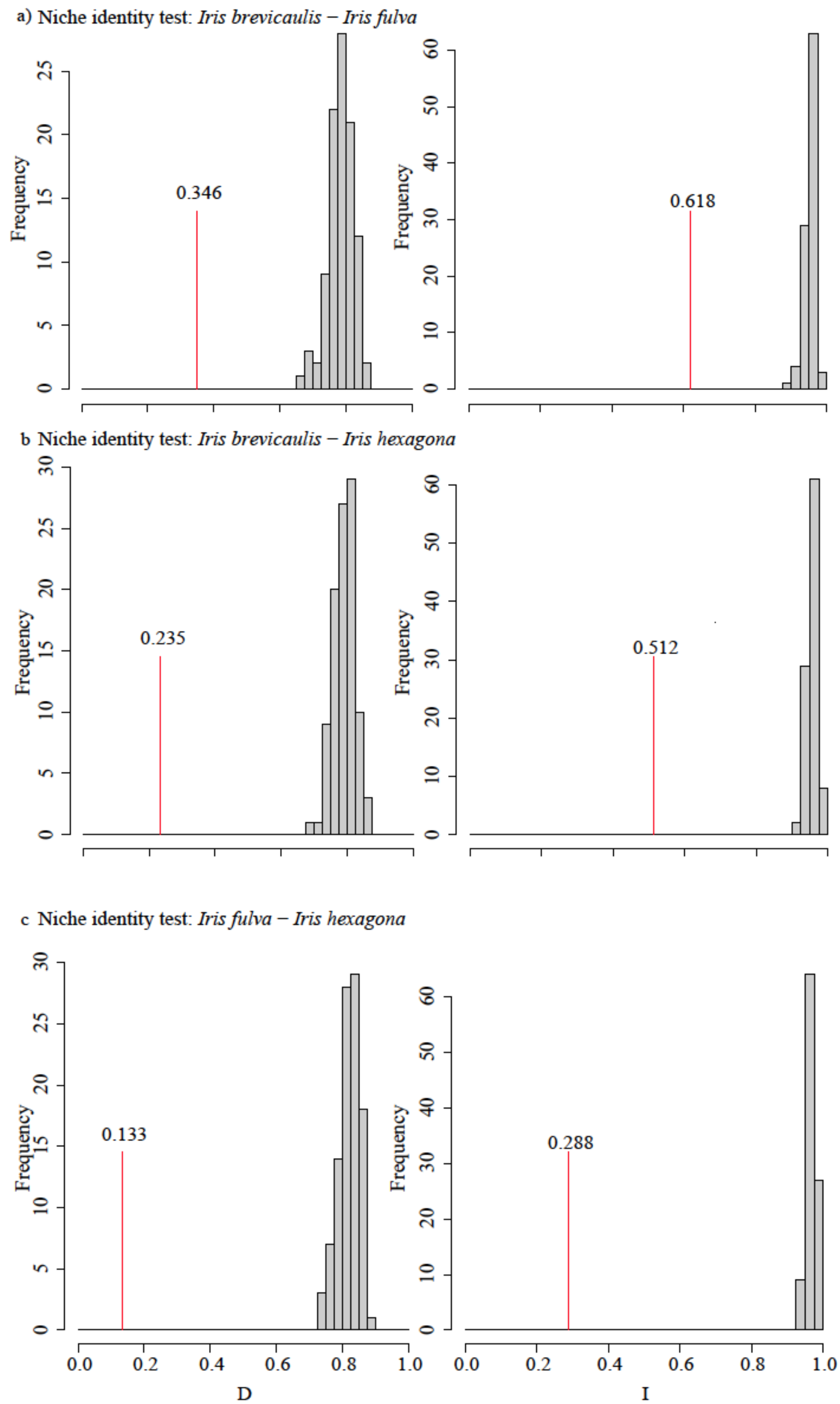

a) Niche identity test: Iris brevicaulis – Iris fulva
b Niche identity test: Iris brevicaulis – Iris hexagona
c Niche identity test: Iris fulva – Iris hexagona
Frequency
0.346
0.618
0.235
0.512
0.133
0.288
D
I

Figure S1. Frequency histograms from tests of niche identity for both metrics of similarity (*D* and *I*). The red line represents the observed value of niche overlap and when that value is smaller than the null distribution than it indicates the species niches are not equivalent. a) *Iris brevicaulis* compared to *Iris fulva*; b) *Iris brevicaulis* compared to *Iris hexagona*; c) *Iris fulva* compared to *Iris hexagona*.