THE GENOMIC ARCHITECTURE OF REPRODUCTIVE ISOLATION IN A LOUISIANA IRIS HYBRID ZONE

by

Cheng-Jung Sung, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biology
August 2016

Committee Members:

Noland Martin, Chair

Chris Nice

James Ott

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

I earnestly thank my advisor, Dr. Noland Martin who guided me on the scientific research, and always cares about his students in all aspects in life. I also thank my committee members, Dr. Chris Nice and Dr. James Ott for their help and suggestions on this thesis. Particularly, I thank Kate Bell for her great help on many concepts and techniques used in this thesis. I also thank Sunni Taylor and Alex Zalmat for advising me a lot on the project. Of course, I thank my family members who always support me on my decisions and everything to take care of my studying. Also, many thanks to the SF-trip friends who always mentally support each other with respect to life in USA, and also thank all my roommates and friends. Specially, I deeply thank my lovely pet, Hanchi, for all the mental support. To have this thesis finished, I give all my thanks to everyone and everything that ever helped me during my endeavor to perfect this research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

I. INTRODUCTION

Speciation is thought to primarily occur through the accumulation of a diverse array of reproductive isolating barriers that arise between genetically diverging lineages (Coyne and Orr 1989; Orr 2001). Because the evolution of complete reproductive isolation is rarely an instantaneous process, natural hybridization and gene flow can occur during speciation (Arnold 1997; Buerkle and Lexer 2008; Gompert and Buerkle 2010), and such natural hybridization can have a wide array of evolutionary consequences, including amalgamation of the divergent lineages' genomes (Arnold and Meyer 2006), further evolution of prezygotic isolation through reinforcement (Coyne and Orr 1989, 2004; Diaz and Macnair 1999; Lowry et al. 2008), introgression of selectively advantageous – or neutral – alleles across species boundaries (Wang et al. 1997; Arnold 2006; Martin et al. 2006; Arnold and Martin 2009; Gompert and Buerkle 2009), or even the formation of new reproductively isolated hybrid species (Arnold et al. 1990b; Rieseberg et al. 1990; Gompert and Buerkle 2009). It is now widely understood that divergence often occurs even in the face of natural hybridization and gene flow, with some regions of the genome being quite resistant to introgression while other regions remain "porous" to gene flow (Gompert and Buerkle 2011). New sequencing technologies and analytical techniques now enable a better understanding of not only the genetic architecture of reproductive isolation and introgression, but also the very barriers that are responsible for effecting reproductive isolation at the genomic scale (Gompert and Buerkle 2009; Cruickshank and Hahn 2014).

Natural hybrid zones, areas where genetically divergent groups of organisms meet, mate and produce offspring, provide opportunities to examine a diverse set of

evolutionary processes that occur during species divergence. While a great deal of progress has been made examining the genetic architecture of a wide variety of reproductive isolating barriers using laboratory crosses across a wide range of taxa (e.g. QTL mapping - Rieseberg et al. 1999; Martin et al. 2007; Taylor et al. 2009; Ballerini et al. 2012), it is often not realistic to perform crosses for species-pairs that either have long generation times or are not easily reared in laboratory conditions (Mandeville et al. 2015). Natural hybrid zones also provide an additional advantage for studying the genomic architecture of reproductive isolation and adaptive introgression in that hybrids have been "tested" by natural selection for several generations, and the fitness consequences of specific genomic combinations as they occur in nature can be assessed (Burke et al. 1998; Gompert and Buerkle 2009, 2011). By taking advantage of the wide variety of late-generation genotypic combinations that result from several generations of recombination and backcrossing, it is possible to identify the genomic distribution of specific loci that are responsible for restricting gene flow – or conversely loci that are prone to introgress across species boundaries at non-neutral rates (Gompert and Buerkle 2009; Rieseberg and Buerkle 2002). Further, using genome-wide association mapping (GWAS) techniques, it is now even possible to not only identify genetic markers that are associated with phenotypic variation but also examine such markers are associated with the genetic architecture of known reproductive isolating barriers at a truly genomic scale – all in natural settings (Parchman et al. 2012; Johnston et al. 2014).

The Louisiana Iris species complex provides a unique opportunity to study the evolutionary, genetic, and ecological dynamics of hybridization and species divergence. The Louisiana Iris clade (*Iris* series *Hexagonae*) contains four closely related and phenotypically differentiated species - *Iris hexagona*, *Iris fulva*, *Iris*

brevicaulis and *Iris nelsonii* - that exhibit incomplete reproductive isolation and form large natural hybrid zones where they co-occur in sympatry. A broad suite of traits are known to influence both pre and post-zygotic isolation between these four species, and a number of quantitative genetic studies have been performed that describe the genetic architecture of these traits (Arnold 1993; Martin et al. 2005, 2006, 2007; Taylor et al. 2009, 2012a; Arnold et al. 2010; Tang et al. 2010; Ballerini et al. 2012; Hamlin and Arnold 2014). Because of the extensive genetic and ecological work on this species complex, Louisiana Iris is now considered a "model system" for investigating speciation and divergence when gene flow is present (Lexer and Widmer 2008). The current study focuses on two widespread Louisiana Iris species that hybridize where they co-occur in nature: *Iris fulva* and *Iris hexagona*. These two species have a number of morphological and ecological factors that differentiate them, including differences in habitat preference, floral and vegetative morphological traits, as well as pollination syndromes. These differences contribute to the partial reproductive isolation that exists between the two species even when they occur in sympatry (Martin et al. 2005, 2006, 2007; Arnold et al. 2010; Tang et al. 2010; Ballerini et al. 2012).

In Louisiana, *I. fulva* is predominantly encountered in lower-elevation sites in both shade and sun alongside rivers and bayous (Taylor et al. 2011), its flowers are copper red and attractive to hummingbirds - the primary pollinators (Emms and Arnold 2000; Martin et al. 2008; Taylor et al. 2012a, 2012b) - and the whole plant is relatively short with thinner leaves than *I. hexagona* (Taylor et al. 2011). In contrast, *Iris hexagona* occurs primarily as large sunny populations in coastal wetlands. It is much taller than *I. fulva*, and the blue flowers are about twice as large with predominate yellow nectar guides on the sepals. These flowers are visited - and

pollinated - primarily by native bumblebees (Taylor et al. 2012a, 2012b). A large amount of prezygotic isolation thus exists between these two species such that F_1 hybridization is exceedingly rare in sympatric populations (Arnold 2000; Martin et al. 2007; Taylor et al. 2009; Ballerini et al. 2012). However, despite the rarity at which they are formed, F_1 hybrids exhibit high fitness (Arnold et al. 1990a, 1990b; Burke et al. 1998; Taylor et al. 2009; Ballerini et al. 2012) and once formed can readily backcross with either parental species resulting in extensive hybrid zones where their ranges overlap in southern Louisiana (Burke et al. 1998; Arnold et al. 2010; Brothers et al. 2013). Within many of these natural hybrid zones, there has been extensive admixture (Gompert and Buerkle 2009) and late generation backcrossing which can allow for high scale resolution of genomic admixture (Arnold et al. 1990b; Arnold and Meyer 2006; Arnold and Martin 2009).

In the current study we aim to extend our understanding of the genomic architecture of reproductive isolation and adaptive introgression in Louisiana Iris by utilizing a large and phenotypically diverse hybrid zone between *I. fulva* and *I. hexagona* located in an area immediately surrounding Lake Martin in St. Martin Parish in Southern Louisiana. In the current study, genome wide sequence data along with Bayesian genomic cline (BGC) analyses were utilized in this hybrid zone to detect genomic regions that reveal exceptional patterns of introgression in order to identify loci that are responsible for both reproductive isolation and adaptive introgression. In addition, a large number of individual hybrid plants were phenotyped for several floral, vegetative, and ecological traits that are putatively associated with reproductive isolating barriers (Martin et al. 2005, 2007; Taylor et al. 2009, 2012A; Brothers et al. 2013). Genome wide association mapping (GWAS) studies were then by using the program pi-MASS, a Bayesian variable selection regression (BVSR)

model, to examine the genomic architecture of reproductive isolation and adaptive introgression. Further, tests were performed to determine whether there were significant associations between loci that reveal exceptional patterns of introgression (e.g. “outlier loci”) and loci that are associated with phenotypic trait variation, as this would be consistent with the hypothesis that the phenotypic traits examined act as reproductive isolating barriers and/or facilitate introgression in nature.

II. METHODS

Study Site and Sample Collection

A single large *I. hexagona* × *I. fulva* hybrid zone was studied along the perimeter of Lake Martin (30.221015° N, 91.910969° W) - a Nature Conservancy operated preserve in St. Martin Parish in Southern Louisiana. Leaf tissue samples and phenotypic measurements were taken from 400 different iris genotypes at this location. This natural hybrid zone is composed of a wide variety of morphological and late-generation hybrids that exhibit a diverse array of flower colors and plant-growth morphologies. Pure *I. fulva* individuals can be found at drier margins of the lake, owing to their preference for drier soils, while pure *I. hexagona* individuals are encountered more towards the interior of the lake where there is deeper-water habitat and lower tree canopy cover. Most of the iris individuals in this collection locale were determined to be of hybrid origin between *I. fulva* and *I. hexagona*, on the basis morphological and genomic characteristics. In addition, two nearby allopatric *I. fulva* sites were identified from Lottie, LA (30.55676° N, 91.64553° W - 8 individuals sampled) and Livonia, LA (30.55491° N, 91.57348° W - 11 individuals sampled), and 19 individuals were sampled from a single allopatric *I. hexagona* site near Abbeville, LA (29.48013° N, 91.47288° W). Young leaves were collected from each of the identified non-clonal iris genotypes by cutting off approximately 1 cm × 4 cm of tissue and placing the leaf samples in individually-labeled coin envelopes. GPS locations were recorded for each genotype sampled. The samples were placed in sealed plastic Tupperware ® containers that were filled with clay pet litter as a desiccant and stored at room temperature until DNA was extracted from the dried leaves.

Phenotypic Variation

In early March to the end of April of 2013, a single habitat variable (canopy cover) and a number of floral and vegetative phenotypes were assessed for all 400 Louisiana Iris individuals sampled located in the hybrid zone at Lake Martin. Individual Louisiana Iris flowers typically persist for up to three days, and because floral morphologies change as the flowers mature (Smyth et al. 1990), measurements of floral morphology were made using the first apical flower on the second day that the flower was open. Flowers were removed from the stalk, and the individual floral organs including the sepals, petals, stigmas and anthers with filaments still attached, were individually dissected and photographed on grid paper (0.25 inch × 0.25 inch grid size). The following 10 floral morphological traits were measured using a hand-held ruler to the nearest 0.5 mm (Figure 1, Table 1): (1) nectar guide area (NGA) is the yellow roughly-triangular area on the iris sepal, calculated as one half of the length times the width of this triangle. "Pure-species" *I. fulva* individuals do not normally reveal a nectar guide, and individuals with no nectar guide received a measurement of "zero" for this measure; (2) (ANL) is the length of the anther; (3) stylar branch length (STBL) was measured from the base of calyx to the tip of the stigma; (4) anther extension (ANEX) was measured as the distance from the tip of the anther to the tip of the stigma, and was calculated as the length of the entire stamen (anther length + filament length) minus the STBL; (5) stylar branch width (STBW) was measured at the widest section observed across the stylar branch; (6) petal total length (PETTL) was measured from the base of calyx to the distal end of the petal; (7) petal width (PETW) was measured at the widest distance observed across the entirety of the petal; (8) sepal total length (SEPTL) was measured form the base of calyx to the distal end of the sepal and was calculated as the summation of the sepal stalk

length and the sepal blade length; and (9) sepal width (SEPW) was measured at the widest distance observed across the sepal blade. In Louisiana Iris, the shape of the sepals vary from spatulate (*I. hexagona*-like) to pendate (*I. fulva*-like - Bouck et al. 2007), and (10) sepal shape (SEPS) is reflected in - and measured by - the ratio of the sepal stalk length to the sepal blade length (Figure 1, Table 1).

Three vegetative traits were also measured to the nearest 0.159 cm (1/16 inch) using a yardstick (Table 1): (1) flower stalk height (STALK) is the height of the flower stalk that is measured from the base of the stalk at the rhizome to the base of the calyx of the apical flower; (2) leaf height (LEAF) is the height of the tallest leaf that was not associated with the flowering stalk, measured from the top of the rhizome to the tip of the leaf. (3) In both *I. hexagona* and *I. fulva*, flowering stalks extend higher than the leaves encountered on exclusively vegetative growth points. Relative stalk height (RLTH) was measured as the ratio of STALK relative to LEAF. *Iris hexagona* is typically encountered in sunnier locations than *I. fulva*, and canopy open space (CNP) measurements were taken at four cardinal directions around each individual plant using a convex densiometer held at chest height (Lemmon 1956). The densiometer used has 24 squares drawn onto a mirror. The number of squares were divided into four equal and smaller squares, and the presence and absence of canopy cover within each of these 96 squares was determined. Mean CNP was assessed for each plant by averaging across four measurements taken at each plant. To reduce observational error, all 14 traits were measured by the same person. Additionally, because trait correlations are common and may affect interpretations of studies that examine the genomic architecture of quantitative traits (Taylor et al. 2012A), correlations were quantified among all of the phenotypic and environmental traits measured in this study.

DNA Sequence Generation, Assembly and Variation

Genomic DNA was isolated from 346 of the 400 leaf samples collected at the Lake Martin hybrid zone (those with the most phenotypic data obtained were chosen for genotyping), as well as the 19 allopatric *I. fulva* and 19 allopatric *I. hexagona* genotypes (384 individuals total). DNA extractions were performed in 96-well format following a modified cetyltrimethyl ammonium bromide DNA extraction protocol (Doyle 1991). A single reduced-complexity genomic library was prepared for genotyping by sequencing following protocols modified from Meyer and Kircher (2010), Gompert et al. (2012), Parchman et al. (2012), and Mandeville et al. (2015). DNA from each sample was separately digested with the restriction enzymes *EcoRI* and *MseI* (New England Biolabs; NEB, Inc.). Fragments from each sample were then labeled by ligating 8-10 base pair adaptor oligonucleotides (barcodes) to the DNA fragments. Barcoded restriction-ligation products were run through two separate rounds of PCR amplification using standard Illumina primers, and the final PCR products of all sampled individuals were pooled. DNA fragments from the pooled PCR products were size-separated on a 2% agarose gel and fragments 300-380 base pairs in length were excised from the gel and subsequently purified using Qiaquick Gel Extraction kits (Qiagen Inc.). This final purified DNA library was sent to the University of Texas Genomic Sequencing and Analysis Facility (Austin, TX, USA), and three separate lanes were sequenced on an Illumina HiSeq 2500 platform after DNA quality and concentration was verified. After sequencing across the three lanes, a total of 323,074,637 reads with 84-86 bp DNA sequences were obtained (barcode and adapter sequences were removed).

Because a reference genome is not available for *Iris*, an artificial reference genome was created by choosing a random subset of 30 million reads and performing

a *de novo* assembly using the software SeqMan NGen ver. 11.0.0.172 (DNASTAR, Inc. - and following protocols modified from Gompert et al. 2012, 2014). A minimum match percentage of 92%, a match size of 71 bp, and a match spacing of 100 bp was used. Based on these criteria a total of 16,005,165 sequences assembled onto 2,180,875 separate contigs, and from these contigs consensus sequences were extracted. In order to identify possible homologs or recent paralogs, consensus sequences were then assembled to each other with a lower minimum match percentage of 83%. The sequences that assembled under these less-stringent conditions were removed from the dataset, and a total of 2,023,227 consensus sequences were retained as the reference sequence set. The full 323,074,637-read dataset was then assembled to the artificial reference genome using SeqMan xng ver. 11.0.0.172 (DNASTAR, Inc.).

Variable sites (SNPs) were then identified using samtools and bcftools ver. 0.1.18 to calculate the Bayesian posterior probability that each nucleotide was variable (Li et al. 2009). SNPs were further designated using two criteria. In the first, a minimum of 50% of all individuals must have had at least one read at a particular nucleotide site (i.e., the "d" parameter in bcftools was set at 0.5), while in the second, at least 90% (d = 0.9) of all individuals must have had at least one read at that site. It is important to note that genotypes were not "called" but rather genotype uncertainty was incorporated by using genotype likelihood estimates as the data. The former (d = 0.5) resulted in 153,748 variable sites, while the latter more stringent inclusion criteria (d = 0.9) resulted in 18,902 variable sites. All subsequent analyses were performed separately on these two datasets. Allele frequencies were estimated directly from genotype likelihood estimates and sorted to exclude loci where minor allele frequency (MAF) $\leq$ 5%. To reduce the effect of non-independence among physically linked loci,

only a single randomly-chosen SNP was included per reference consensus sequence (Gompert et al. 2014). In all, genotype likelihood data were obtained for a total of 45,384 loci where MAF ≥ 5% for the more inclusive dataset (d = 0.5), while 3,699 loci (MAF ≥ 5%) were identified for the dataset in which the d parameter was set at 0.9.

To quantify the genomic composition of the Louisiana Iris hybrid zone and the three allopatric pure-species collection locales, population genetic parameters were estimated using ENTROPY, an admixture proportion statistical model developed by Gompert et al. (2014; Mandeville et al. 2015). ENTROPY is a hierarchical Bayesian model similar to the correlated allele frequencies admixture model in STRUCTURE (Pritchard et al. 2000; Falush et al. 2003). Both models require prior specification of the number of ancestral populations to be modeled, yet do not require a priori assumptions as to which populations individuals might belong (Mandeville et al. 2015). The most important difference between the two models is that rather than employing a threshold of sequence coverage used to "call" genotypes, ENTROPY incorporates a variety of forms of genotypic uncertainty into its models by utilizing the genotype likelihood estimates derived from bcftools, which ultimately allows for a robust inference of genotype probabilities as well as a variety of other population parameters (Gompert et al. 2012a, 2014; Mandeville et al. 2015). In addition, ENTROPY utilizes a deviance information criterion (DIC) approach to compare among models with different numbers of ancestral populations (*k*), and here statistical support was examined in ENTROPY for $k = 1$ to $k = 8$ ancestral populations.

In the current study, posterior distributions of admixture proportions were calculated separately for $k = 2$ to $k = 8$ by performing 50,000 MCMC step chains, discarding the first 5,000 iterations as a burn-in, and sampling every 10^{th} iteration.

Posterior means, medians, and 95% credible intervals (CIs) were measured for a variety of parameters of interest, and in order to ensure proper mixing, sequential MCMC steps were plotted for a number of parameter estimates. To summarize the distribution of genotypic variation across all of the sampled hybrid individuals and the allopatric samples, principal component analysis (PCA) was performed using the PRCOMP function in R (R Development Core Team 2012) with genetic covariance matrices calculated from the genotype estimates generated by ENTROPY (Gompert et al. 2014). In addition, medians of admixture proportions (examined separately for k = 2 to k = 8) were determined for all individuals and plotted with structural plots (Gompert et al. 2014).

Identification of Genomic Introgression

To quantify genome-wide variation in introgression among admixed individuals of *I. fulva* and *I. hexagona* ancestry located in the Lake Martin Louisiana Iris hybrid zone, the Bayesian genomic cline (BGC) model developed by Gompert and Buerkle (2011, 2012) was used. This model uses genomic clines to describe patterns of introgression between parental species-pairs at individual loci, and it examines the probability of ancestry of one species (ranging from 0 to 1) as a function of hybrid index (h; also ranging from 0 to 1). This hierarchical model includes two basic locus-specific genomic cline parameters - α and β. In the current study, the parameter α reflects either an increase (positive α) or decrease (negative α) in the probability of *I. hexagona* ancestry for a locus relative to the null expectation that the probability of hybrid ancestry equals that of the hybrid index. Thus, positive values of α indicate directional movement of *I. hexagona* alleles into a largely *I. fulva* genomic background, while negative values of α indicate movement of *I. fulva* alleles into an *I. hexagona* background. The parameter β, specifies an increase (positive β, steeper

clines) or decrease (negative β, wider clines) in the rate of change, with positive values indicating limited amount of introgression between species while negative values indicate increasing introgression (Gompert and Buerkle 2011; Gompert et al. 2012b; Parchman et al. 2013). Marginal posterior probability distributions for α, β and *h* were estimated by performing two independent chains of MCMC with 100,000 steps each and 50,000 step burn-ins, sampled every 50th step. The outputs of the two chains were then inspected to determine whether both converged to the stationary distribution, and then combined. Posterior point estimates and 95% CIs for the two cline parameters, α and β, were calculated from the two merged chains using estpost. Outlier loci that had extreme values of α or β (when 95% CIs of the parameter value did not intersect 0) were identified.

Genomic Architecture of Phenotypic Variation

To quantify aspects of the genomic architecture of various traits believed to be important in affecting reproductive isolation and / or adaptive introgression between *I. fulva* and *I. hexagona*, Bayesian variable selection regression (BVSR) analyses were performed in the computer software pi-MASS (version 0.9) with SNPs as covariates (Guan and Stephens 2011). SNPs statistically associated with phenotypes of interest were identified, and their effect sizes were estimated. Additional model parameters that were estimated from the SNP and phenotypic data include the proportion of variance explained by all of the SNPs in the model (PVE), the conditional prior probability of a single SNP being included in the model (P_{SNP}), the number of SNPs included in the regression model (N_{SNP}), and the average phenotypic/additive effect associated with a SNP (σ_{AVE}).

In the current study, all 14 phenotypic datasets were normal quantile-transformed prior to analyses in order to standardize all traits to have a mean of 0 and

variance of 1 and allow for the comparison of phenotypic effect sizes of SNPs across all traits (Guan and Stephens 2011). For each phenotypic trait, a single MCMC chain was run in pi-MASS by performing a 1,000,000-step burn-in followed by sampling every 400th step of an additional 8,000,000 steps. A number of different prior options (see "Setup other parameters" section of pi-MASS user manual v. 0.9) were examined and the posterior probability traces of numerous parameters and hyperparameters were examined. Ultimately the priors that were specified in the results presented here restricted the hyperparameter *h* (the proportion of variance explained by the model) to be between 0.01 and 0.9, and the hyperparameter *p* (the conditional prior probability that specifies the sparsity of the model) to be between one and 1,000. Additionally, the minimum and maximum numbers of SNPs (*n*) that were included in the model were set to one and 100, respectively (Comeault et al. 2014).

Analyses were run separately for each trait, and the means and 95% equal-tail probability CIs were reported as point estimates for each of the following parameters: PVE, P_{SNP}, N_{SNP}, and σ_{AVE}. In addition, the Rao-Blackwellized posterior inclusion probability (PIP - the probability that a particular SNP is associated with phenotypic variation) was calculated for each locus (Gompert et al. 2012a; Comeault et al. 2014). To identify the SNPs that have the strongest association with phenotypic variation for each trait, all SNPs were sorted by the magnitude of their PIPs, and β (the Rao-Blackwellized estimated magnitude of their phenotypic effect) were reported for the three SNPs with the highest PIPs for each trait.

In order to explore associations between the genomic architecture of different traits and to determine whether genetic regions associated with phenotypic variation were also significantly associated with regions found to have significant α and β parameters from the genomic cline (BGC) analyses, the mean N_{SNP} (rounded to the

next-highest integer) was calculated for each trait, and only those SNPs with the highest PIP scores (N = mean N_{SNP}, hereafter referred to as "top SNPs") were considered for these comparative analyses. For each trait-by-trait comparison, the actual number of shared top SNPs (i.e. SNPs that influenced both traits) was identified, and the probability *p*, that the number of shared top SNPs were more common than expected by chance, was calculated using the following formula (modified from Moyle and Nakazato 2008):

$$\sum_{p=m}^{s} p = \frac{\begin{bmatrix} l \\ m \end{bmatrix} \cdot \begin{bmatrix} n-l \\ s-m \end{bmatrix}}{\begin{bmatrix} n \\ s \end{bmatrix}}$$

where *l* is the number of top SNPs in the larger sample (i.e. the trait that had the highest of the two mean N_{SNP} value), *s* is the number of top SNPs in the smaller sample (i.e. the trait that had the lowest mean N_{SNP} value), *m* is the number of top SNPs that were shared across both traits, and *n* is the total number of SNPs in the sample (45,384 in the current study).

Associations between top SNPs (from pi-MASS) and both the genomic cline parameters α and β (compared separately *and* together from BGC analyses) were also assessed using the same formula above. Because association mapping analyses consistently identified fewer top SNPs (for all traits) than markers that were identified as having significant α or β cline parameters, in the current study, *l* is the number of markers with significant α (or β - tested separately) identified from BGC analyses, *s* is the number of top SNPs identified from the association mapping study, *m* is the number of SNPs that were identified as significant in both BGC analyses and the association mapping study, while *n* is the total number of SNPs in the sample (again 45,384 in the current study).

III. RESULTS

Phenotypic Variation

All 14 trait measurements were variable and revealed continuous distributions within the hybrid population examined in this study (Figure 2). Both positive and negative correlations were noted between the 91 possible trait combinations with r ranging from -0.91 and 0.92, and 85 correlations being statistically significant ($P < 0.05$; Figure 3), which is consistent with previous studies performed in Louisiana Iris (Brothers et al. 2013). The weakest - and often nonsignificant - correlations were found when measures of CNP and RLTH were examined, while strong correlations greater than |0.8| were observed among many floral length measurements, such as STBL, PETTL and SEPTL (Figure 3). ANEX is a trait that included the other trait measurement (STBL) as a component measure, and as expected, the result showed a significant negative correlation between the two traits (correlation coefficient = -0.91; Figure 3). All traits revealed consistently positive correlations except for ANEX and RLTH which revealed several significant negative correlations among traits (Figure 3). This is consistent with previous findings that *I. fulva* plants are smaller - with much smaller flowers - than *I. hexagona* plants, yet the anther extension in *I. fulva* is greater than that of *I. hexagona*. RLTH is positively correlated with STALK because STALK is a component measure of RLTH (correlation coefficient = 0.34; Figure 3). However, all other correlations with this trait were negative because *I. fulva* plants have flower stalks that are significantly taller than the leaves (Figure 3). Trait combinations with CNP showed weaker correlations than the other traits, and CNP was not significantly correlated with four traits, including PETW and the three vegetative traits, STALK, LEAF and RLTH (Figure 3). However, significant positive correlations between CNP and eight floral measures were detected consistent with the observation that *I.*

hexagona plants (with larger flowers) are generally found in more open space than *I. fulva* throughout the species ranges.

Genetic Variation

All genetic analyses in this study were performed separately on two different datasets. One dataset utilized the more stringent inclusion-criteria for loci (d = 0.9) and resulted in 3,699 SNPs, while the more inclusive dataset included an order of magnitude more loci but with lower coverage (d = 0.5; 45,384 SNPs). Results obtained from analyses of the two datasets were very similar, however the d = 0.5 dataset, because of the larger number of markers used relative to the more stringent d = 0.9 dataset, resulted in more precise measures of hybrid index (*h*), the genomic cline parameters (α and β), as well as the model parameters estimated in the genomic architecture analyses. As an example, the distribution hybrid index estimates are very similar across both datasets. However, the range of the 95% credible intervals (CIs) for each of the 346 hybrid individuals is much narrower for the d = 0.5 dataset relative to that of the d = 0.9 dataset. This pattern was observed for all parameter estimates in this study, so only the d = 0.5 results are reported and discussed below.

When models of different numbers of *k* (1-8) were compared, DIC scores sharply decreased from *k* = 1 (DIC = 79543251.41) to *k* = 2 (DIC = 72823315.58) and from there decreased incrementally when additional populations were added to the model (for *k* = 8, DIC = 67209363.59) (Figure 4). Because *k* = 2 has similar statistical support to *k* = 3-8, and because it makes the most biological sense given what is known about the hybrid zone and nearby populations, only the results of *k* = 2 model are reported and discussed below.

Principal component analysis (PCA) clearly separated the two parental species, *I. fulva* and *I. hexagona* sampled from the nearby allopatric populations, while the PC

loadings of the putative hybrid individuals sampled from the Lake Martin hybrid zone were intermediate between the parental species (Figure 5). Distinct clusters were formed primarily based on PC1 which accounts for 38.7803% of the variation in the matrix of genotype covariance (Figure 5). The two allopatric *I. fulva* collection locales corresponded to the same cluster - not differentiated from one another - at one side of PC1 space, while individuals from the allopatric *I. hexagona* collection locale were clustered at the other side. The PC1 scores of Lake Martin hybrid zone individuals spanned the full range between parental species (Figure 5). Notably, a number of hybrid zone individuals were indistinguishable from the allopatric *I. fulva* cluster, as well as morphologically indistinguishable from putatively "pure" *I. fulva* plants.

Admixture proportions estimated using ENTROPY revealed similar results as the PC plots. The putatively "pure" *I. fulva* and *I. hexagona* allopatric individuals were assigned to two distinct groups which corresponded to the two distinct clusters in PC space. A single putative *I. hexagona* allopatric individual was revealed to have mixed *I. fulva* X *I. hexagona* ancestry, and this individual also was easily identifiable in the PC plot (Figure 5). Most Lake Martin individuals had admixed genomes from both of the two parental species, except for some individuals that were morphologically indistinguishable from individuals in the *I. fulva* allopatric collection locales (i.e. *I. fulva*; Figure 6). This is consistent with the PCA plot interpretation which also indicated that the Lake Martin hybrid zone contained an abundance of admixed hybrid individuals derived from the two ancestral species, *I. fulva* and *I. hexagona*, while there were a few putatively "pure" *I. fulva* genomes within Lake Martin (Figure 5 and 6).

Identification of Genomic Isolation and Introgression

Hybrid indexes for individuals sampled in the Lake Martin hybrid zone ranged

from 0.0972 to 0.7481 with $h = 0$ and $h = 1$ indicating "pure" *I. fulva*, and "pure" *I. hexagona* respectively. The medians and 95% CIs of these posterior estimates are plotted in Figure 7. As mentioned above, the posterior estimates of hybrid indexes are similar across individuals for the d = 0.5 and d = 0.9 datasets, but as the former utilized a larger number of SNPs, this resulted in dramatically smaller 95% CIs (Figure 7), and thus the genomic clines generated from this dataset are discussed in the current manuscript. Similar to the distribution of admixture proportion from ENTROPY (Figure 6), the hybrid indexes (Figure 7) showed an identical distribution for hybrids in between the two parental species (Figure 8). By the very similar patterns, the two distributions for the hybrid individuals gave us consistent results and robust interpretation on the hybrid genomic proportion derived from the two parental species, but noticeably, due to the two different models used for analyses, the range of the distributions are different (0.0000 – 0.9994 for the admixture proportion of *I. hexagona* ancestry of hybrids in ENTROPY and 0.0972 – 0.7481 for hybrid indexes of *I. hexagona* ancestry; Figure 8). Interestingly, the small *I. fulva*-like group (of 36 hybrid individuals collected from the Lake Martin) was identified from both analyses, but the admixture proportion of *I. hexagona* genome for those 36 individuals ranged from 0.0000 to 0.0003 in ENTROPY analyses while showing a range from 0.0972 to 0.1393 for *h* in BGC analyses. The possible reason to explain this is that those 36 *I. fulva*-like individuals might not be the same as the allopatric parental genomes for the hybrids, in other words, the allopatric putative "pure" *I. fulva* individuals are actually not the real source genomes of gene flow for the Lake Martin hybrid population. The other possible reason is that the real parental genomes were included in the model while the *I. fulva*-like individuals were actually the same as the "pure" parental genomes but were assigned to be hybrids, so that these 36 individuals were forced to

be calculated as hybrids to not having 0 as their hybrid indexes.

The genomic cline parameters, α and β, were quite variable across the loci examined in this study with posterior estimates of the median values α ranging from -1.47 to 1.33 and β ranging from -0.88 to 1.92. A number of outlier loci with values for α and/or β where the 95% CIs failed to intersect 0 were identified. In all, 14,261 of 45,384 sampled loci (31.4%) deviated significantly from the genome wide average, with 13,468 loci (29.7%) having extreme α values, and 1,569 loci (3.5%) having extreme β values. Only 776 loci (1.7%) revealed extreme values for both α and β.

Among the 13,468 α outlier loci, 7,885 exhibited excess *I. hexagona* ancestry (the lower-bound 95% CI did not overlap with zero, Figure 9A), and 5,583 loci exhibited excess *I. fulva* ancestry (the upper-bound 95% CI did not overlap with zero, Figure 9A). An exact binomial sign test revealed that the number of positive α outliers is significantly greater than the number of negative α outliers ($P < 0.0001$), indicating that *I. hexagona* alleles are generally more favored to introgress compared to *I. fulva* alleles in the Lake Martin hybrid population. Of the 1,569 β outlier loci, 1,339 revealed exceptionally steep clines (positive β values) indicating significantly reduced gene flow and possible association with reproductive isolation, while only 230 loci revealed exceptionally shallow clines (negative β values) indicating that heterospecific alleles at these loci might be favored to introgress into both species backgrounds (Figure 9B). An exact binomial sign test revealed that this nearly seven-fold difference in the number of positive versus negative β outliers was significant ($P < 0.0001$), demonstrating that reproductive isolation happened more than the bidirectional adaptive introgression.

Genomic Architecture of Phenotypic Variation

The leaves and flower stalks of *Iris fulva* are generally shorter than those of

pure-species *I. hexagona*; the flowers of *I. fulva* are generally smaller than those of *I. hexagona*; and the anthers of *I. fulva* generally extend past the stigma much farther than those of *I. hexagona* (Brothers et al., 2013). In the Lake Martin hybrid zone, strong associations between hybrid indexes and phenotypic variation were found. Each of the 14 individual trait values were regressed against hybrid index for the 346 hybrid individuals (with $h = 0$ reflecting pure-species *I. fulva*, and $h = 1$ reflecting pure-species *I. hexagona*; Figure 10). All linear regressions are statistically significant ($P < 0.05$), and all results are consistent with the phenotypic differences that exist among the species that have been documented in other studies of Louisiana Iris (e.g. plants with higher hybrid indexes tending to display *I. hexagona*-like phenotypes).

Variation for each of the 14 phenotypic traits was modeled as a function of 45,384 SNPs using BVSR. The model parameters obtained from pi-MASS outputs included PVE, P_{SNP}, N_{SNP} and σ_{AVE}, and the point estimates of each of the means along with their 95% CIs are shown in Table 2. The means of PVE were large and varied among traits from 0.5005 (RLTH) to 0.8616 (STBL; Table 2). The model explained the largest amounts of phenotypic variation for floral length measurements ($PVE_{ANL} = 0.8240$; $PVE_{STBL} = 0.8616$; $PVE_{PETTL} = 0.8491$; $PVE_{SEPTL} = 0.8360$) as well as ANEX (which has STBL as a component measurement - $PVE_{ANEX} = 0.8601$; Table 2). Estimates of the mean number of SNPs included in the models (N_{SNP}) varied among all 14 traits and ranged from 44.8871 (RLTH) to 80.9265 (LEAF), though the CIs were quite large for some of the traits (Table 2). The average effects of the associated SNPs (σ_{AVE}) also varied among traits and ranged from 0.3741 (PETW) to 0.6849 (ANEX), and these measures varied similarly to those of PVE, where larger measures of PVE were associated with larger measures of σ_{AVE} (Table 2). Phenotypic effect sizes (β) for individual SNPs were very small for the vast majority of top SNPs

across all traits. Measures of |β| for the top SNPs ranged from 0.0050 to 0.4283, with the vast majority of |β| measures being less than 0.05 for all traits (Figure 11). This indicates that a large number of genes of small effect are responsible for the phenotypic differences observed between the two species. Furthermore, top SNPs with |β| greater than 0.05 were compared among 14 traits, and there are two SNPs shared between STBL and ANEX (locus 17152 - $|\beta|_{STBL} = 0.1060$, $|\beta|_{ANEX} = 0.3732$; and locus 27191 - $|\beta|_{STBL} = 0.0735$, $|\beta|_{ANEX} = 0.1182$), and one SNP shared between PETTL and SEPTL (locus 31243 - $|\beta|_{PETTL} = 0.1891$, $|\beta|_{SEPTL} = 0.1686$). These three SNPs that have $|\beta| > 0.05$ (i.e., stronger phenotypic effects) seem to work and influence on the morphologies more than others, noticeably, locus 31243 is also one of the 12 top SNPs that are shared by three traits, and it implies that this SNP might function well on several morphologies simultaneously and give greater effects on them, thus it becomes one of the most responsible genomic regions to speciation.

Many top SNPs were associated with multiple traits, and these associations occurred more than expected by chance (i.e., out of a total of 91 pairwise comparisons, 33 traits pairs shared at least one top SNP (Table 4, 5). This result is consistent with the significant phenotypic correlations (Figure 3). In particular, the floral length measurements, STBL, PETTL, and SEPTL, which exhibited the highest trait correlations (Figure 3) had several top SNPs that were significantly associated with respect to the pairwise comparisons among the three traits (Table 4). The highly correlated floral width measurements STBW, PETW, and SEPW also shared a large number of top SNPs, as did the length and width measurements of the same floral parts (i.e., STBL and STBW, PETTL and PETW, and SEPTL and SEPW; Table 4). Also consistent with phenotypic correlations, traits that represented combination measures shared top SNPs with their component traits. For example, there were eight

shared top SNPs detected between the traits ANEX and STBL. Furthermore, the three vegetative traits measurements (STALK, LEAF, and RLTH) shared top SNPs among them as well (Table 4). Notably, SEPS was the only trait that did not share top SNPs with other traits (Table 4). Twelve SNPs were significantly associated with variation in three traits (Table 5). SNPs 1475, 31243, and 36350 were associated with the three floral traits PETTL, SEPTL, and SEPW, while SNPs 11462 and 19160 were associated with the floral traits PETTL, SEPTL, and PETW. The remaining seven SNPs that were associated with variation in three traits each revealed unique three-trait combinations - though each included at least one of the floral traits PETTL, PETW, or SEPTL.

Genomic Introgression and Genomic Architecture × Trait Associations

Significant relationships were found among loci that revealed extreme parameter estimates for α and/or β in the genomic cline analyses and those loci with the highest PIPs included in the models describing the genomic architecture of the 14 floral and plant stature traits (Table 7). Evidence for associations between those highest PIP regions for SEPS and |α| (especially +α; P = 0.0055 Table 7) were found. A similar pattern in which a significant association between highest PIP regions for ANL and +α was also found (P = 0.0375 Table 7). These results indicate that at genomic regions affecting both SEPS and ANL, *I. hexagona* alleles are more likely to be favored to introgress. There was also evidence for weak associations between the highest PIP loci affecting STALK and the genomic cline parameter |β| (P = 0.0383). This association was primarily driven by loci with +β estimates and indicates that STALK may be an important phenotypic trait causing reproductive isolation between these two species. Highest PIP loci affecting the traits PETTL and SEPTL were also significantly associated with loci that revealed significant |α| *and* |β| estimates. A total

of 94 tests for genomic architecture × genomic cline parameter associations were performed. Were these tests independent, one would expect 4.7 of the tests to reveal significant associations by chance at $\alpha = 0.05$ (We found seven tests to be significant). However, these tests are decidedly not independent as traits are highly correlated and the tests examining $|\alpha|$ or $|\beta|$ are not independent from tests examining $+/-\alpha$ and $+/-\beta$. For this reason, we report uncorrected P-values.

IV. DISCUSSION

Understanding the genomic architecture of prezygotic and postzygotic barriers that lead to speciation is an important topic of research (Brothers et al. 2013; Turelli et al. 2013; Mandeville et al. 2015). Barriers that prevent gene flow can be categorized based on the timing at which they occur during the life cycles of hybridizing populations, with prezygotic barriers - such as habitat isolation and pollinator isolation - potentially playing the most "important" role in reducing introgression because of the fact that they act earlier in the life cycle of the hybridizing organisms (Martin et al. 2008; Taylor et al. 2009; Tang et al. 2010; Brothers et al. 2013; Singhal and Moritz 2013). In Louisiana Iris, F1 and late-generation hybrid fertility and viability is quite high, and the total isolation observed between these species is influenced largely by prezygotic barriers (Martin et al. 2007; Arnold and Martin 2009; Taylor et al. 2009; Ballerini et al. 2012). However, despite this strong prezygotic isolation – and because of the high fitness of Louisiana Iris hybrids - hybrid zones can often be encountered where species come into geographic contact. These hybrids have been shown to act as "bridges" for gene flow – even adaptive gene flow – across Louisiana Iris species boundaries despite much of the genome being resistant to introgression (Martin et al. 2005, 2008; Taylor et al. 2009; Ballerini et al. 2012).

Standard QTL mapping techniques using laboratory crosses have allowed for a dissection of the genetic architecture of a number of prezygotic and postzygotic isolating barriers that prevent gene flow between Louisiana Iris species (e.g., Slotman et al. 2004; Martin et al. 2005, 2006, 2007, 2008; Taylor et al. 2009; Tang et al. 2010; Ballerini et al. 2012). Studies in natural hybrid zones using a small number of genetic markers have also revealed that interspecific introgression occurs between Louisiana

Iris species (Arnold et al. 1990A, B; Arnold 1993). Despite this introgression, however, these species largely maintain their phenotypic integrity likely as a result of the various reproductive barriers that exist. The genetic architecture of those barriers points to the fact that a large number of genomic regions (i.e. QTLs) are scattered in isolated locations throughout the genome, and because of this fact, Louisiana Iris have been considered a "model system" for studying a "genic view" of speciation (Arnold et al. 1990; Arnold 1993; Taylor et al. 2009; Tang et al. 2010; Ballerini et al. 2012; Hamlin and Arnold 2014). Here, the degree to which 45,384 SNPs were associated with genomic introgression and phenotypic trait variation were examined at a very large and phenotypically diverse *I. fulva* × *I. hexagona* Louisiana Iris population.

BGC (Identification of Genetic Introgression)

In the current study, reproductive isolation and introgression were examined at a genomic scale, and the rates and form of introgression were found to be variable across the genome. Of the 45,384 loci examined, almost a third (31.4%) revealed significant deviations with respect to the genomic cline parameters α and/or β. These loci are thus potentially linked to genomic regions responsible for reproductive isolation and/or adaptive introgression between *I. fulva* and *I. hexagona.* Among the 13,468 loci that revealed significant deviations for cline parameter α, 7,885 revealed excess ancestry for *I. hexagona* while 5,583 revealed excess ancestry for *I. fulva*. This asymmetry, in which *I. fulva* alleles are slightly (but significantly) underrepresented relative to those of *I. hexagona*, runs counter to some studies in Louisiana Iris wherein *I. fulva* alleles have generally been shown to be selectively advantageous in hybrid genomic backgrounds, as well as studies that reveal *I. fulva* alleles tend to introgress more often than those of *I. brevicaulis* or *I. hexagona* in natural hybrid zones (reviewed in Arnold et al. 2010). However, this is not universally the case. For

example, in experiments examining populations of *I. brevicaulis* × *I. fulva* reciprocal backcross hybrids, *I. brevicaulis* alleles were generally favored (in both reciprocal backcross populations) and resulted in increased survivorship of adult plants in dry greenhouse environments (Martin et al. 2005), while *I. fulva* alleles (again using the same reciprocal backcross populations) tended to increase survivorship in flooded field conditions (Martin et al. 2006). Thus, hybrid fitness (and directionality of introgression) is likely to be habitat-dependent.

There were a number of loci in the present study that deviated significantly with respect to the genomic cline shape parameter β as well, with 1,569 loci (3.5%) having extreme β values. Of these, over 85% were revealed to have steep clines indicating low rates of introgression at these markers. These results are consistent with a number of QTL mapping studies examining the genetic architecture of a diverse array of reproductive isolating barriers between Louisiana Iris species which show that loci that are important for reproductive isolation are scattered throughout the Iris genome (Martin et al. 2006, 2007, 2008; Taylor et al. 2009, 2012A; Ballerini et al. 2012). Only a very small number of loci revealed extreme negative β parameters (i.e. those loci in which heterospecific alleles - and thus bidirectional introgression are favored in the genomic background of both species), and this is also consistent with the fact that introgression is largely asymmetric when observed in Louisiana Iris (Arnold et al. 2010). Overall, genomic cline analyses indicates (reaffirms) that reproductive isolation and introgression in this Louisiana Iris hybrid zone have a complex genetic basis.

Genomic Architecture of Phenotypic Variation

Iris hexagona and *Iris fulva* differ with respect to the floral, vegetative, and ecological characters examined here, and strong associations were found between

phenotypic variation and the hybrid indexes. Strong phenotypic correlations among all the traits were also found as well, which has been observed in all QTL mapping studies examining the genetic architecture of ecologically important traits performed in Louisiana Iris (Arnold et al. 1990a; Taylor et al. 2009; Ballerini et al. 2012). Also consistent with previous QTL mapping studies, a large number of SNPs were found to explain a modest to large proportion of variation in morphology (Table 2). PVE was high (mean for all traits ranged from 0.5 - 0.86), with relatively small credible intervals for most traits, though for RLTH and PETW credible intervals were quite large (0.119 - 0.815 and 0. 209 - 0.853 respectively). For those two traits, there was also considerable uncertainty surrounding the number of SNPs in the model as well, with the credible intervals being 3 - 97 for RLTH and 5 - 99 for PETTL. These credible intervals nearly span the entire prior range (1 - 100). Such large credible intervals for these two traits are consistent with other studies using pi-MASS to perform GWAS. For example, in a study examining genomic architecture of male genetalic morphology and oviposition preference in two separate populations of *Lycaeides* butterflies, credible intervals for *N*snp estimates for all traits ranged from 1 to 83 and 1 to 94 with priors set at a maximum of 100 (as in the current study), and mean PVE estimates ranged from 0.049 to 0.241 (with the lowest credible intervals approaching zero for each character examined - Gompert et al. 2012). In a study examining the genetic architecture of color, size, and shape phenotypes in *Timema* stick insects, similarly large credible intervals with respect to the number of SNPs included in the model were observed (ranging from 1 - 59 and 1 - 97, again with the priors ranging from 1 - 100) for five of the ten traits examined (Gompert et al. 2014). These traits were interpreted to have a complex genomic architecture in which many genes of small to moderate effect were associated with phenotypic variance. The

remaining five traits revealed much smaller *N*snp credible intervals ranging from 1 to 59 and 3 to 26, and these traits were interpreted as each having major-effect loci influencing trait variation. In a third study examining genetic architecture of two great tit (*Parus major*) populations, high credible intervals for the number of SNPs contributing to the phenotypic variation were observed (ranging from 0 - 160 and 25 - 306) for eight phenotypic traits. However, since the authors used different priors from the current study (as well as different from those of Gompert et al. 2012, 2014) comparisons across these studies is difficult. Since the number of individuals phenotyped in the current study is more than twice that of both Gompert et al. (2012, 2014) studies, it is likely that the increases in PVE as well as the much narrower credible intervals for all parameter estimates reflects this increased sampling effort.

Overall, the current study is consistent with previous QTL mapping studies in Louisiana Iris examining the genetic architecture of similar phenotypic traits. For most traits examined here, many loci of small to moderate effect explain a very large proportion of the phenotypic variance. For ten of the traits, none of the estimated effect sizes for the individual SNPs (β) exceeded 0.25 (measured in standard deviations). For the remaining four traits, (NGA, ANEX, PETTL and CNP), a small minority of SNPs included in the models had moderate estimates of individual effect sizes (ANEX had four SNPs with measures of |β| ranging from 0.254 - 0.428, NGA had two SNPs where |β| ranged from 0.279 - 0.280, PETTL measures of |β| ranged from 0.296 - 0.424, and CNP had a single SNP where |β| was 0.292). Noticeably, each of these SNPs that had the largest effect sizes also had the highest posterior inclusion probabilities (PIP > 0.5 for each) in each model. Again, these results are corroborated by QTL mapping studies in *Iris* which identify many QTLs of varying effect sizes distributed throughout the genome that influence similar quantitative traits.

Many of the top SNPs were also shared across multiple traits as shown in Table 4, which is consistent with the strong correlations observed among the traits (Figure 3). Not surprisingly, combination traits were both highly-correlated with their component traits and often shared a significant number of top SNPs (i.e., ANEX and STBL). There were also a traits within the same floral component that were positively correlated and shared a significant number of top SNPs (i.e., PETTL and PETW). Further, there were a number of top SNPs that affected more than one floral component (i.e., several top SNPs were shared across SEPTL and PETTL). In fact, all but two traits (NGA, and SEPS) shared a statistically significant number of top SNPs with one or more other traits. Thus, the phenotypic correlations observed in this study are likely the result of shared genetic architectures among all of the traits examined.

Genomic Introgression and Genomic Architecture × Trait Associations

Almost a third of the SNPs examined in this study revealed significant deviations with respect to the genomic cline parameters α and/or β. Further, phenotypic variation in the traits examined here are associated with a large number SNPs. These traits are thought to be important in causing measurable amounts of reproductive isolation between *I. fulva* and *I. hexagona*, with differences in floral morphology likely attracting different pollinator suites (hummingbirds for *I. fulva* and bumblebees for *I. hexagona* - Wesselingh and Arnold 2000; Martin et al. 2006, 2008; Ballerini et al. 2012; Taylor et al. 2012a; Brothers et al. 2013), and differences in canopy cover, flowering stalk and leaf morphologies likely reflecting of the fact that the two species are adapted to different habitats. Thus an important question to ask is whether loci underlying the genetic architectures of each of these phenotypic traits are also identified (more often than not) as having exceptional genomic cline parameters from the BGC analyses. We found that top SNPs underlying the phenotypes SEPS,

STALK, and ANL were associated with the cline parameters α or β (Table 7). The +β association with STALK is consistent with the hypothesis that combinations of alleles that confer higher (*I. hexagona*-like) stalk heights, while combinations of alleles that confer lower (*I. fulva*-like) stalk heights were both favored by selection. The +α associations with SEPS and ANL are consistent with the hypothesis that *I. hexagona* alleles at loci affecting these traits were favored by selection over *I. fulva* alleles. The latter finding is perhaps consistent with asymmetric introgression patterns observed in both natural and experimental *I. fulva* and *I. hexagona* sympatric populations - with *I. fulva* alleles largely being shown to introgress into *I. hexagona* genomic backgrounds more readily than *I. hexagona* alleles introgressing into *I. fulva* (Arnold et al. 2010). When experimental hybrids are placed in sympatric *I. fulva* × *I. hexagona* populations, backcrossing towards *I. hexagona* - presumably mediated by bumblebees - occurs at a nearly ten-fold higher rate than it does towards *I. fulva* (Arnold et al. 2010; Emms and Arnold 2000). Repeated backcrossing towards *I. hexagona* can result in the incorporation of *I. fulva* alleles (even neutral alleles) into a largely *I. hexagona* background, and this has been observed in natural sympatric areas (Arnold et al. 1990b). Thus, SEPS and ANL may be important components of floral morphology where *I. hexagona*-like phenotypes are preferred by bumblebees. A note of caution should accompany these interpretations, however, as 70 different association-tests were performed (Table 7), and only four significant results were found (at $P < 0.05$). Assuming these tests are independent, 3.5 tests would have been expected to be significant at $P < 0.05$. Given these tests are certainly not independent (e.g., traits are correlated), no attempts to correct for multiple comparisons were performed (e.g., Bonferroni corrections) and interpretations are made from uncorrected P-values are reported.

Significant associations between pi-MASS top SNPs and BGC outliers were found as evidence of *I. hexagona* alleles preference by the pollinators that SEPS have significant association with positive α. And the genomic architecture of several traits that have significant associations with pollinators fit the expectation to be QTL traits responsible for reproductive isolation. NGA is an influential trait to attract specific pollinators (i.e., bumble bees) by their yellow triangular areas in Louisiana iris, and several studies have shown that the bumblebee pollinator syndrome is important for gene flow to speciation (Viosca 1935; Wesselingh and Arnold 2000; Martin et al. 2005, 2006, 2008). In this study, about two third of the individuals were observed to have NGA, and this trait was detected to be controlled by several SNPs with small effects with two larger effect loci. STBL (stylar branch length) and STBW (stylar branch width) are symbols of the size of stylar, that the shapes of stylars can attract different pollinators to reach the pollen/nectar under stylars (i.e., a flat stylar is suitable for bumble bees to craw in, and a narrower stylar is easier for hummingbirds to sip through the tube-shape stylar.). ANEX represents the appearance of anther tip that exposed to the uncovered air, that this morphological trait directly influence the ease of pollen spreading or self-pollination, and is controlled by several small effect loci with four larger effect loci. PETTL (petal total length) and SEPTL (sepal total length) explained the major size of the flower by the length measurements, where the greater the flower sizes are, the more attractive by bumble bees (Emms and Arnold 2000; Wesselingh and Arnold 2000). SEPS provide unique spaces for different pollinators to approach the pollen/syrup as food (Wesselingh and Arnold 2000; Bouck et al. 2007), and are controlled by some SNPs with weaker phenotypic effects. Although some traits are not associated with the pollinator syndrome directly, due to the significant correlations between these traits that show continuous variation, there

are still associations for all traits to be connected to the pollinator isolation by their polygenic genomic architecture.

Characters that are correlated with each other might have been influenced by natural selection simultaneously, the phenotypic correlations between traits are ubiquitous and might be able to reflect the natural selection to adaptation and evolution (Lande and Arnold 1983; Martins and Garland 1991). Strong absolute values of correlations greater than 0.8 were generally observed among floral length measurements in this study. All 14 traits used in this study are well-correlated and have strong associations with genotypes, means that the loci detected to be associated with these traits might be the genomic regions responsible for gene flow to speciation, especially the floral morphologies that are usually related to pollinator syndromes more than other traits. Pollinator isolation is one of the very important and potential reproductive isolating barriers (Arnold 2000; Martin et al. 2007, 2008; Taylor et al. 2012b), and the traits morphologies are reasons to attract different pollinators which indirectly lead to pre-zygotic isolation.

As a similar result to Gompert et al. (2012a), the reason why the exact loci might not be significantly associated with each other from the two analyses could be due to the closely linked loci distances that were excluded from the pi-MASS model, therefore cause some PIP values changed. Programs incorporated with single SNP analyses might be able to solve this problem that pi-MASS analyses could lose some closely linked top SNPs.

V. CONCLUSION AND FUTURE WORK

Louisiana Iris is an important system for studying speciation, and several studies have been performed using laboratory crosses, though studies of natural populations are less common (Martin et al. 2007). Here, the genomic architecture of reproductive isolation and adaptive introgression for Louisiana Iris demonstrated that the ecological and floral traits can effect reproductive isolation at the genomic level. In this study, the phenotypes that are associated with prezygotic isolation, especially pollinator isolation, are composed of several SNPs of small to moderate effects and revealed a complex and polygenic genomic architecture. Pollinator isolation is one of the most important barriers to gene flow among *I. fulva* and *I. hexagona,* and this study found that regions of the genome that were strongly associated with producing nectar guides in *I. hexagona* were favored to introgress across species boundaries. Genomics studies within hybridizing populations are both interesting and informative for understand the genomic architecture of reproductive isolation and speciation in Louisiana Iris.. Future work examining locus-specific estimates of genetic differentiation (Fst) between *I. fulva* and *I. hexagona* will be further examined in allopatric populations of both species in order to increase our understanding of the degree to which selection in hybrid zones is informative across a broader geographic context.

TABLES

Table 1. Abbreviations and descriptions for all 14 phenotypes measured. See Figure 1 for floral images.

Trait	Abbreviation	Description
Nectar Guide Area	NGA	The measure of the roughly triangular area of yellow nectar guide (when present). It was calculated as one half the length times the width of the yellow area.
Anther Length	ANL	The length of anther.
Stylar Branch Length	STBL	Measured from the base of calyx to the tip of a stylar branch.
Anther Extension	ANEX	Calculated as the length of stamen minus STBL.
Stylar Branch Width	STBW	Measured from the widest horizontal location of the stylar branch.
Petal Total Length	PETTL	Measured from the base of calyx to the tip of the petal.
Petal Width	PETW	Measured from the widest horizontal location of the petal.
Sepal Total Length	SEPTL	The summation of sepal stalk length and sepal blade length. Sepal stalk length was measured from the base of calyx to the neck of the sepal. Sepal blade length was measured from the neck to the tip of a sepal.
Sepal Width	SEPW	Measured from the widest horizontal location along the sepal.
Sepal Shape	SEPS	The ratio of sepal stalk length to sepal blade length.
Stalk Height	STALK	Measured from the base of the flowering stalk that connected to the rhizome to the base of lowest flower calyx.
Leaf Height	LEAF	Measured from the base of longest leaf without a flower stalk to its terminus.
Relative Height	RLTH	The ratio of STALK to LEAF.
Canopy Open Space	CNP	CNP is the only environmental measurement in this study. Calculated using a densiometer. (higher values equal lower canopy cover)

Table 2. Means of parameter estimates with 95% CIs (equal-tail probability intervals, given in parentheses) for proportion of phenotypic variance explained (PVE), the conditional prior probability of a SNP being in the model (PSNP), the mean number SNPs included in the model (NSNP), and the average effect of a SNP on the phenotype (σAVE). Trait abbreviations are given in Table 1.

Trait	PVE	P_{SNP}	N_{SNP}	σ_{AVE}
NGA	0.8047617 (0.617 – 0.896)	0.001251031 (0.000398084 – 0.002301442)	62.0809 (19 – 98)	0.5604149 (0.336 – 0.831)
ANL	0.8239962 (0.676 – 0.897)	0.001614180 (0.000737904 – 0.002426610)	76.1696 (36 – 100)	0.5236584 (0.330 – 0.737)
STBL	0.8616358 (0.766 – 0.899)	0.001647445 (0.000870964 – 0.002404363)	77.2641 (42 – 100)	0.5924448 (0.409 – 0.818)
ANEX	0.8601266 (0.752 – 0.899)	0.001199555 (0.000549541 – 0.002187762)	57.9874 (28 – 95)	0.6848505 (0.441 – 0.952)
STBW	0.6815612 (0.466 – 0.849)	0.001208309 (0.000341154 – 0.002312065)	60.9641 (17 – 98)	0.4133489 (0.233 – 0.742)
PETTL	0.8491075 (0.709 – 0.899)	0.001435255 (0.000647143 – 0.002328091)	68.6326 (32 – 99)	0.6040688 (0.399 – 0.843)
PETW	0.5678249 (0.209 – 0.853)	0.001000177 (0.000110656 – 0.002312065)	56.1416 (5 – 99)	0.3741356 (0.160 – 0.800)
SEPTL	0.8360089 (0.690 – 0.898)	0.001615128 (0.000770903 – 0.002426610)	76.2589 (37 – 100)	0.5467708 (0.349 – 0.778)
SEPW	0.7252846 (0.464 – 0.885)	0.001501941 (0.000532078 – 0.002393316)	72.6214 (26 – 99)	0.4193031 (0.235 – 0.652)
SEPS	0.7317385 (0.519 – 0.886)	0.001392160 (0.000481948 – 0.002371374)	68.0229 (23 – 99)	0.4420624 (0.252 – 0.843)
STALK	0.7392934 (0.490 – 0.886)	0.001467044 (0.000481920 – 0.002398833)	71.3497 (23 – 99)	0.4390998 (0.248 – 0.678)
LEAF	0.7781419 (0.589 – 0.890)	0.001727178 (0.000860994 – 0.002449063)	80.9265 (41 – 100)	0.4411075 (0.277 – 0.627)
RLTH	0.5004557 (0.119 – 0.815)	0.000695080 (0.000052000 – 0.002259436)	44.8871 (3 – 97)	0.4114431 (0.126 – 1.158)
CNP	0.7236324 (0.447 – 0.882)	0.001231975 (0.000303389 – 0.002338837)	62.7411 (15 – 99)	0.4657737 (0.270 – 0.746)

Table 3. Estimates of the posterior inclusion probabilities (PIP) and the absolute values for the magnitude of the phenotypic effect (|β|) for the first three SNPs identified with the highest PIPs for each trait.

Trait	PIP_1	$\|\beta_1\|$	PIP_2	$\|\beta_2\|$	PIP_3	$\|\beta_3\|$
NGA	0.8030	0.2797	0.6608	0.2791	0.3948	0.1729
ANL	0.4205	0.1155	0.3785	0.1472	0.3707	0.1598
STBL	0.4083	0.1060	0.4025	0.1106	0.3009	0.0676
ANEX	0.9950	0.4283	0.9531	0.3732	0.8471	0.4138
STBW	0.8314	0.2417	0.2942	0.1637	0.2127	0.1131
PETTL	0.9995	0.4243	0.6609	0.2960	0.6509	0.1940
PETW	0.2316	0.1271	0.1415	0.0760	0.1152	0.0523
SEPTL	0.4669	0.1686	0.3937	0.1414	0.2637	0.1190
SEPW	0.4346	0.1699	0.2134	0.0778	0.2125	0.0955
SEPS	0.2750	0.0767	0.2691	0.1020	0.2410	0.0737
STALK	0.2744	0.1176	0.2700	0.1675	0.2282	0.0833
LEAF	0.5432	0.2478	0.4008	0.1835	0.3365	0.1317
RLTH	0.4114	0.1944	0.1992	0.1314	0.1335	0.0933
CNP	0.4981	0.2919	0.4292	0.1763	0.3583	0.1344

Table 4. Identification of shared top SNPs between traits. Below diagonal – the numbers of shared top SNPs between the two traits (blank cells representing no shared top SNPs). Above diagonal – the probability of getting N or more numbers of shared top SNPs between the two traits by chance ($P < 0.05$ are bold).

Traits	NGA	ANL	STBL	ANEX	STBW	PETTL	PETW	SEPTL	SEPW	SEPS	STALK	LEAF	RLTH	CNP
NGA									0.0965					
ANL			0.1242			0.1106			**0.0003**			**0.0004**		
STBL		1		**9.41e^{-14}**	0.0997	**0.0002**		**9.93e^{-6}**						
ANEX			8					0.0938						**0.0030**
STBW			1				0.0738	**0.0049**	0.0936					
PETTL		1	3				**9.28e^{-5}**	**1.11e^{-16}**	**0.0002**		0.1038			
PETW					1	3		**0.0001**	**0.0000**	0.0831	0.0866	0.0969		
SEPTL			4	1	2	15	3		**4.56e^{-11}**		**0.0067**	0.1286		
SEPW	1	3			1	3	10	7			0.1095			
SEPS							1							
STALK						1	1	2	1			**2.04e^{-7}**	**0.0024**	0.0952
LEAF		3					1	1			5		**0.0030**	**0.0057**
RLTH											2	2		**0.0018**
CNP				2							1	2	2	

Table 5. Top SNPs (with the highest PIPs) shared by more than two traits are denoted by asterisks. SNP identification numbers are given in the first column. Trait abbreviations are listed in Table 1.

	NGA	ANL	STBL	ANEX	STBW	PETTL	PETW	SEPTL	SEPW	SEPS	STALK	LEAF	RLTH	CNP
952			*			*		*						
1475						*		*	*					
9562						*		*			*			
11462						*	*	*						
19160						*	*	*						
24190			*	*				*						
24229			*		*			*						
28746							*	*	*					
31243						*		*	*					
35796											*		*	*
36350						*		*	*					
43840							*				*	*		

Table 6. Estimates of the posterior inclusion probabilities (PIPs) and the absolute values of the magnitude of the phenotypic effect (|β|) for the top SNPs shared by more than two traits. Trait abbreviations are given in Table 1.

Trait	SNP	PIP	|β|	Trait	SNP	PIP	|β|
STBL	952	0.0485	0.0125	STBL	24229	0.0794	0.0277
PETTL	952	0.1599	0.0639	STBW	24229	0.2942	0.1637
SEPTL	952	0.0924	0.0323	SEPTL	24229	0.0751	0.0287
PETTL	1475	0.0493	0.0135	PETW	28746	0.0263	0.0105
SEPTL	1475	0.0837	0.0244	SEPTL	28746	0.0658	0.0182
SEPW	1475	0.0360	0.0120	SEPW	28746	0.0303	0.0110
PETTL	9562	0.0468	0.0112	PETTL	31243	0.5606	0.1891
SEPTL	9562	0.0331	0.0086	SEPTL	31243	0.4669	0.1686
STALK	9562	0.0332	0.0122	SEPW	31243	0.0637	0.0232
PETTL	11462	0.1865	0.0603	STALK	35796	0.0323	0.0106
PETW	11462	0.0437	0.0184	RLTH	35796	0.0174	0.0054
SEPTL	11462	0.1335	0.0407	CNP	35796	0.0327	0.0101
PETTL	19160	0.0625	0.0210	PETTL	36350	0.0422	0.0112
PETW	19160	0.0266	0.0118	SEPTL	36350	0.1647	0.0547
SEPTL	19160	0.1932	0.0774	SEPW	36350	0.0544	0.0202
STBL	24190	0.1664	0.0417	PETW	43840	0.0483	0.0171
ANEX	24190	0.3392	0.1203	STALK	43840	0.0494	0.0167
SEPTL	24190	0.1445	0.0468	LEAF	43840	0.0444	0.0123

Table 7. Associations between top SNPs (from pi-MASS) and both the genomic cline parameters α and β (compared separately) were assessed by estimating the probabilities of getting the number of matches or more for SNPs that were identified as significant in the BGC analysis and the association mapping study by chance (significant ones with $P < 0.05$ are bold, while the blank ones represent no matches).

Trait	α outliers	+α outliers	-α outliers	β outliers	+β outliers	-β outliers	α and β	α or β
NGA	0.6227	0.2934	0.9007	0.6453	0.5583		0.6629	0.6319
ANL	0.1803	**0.0375**	0.8504	0.2761	0.1926		0.3801	0.1454
STBL	0.7411	0.6120	0.7585	0.7564	0.6742		0.7398	0.7665
ANEX	0.7799	0.8123	0.5813	0.3247	0.2445		0.2612	0.7775
STBW	0.5596	0.3677	0.7770	0.6278	0.5408			0.4568
PETTL	0.7816	0.6719	0.7602	0.2158	0.1466		**0.0306**	0.8619
PETW	0.6530	0.7966	0.4030	0.3152	0.5044	0.2516		0.4266
SEPTL	0.1803	0.2553	0.3466	0.1278	0.0773		**0.0431**	0.2070
SEPW	0.2311	0.2795	0.4096	0.4647	0.3655		0.3555	0.2562
SEPS	**0.0194**	**0.0055**	0.6258	0.9119	0.8736		0.6961	**0.0408**
STALK	0.4793	0.1745	0.8914	**0.0383**	0.0616	0.3066	0.7114	0.1622
LEAF	0.1983	0.3282	0.2909	0.4120	0.6942		0.7530	0.2306
RLTH	0.8235	0.6864	0.8207					0.8808
CNP	0.8099	0.8774	0.3677	0.8912	0.8486			0.7106

FIGURES

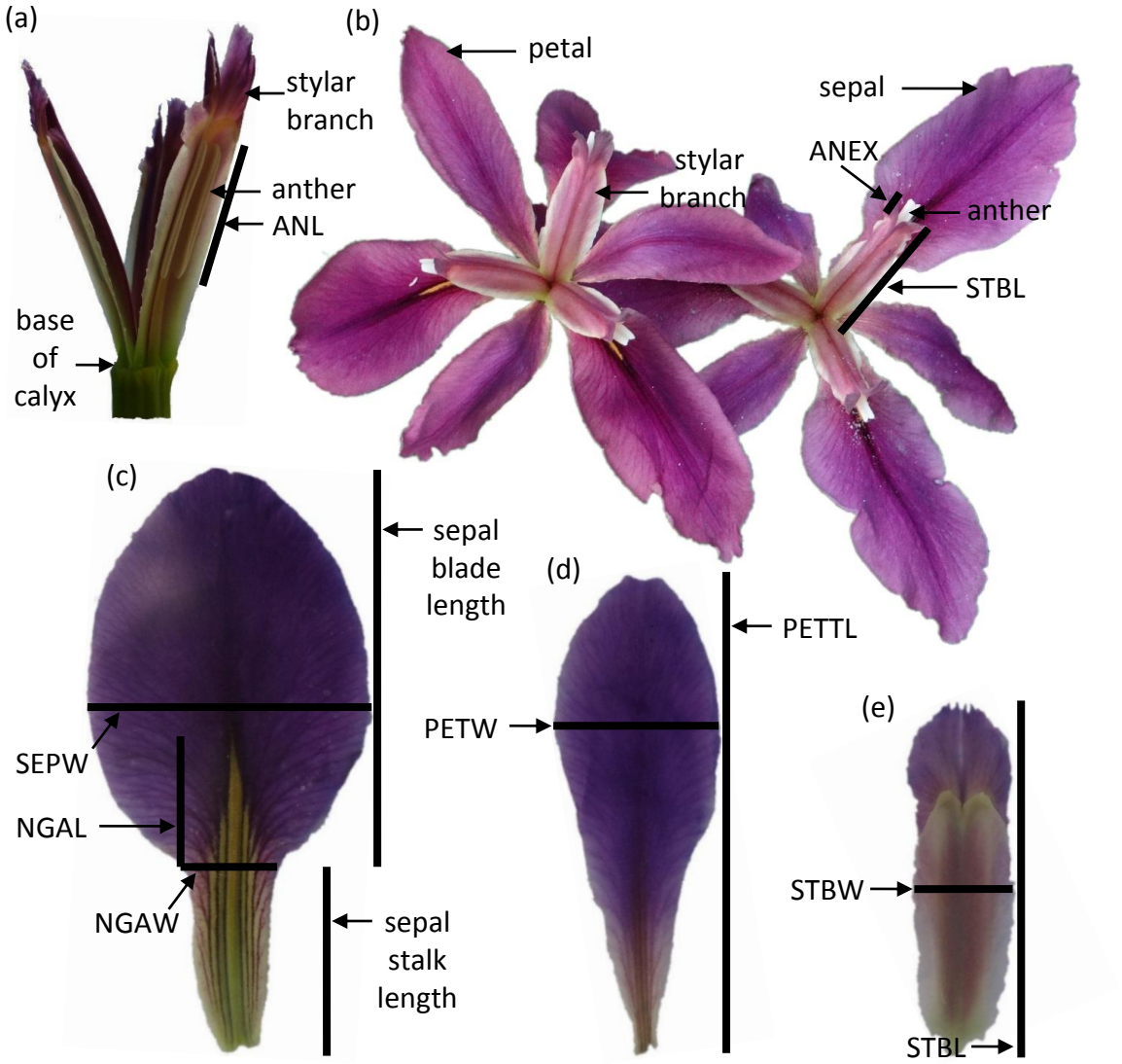

Figure 1. Flower measurements of *I. hexagona* X *I. fulva* hybrids: (a) lateral view of a hybrid flower with sepals and petals removed; (b) an aerial view of hybrid flowers; (c) a hybrid sepal - NGA was measured by estimating the yellow triangular area as 1/2 NGAL multiplied by NGAW; (d) a hybrid petal; and (e) a hybrid stylar branch. See Table 1 for description of all 14 traits measured.

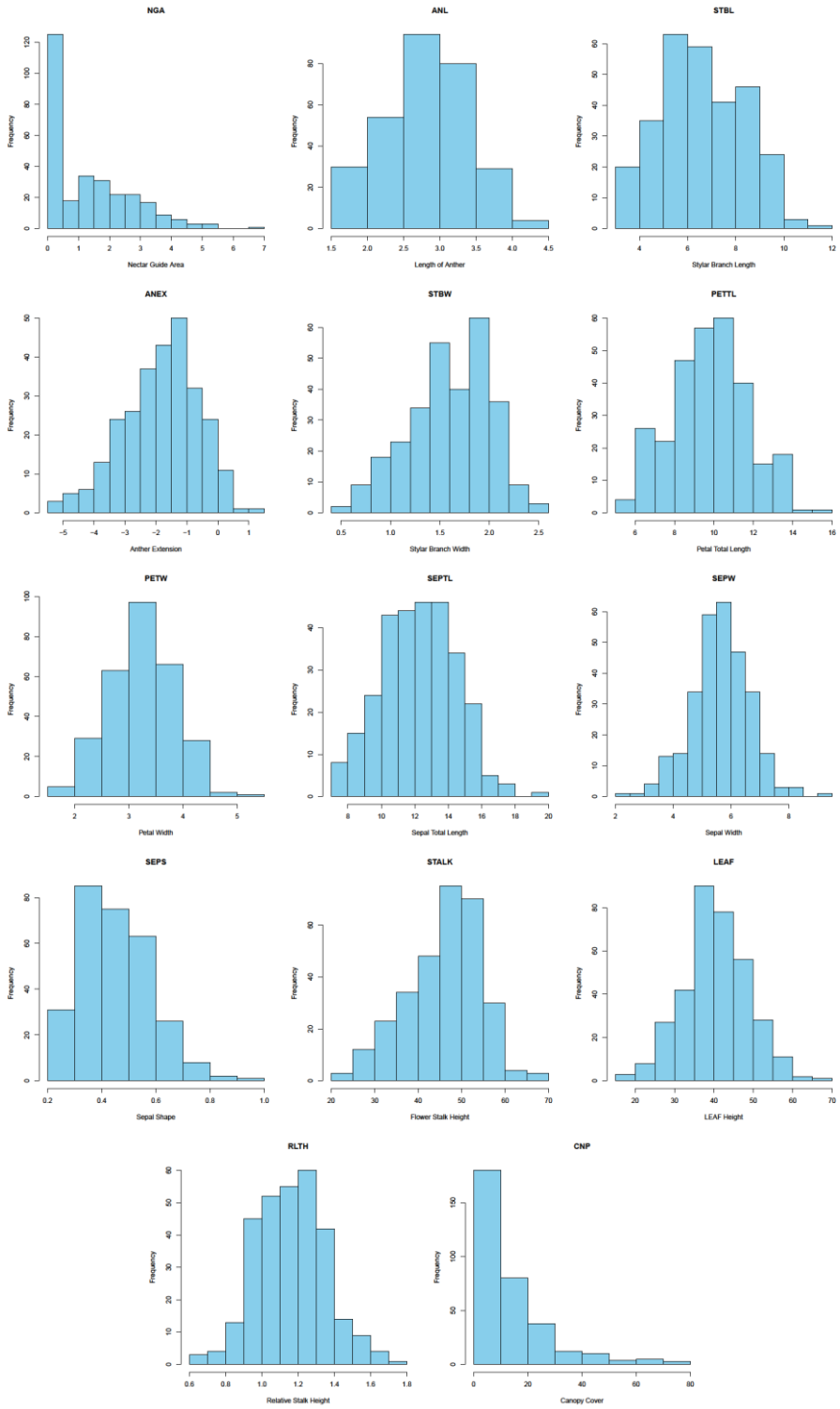

Figure 2. Histograms showing variation in each of the 14 traits measured in the Lake Martin hybrid zone. See Table 1 for descriptions and abbreviations for each trait.

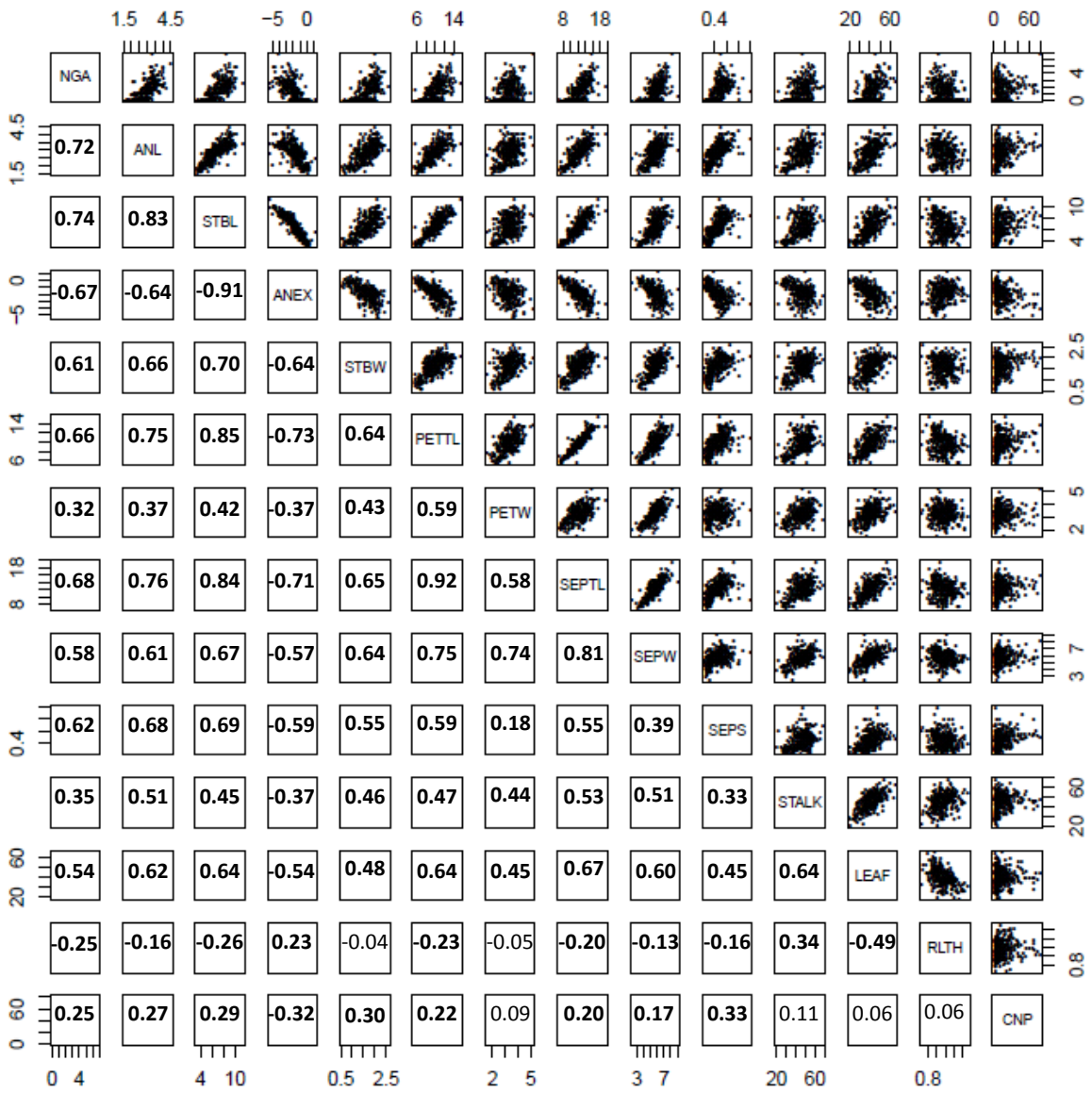

Figure 3. Correlations among traits: scatter plots of the relationship between trait-pairs above the diagonal, trait abbreviations (See Table 1) along the diagonal, and correlation coefficients (r) below the diagonal. Significant coefficients ($P < 0.05$) are listed in bold. All correlation coefficients (with the exception of SEPW x RLTH, $P = 0.02$) are significant with or without Bonferroni correction for multiple tests ($P < 0.0059$).

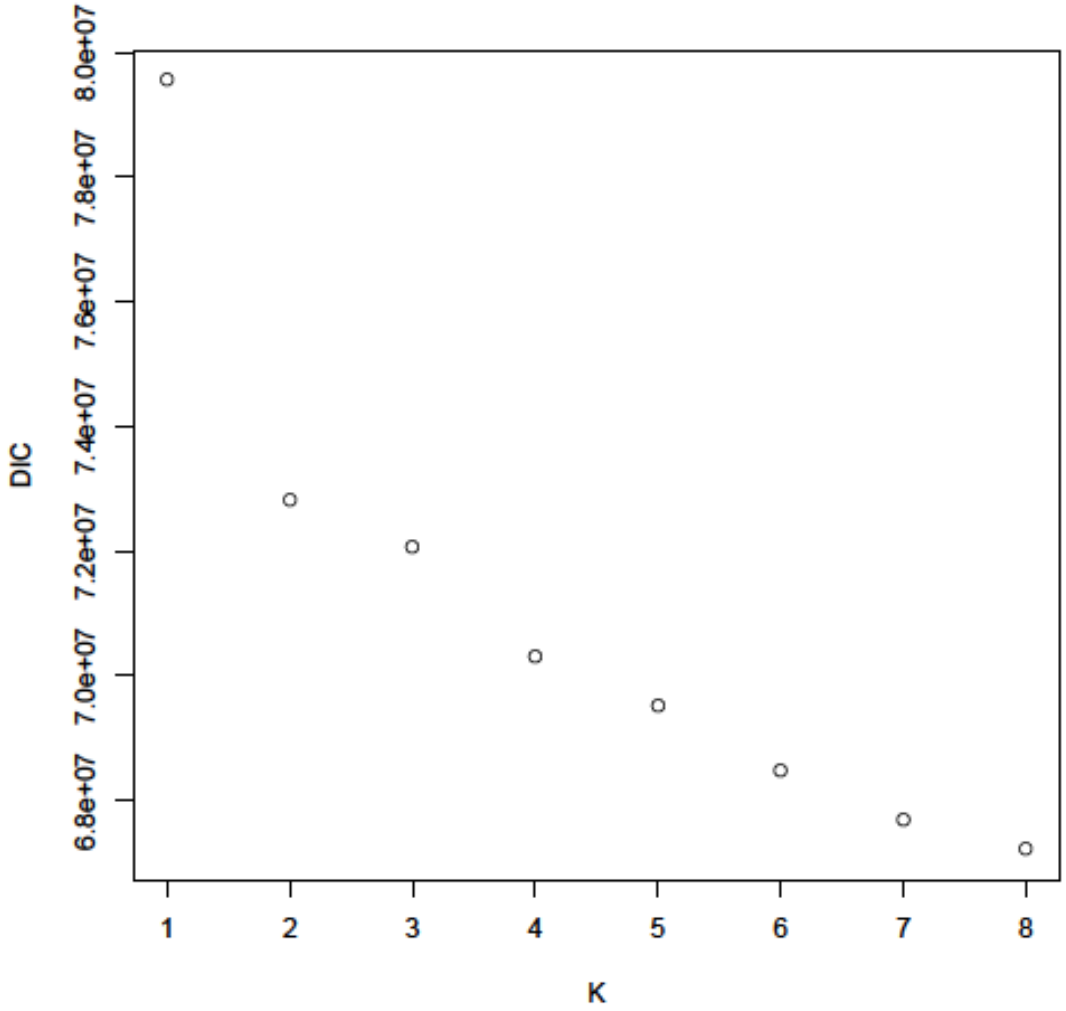

Figure 4. DIC distribution of K = 1 to K = 8 (d = 0.5).

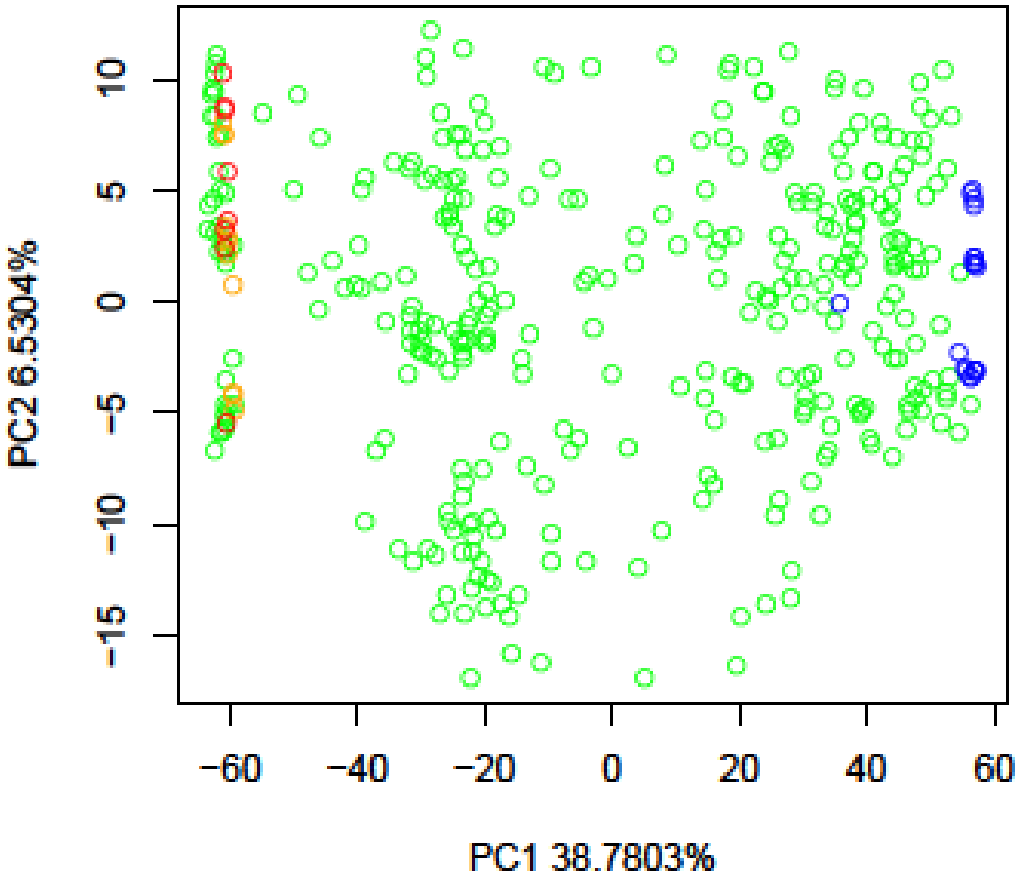

Figure 5. Visualization of genetic structure in a Louisiana Iris hybrid zone using Principal Component Analysis (PCA; K=2). The X axis (PC1) explains 38.7803% and the Y axis (PC2) explains 6.5304% of the variation in genotype estimates. The red circles represent the 8 allopatric *I. fulva* individuals sampled from Lottie, LA. The orange circles represent the 11 allopatric *I. fulva* individuals sampled from Livonia, LA, and the blue circles represent the 19 allopatric *I. hexagona* sampled from Abbeville, LA. The green circles are the 346 hybrid individuals sampled and genotyped from the Lake Martin hybrid zone.

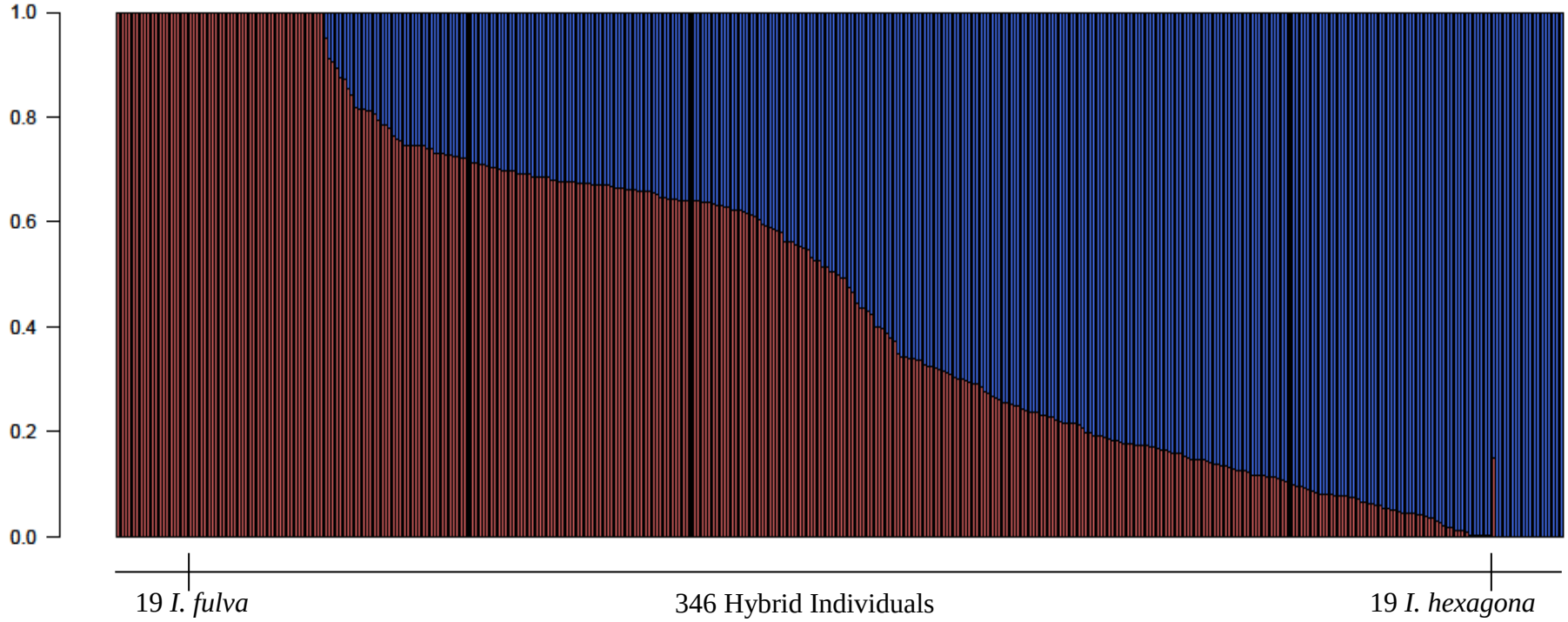

Figure 6. ENTROPY results for K=2 showing the estimated admixture proportions for all 384 individuals sampled for this study. Individuals at the far left of the figure represent the 19 allopatric *I. fulva* individuals, while individuals at far right denote allopatric *I. hexagona* individuals (N = 19). The remaining 346 hybrid individuals are sorted by estimated admixture proportion with the most "*I. fulva*-like" individuals on the left and the most *I. hexagona*-like individuals on the right.

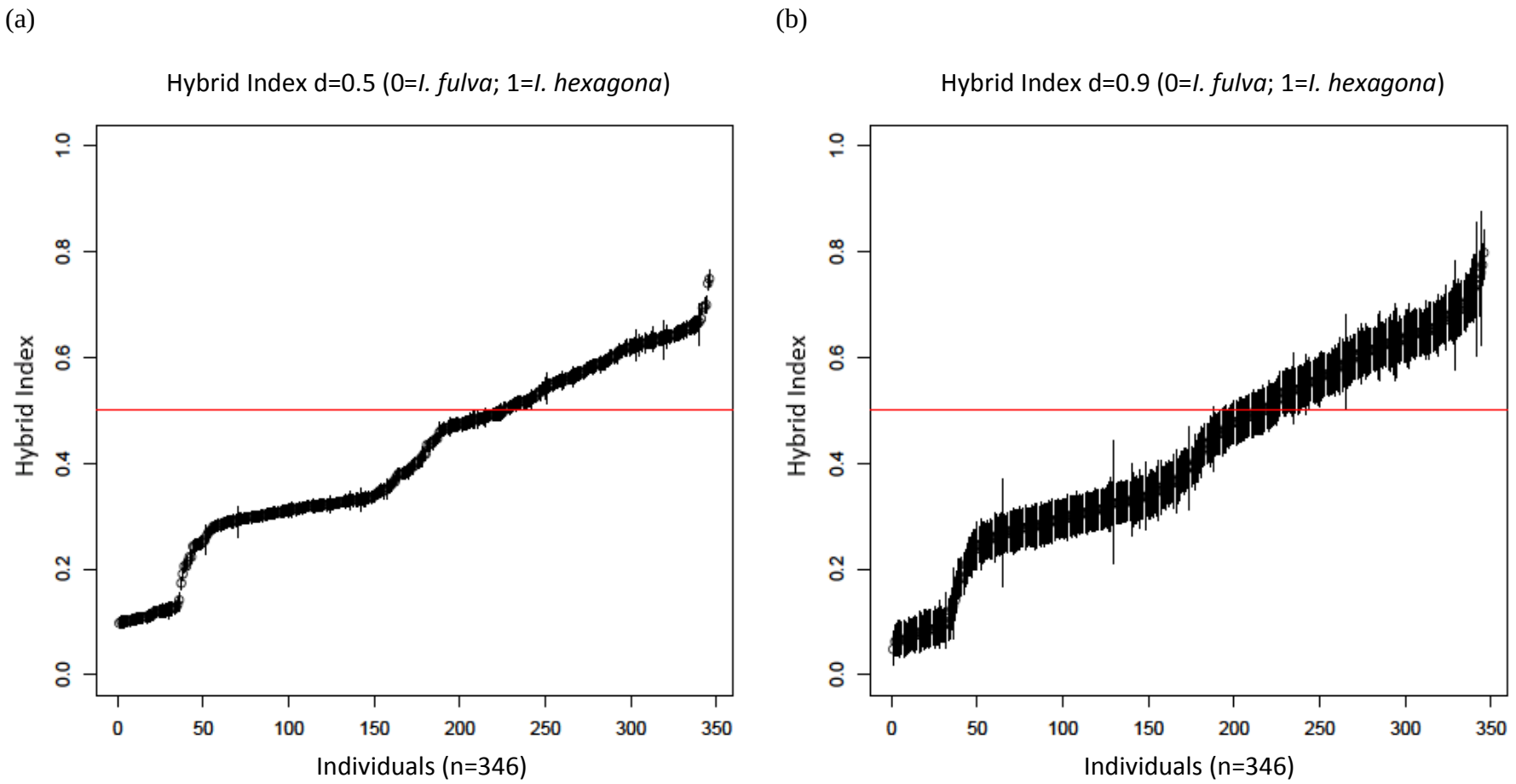

Figure 7. Estimated median hybrid index (and 95% CI) for all 346 hybrids. Hybrid index ranges from 0 to 1, and indicates the proportion of the genome that is comprised of *I. hexagona*. A. Hybrid indexes estimated using 3,699 high coverage SNPs (d = 0.9). B. Hybrid indexes estimated using 45,384 loci with lower coverage (d = 0.5). Note the smaller CIs observed in B.

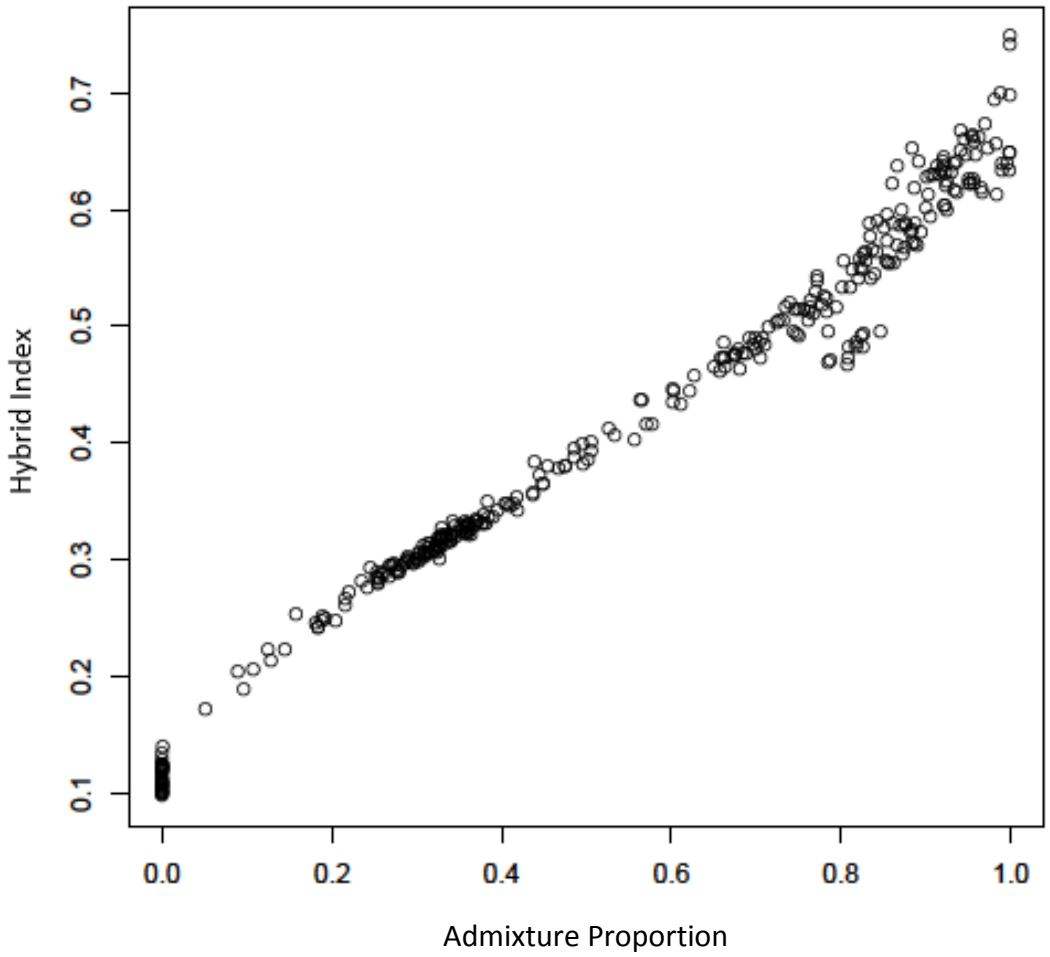

Figure 8. Plot of the admixture proportion of *I. hexagona* ancestry from ENTROPY and the hybrid indexes from BGC of 346 hybrid individuals.

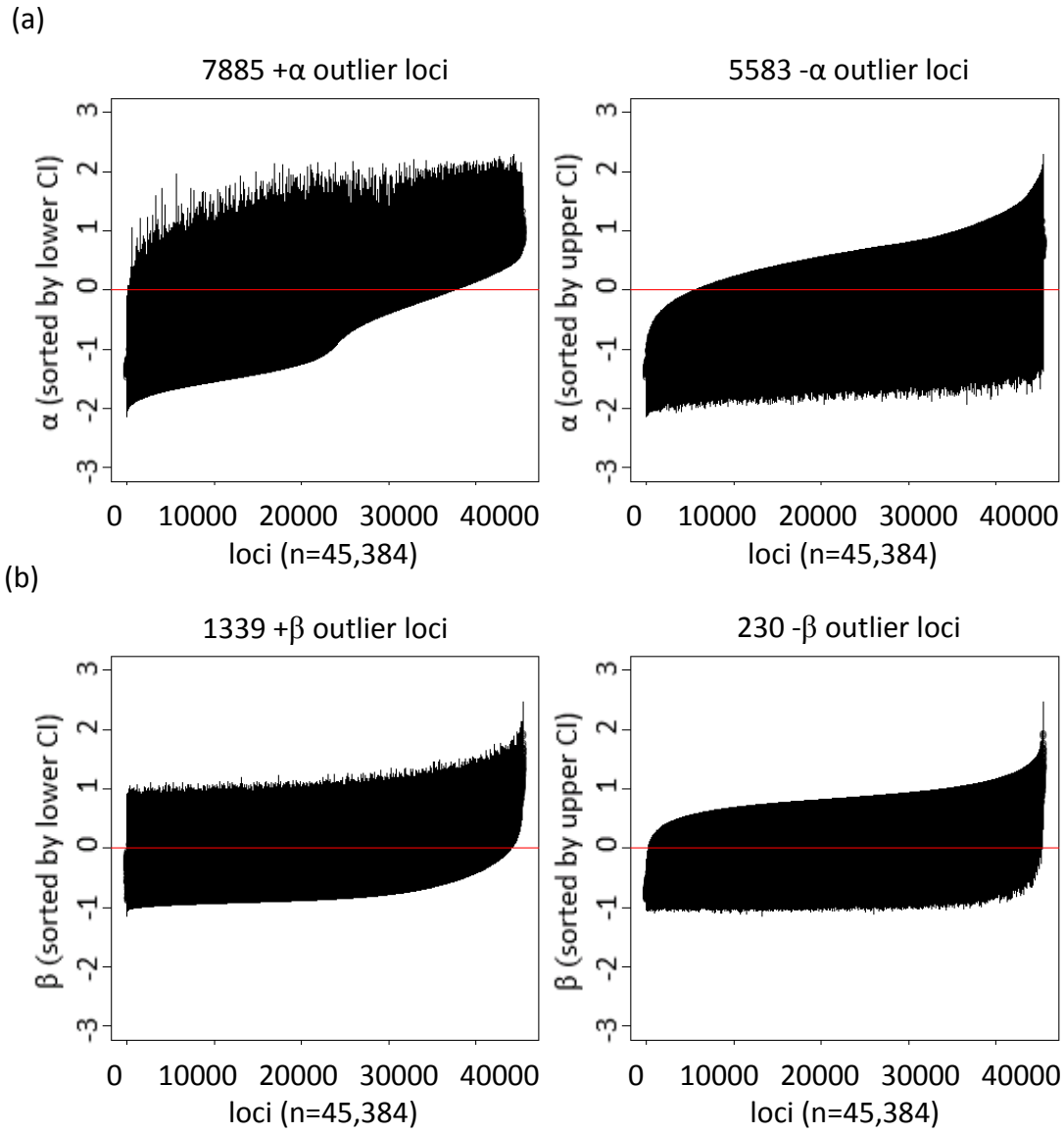

Figure 9. Median (+/- 95% CIs) of BGC cline parameters, α (a) and β (b), for all 45,384 loci. (a-left) Loci were sorted by the lower bound of 95% CIs of α. 7,885 positive outliers which the CIs didn't intersect were identified at the right-hand side. (a-right) Loci were sorted by the upper bound of 95% CIs of α, and 5,583 negative outliers which the CIs didn't intersect zero were identified at the left-hand side. (b-left) Loci were sorted by the lower bound of 95% CIs of β. 1,339 positive outliers where CIs did not intersect zero were identified at the right-hand side. (b-right) Loci were sorted by the upper bound of 95% CIs of β. 230 negative outliers where the CIs did not intersect zero were identified at the left-hand side.

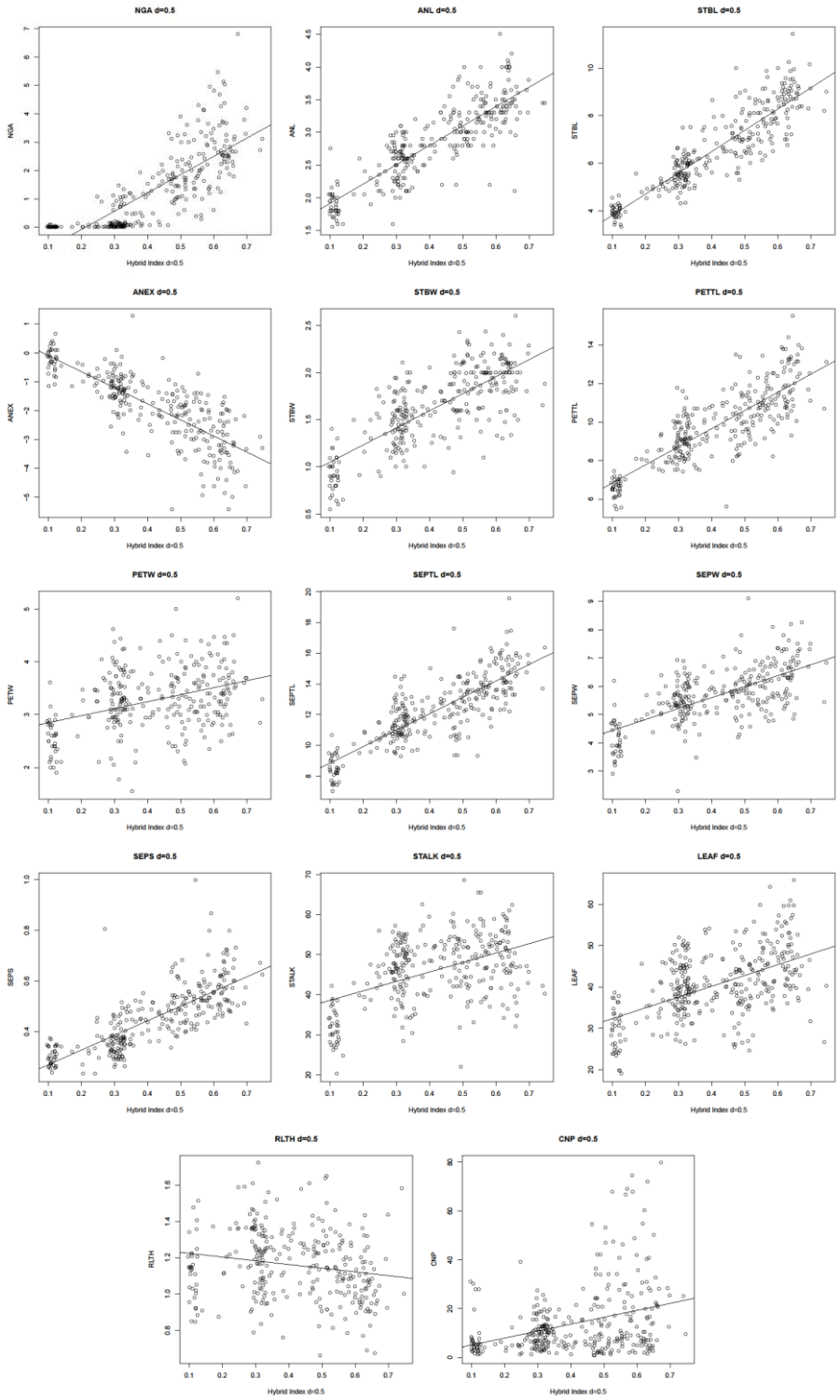

Figure 10. Linear regressions examining the relationship of hybrid index (X-axis) and phenotype (Y-axis) for each of 14 traits measured in the 346 hybrids.

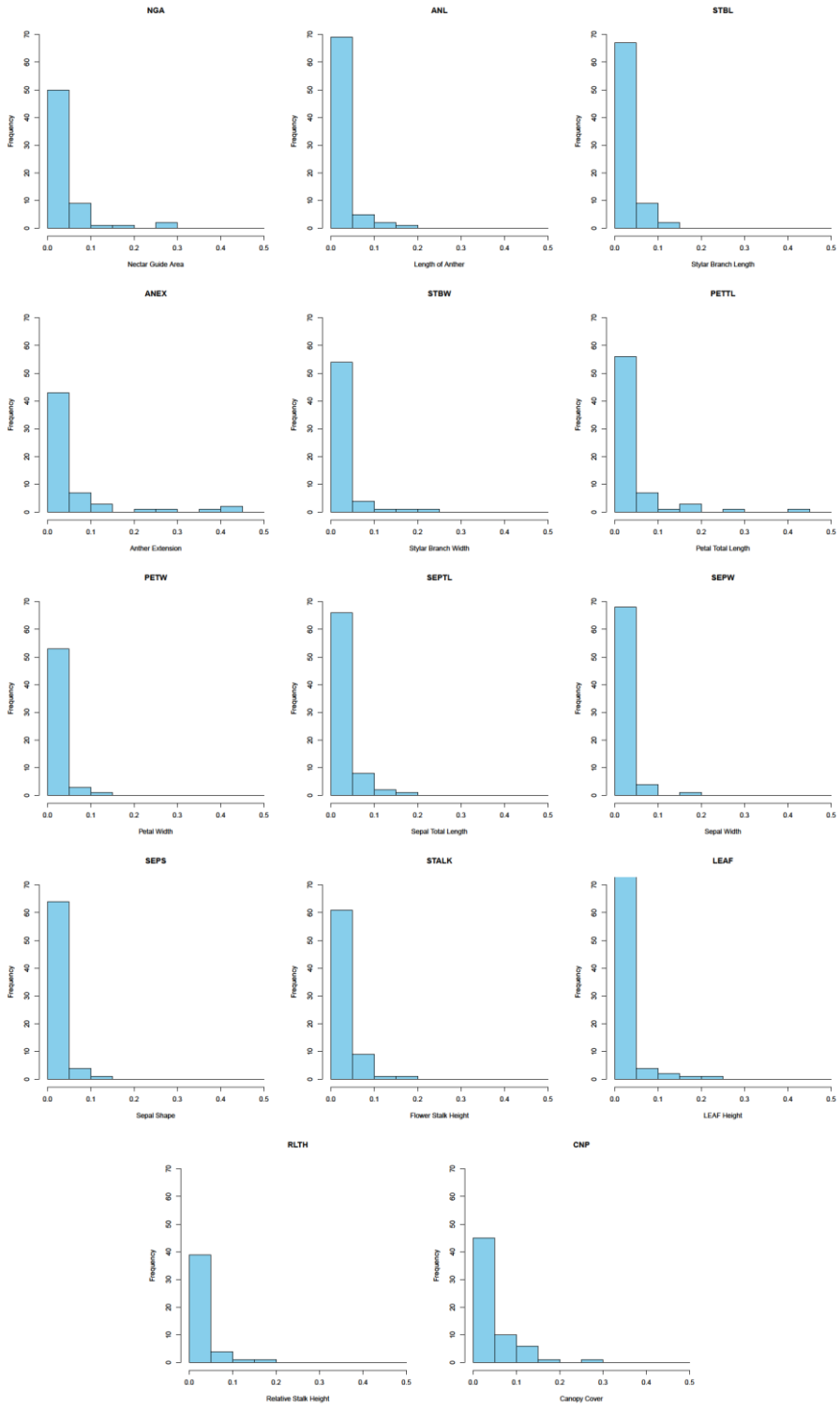

Figure 11. Histograms depicting phenotypic effect sizes (β – in standard deviations) of top SNPs included in individual models. N represent the number of top SNPs included in each model.

REFERENCES

Arnold, M. L., Bennett, B. D. and Zimmer, E. A. 1990 A. Natural hybridization between *Iris fulva* and *Iris hexagona*: pattern of ribosomal DNA variation. *Evolution* **44**: 1512-1521

Arnold, M. L., Hamrick, J. L. and Bennett, B. D. 1990 B. Allozyme Variation in Louisiana Irises: A Test for Introgression and Hybrid Speciation. *Heredity* **65**: 297-306

Arnold, M. L. 1993. Iris nelsonii (Iridaceae): Origin and Genetic Composition of a Homoploid Hybrid Species. *American Journey of Botany* **80-5**: 577-583

Arnold, M. L. 1997. *Natural Hybridization and Evolution.* Oxford University Press, New York

Arnold, M. L. 2000. Anderson's paradigm: Louisiana Irises and the study of evolutionary phenomena. *Mol Ecol* **9**: 1687-1698

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

Arnold, M. L. and Meyer, A. 2006. Natural hybridization in primates: One evolutionary mechanism. *Zoology* **109**: 261-276

Arnold, M. L. and Martin, N. H. 2009. Adaptation by introgression. *Journey of Biology* **8**: 82

Arnold, M. L. and Martin N. H. 2010. Hybrid Fitness across Time and Habitats. *Trends in Ecology and Evolution* **25**: 530-536

Arnold, M. L. and Tang S., Knapp S. J. and Martin N. H. 2010. Asymmetric Introgressive Hybridization among Louisiana Iris Species. *Genes* **1**: 9-22

Ballerini, E.S., Brothers A. N., Tang S., Knapp S. J., Bouck A., Taylor S. J., Arnold M. L. and Martin N. H. 2012. QTL mapping reveals the genetic architecture of loci affecting pre- and post-zygotic isolating barriers in Louisiana Iris. *BMC Plant Bio* **12**: 91

Blount, Z. D., Borland C. Z. and Lenski R. E. 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. PNAS 105-23: 7899-7906

Bouck, A., Wessler S. R. and Arnold M. L. 2007. QTL Analysis of Floral Traits in Louisiana Iris Hybrids. *Evolution* 61-10:2308-2319

Brothers, A. N., Barb J. G., Ballerini, E. S., Drury, D. W., Knapp, S. J., and Arnold, M. L. 2013. Genetic Architecture of Floral Traits in *Iris hexagona* and *Iris fulva*. *Journal of Heredity* 104(6): 853-861

Buerkle, C. A. and Lexer, C. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution* **23**: 686-694

Burke, J. M., Carney, S. E. and Arnold, M. L. 1998. Hybrid fitness in the Louisiana irises: Analysis of parental and F-1 performance. *Evolution* **52**: 37-43

Comeault, Soria-Carrasco, Gompert, Farkas, Buerkle, Parchman and Nosil. 2014. Genome-Wide Association Mapping of Phenotypic Traits Subject to a Range of Intensities of Natural Selection in *Timema cristinae*. The American Naturalist 183:711-727

Coyne, J. A. and Orr, H. A. 1989. Patterns of speciation in Drosophila. *Evolution* **43(2)**: 362-381

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

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

Diaz, A. and Macnair, M. R. 1999. Pollen tube competition as a mechanism of prezygotic reproductive isolation between *Mimulus nasutus* and its presumed progenitor *M. guttatus*. *New Phytol.* **144**: 471-478

Doyle, J. 1991. DNA protocols for plants: a CTAB total DNA isolation. In: *Molecular Techniques in Taxonomy* (eds Hewitt GM, Johnson A), pp. 283-293. Springer, New York

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

Gompert, Z. and Buerkle, C. A. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18**: 1207-1224

Gompert, Z. and Buerkle, C. A. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**: 378-384

Gompert, Z. and Buerkle, C. A. 2011. Bayesian estimation of genomic clines. *Molecular Ecology* **20**: 2111-2127

Gompert, Lucas, Nice and Buerkle 2012a Genome Divergence and the genetic architecture of barriers to gene flow between L Idas and L Melissa Evolution 67-9: 2498-2514

Gompert, Lucas, Nice, Fordyce, Forister and Buerkle. 2012b. Genomic Regions with A History of Divergent Selection Affect Fitness of Hybrids between Two Butterfly Species. Evolution 66-7: 2167-2181

Gompert, Lucas, Buerkle, Forister, Fordyce and Nice. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Molecular Ecology 23: 4555-4573

Guan, Y., and M. Stephens. 2011. Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. Annals of Applied Statistics 5:1780-1815

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

Lowry, D. B., Rockwood, R. C. and Willis, J. H. 2008. Ecological reproductive isolation of coast and inland races of *Mimulus Guttatus*. *Evolution* **62-9**: 2196-2214

Mandeville, E. G., Parchman, T. L., McDonald, D. B. and Buerkle C. A. 2015. Highly variable reproductive isolation among pairs of catostomus species. *Mol Ecol* **24**: 1856-1872

Martin, N. H., Bouck, A. C. and Arnold, M. L. 2005. Loci affecting long-term hybrid survivorship in Louisiana irises: implications for reproductive isolation and introgression. 2005. *Evolution* **59(10)**: 2116-2124

Martin, N. H., Bouck, A. C. and Arnold, M. L. 2006. Detecting Adaptive Trait Introgression between *Iris fulva* and *I. brevicaulis* in Highly Selective Field Conditions. *Genetics* **172**: 2481-2489

Martin, N. H., Bouck, A. C. and Arnold, M. L. 2007. The Genetic Architecture of Reproductive Isolation in Louisiana Irises: Flowering Phenology. 2007. *Genetics* **175**: 1803-1812

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

Mayer, M., Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448-

Moyle, L. C. and Nakazato T. 2008. Comparative Genetics of Hybrid Incompatibility: Sterility in Two Solanum Species Crosses. *Genetics* **179**: 1437-1453

Orr, H. A. 2001. The genetics of species differences. *Trends Ecol Evol* **16**: 343-350

Parchman, Gompert, Braun, Brumfield, McDonald, UY, Zhang, Jarvis, Schlinger and Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology 22: 3304-3317

Ramsey, J., Bradshaw, H. D. and Schemske, D. W. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520-1534

Randolph, L. F. 1966. *Iris nelsonii*, a new species of Louisiana Iris of hybrid origin. *Baileya* **14**: 143-169

Rieseberg, L. H., Carter, R. and Zona, S. 1990. Molecular tests of the hypothesis hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, (in press)

Rieseberg, L. H., Whitton, J. and Gardner, K. 1999. Hybrid Zones and the Genetic Architecture of a Barrier to Gene Flow between Two Sunflower Species. *Genetics* **152**: 713-727

Rieseberg, L. H. and Buerkle, C. A. 2002. Genetic Mapping in Hybrid Zones. *The American Naturalist* **159**: S36-50

Singhal S. and Moritz C. 2013. Reproductive isolation between phylogeographic lineages scales with divergence. *Proc R Soc B* **280**: 20132246

Slotman, M., Torre, A. D. and Powell, J. R. 2004. The genetics of inviability and male sterility in hybrids between Anopheles gambiae and An. arabiensis. *Genetics* **167**: 275-287

Smyth, D. R., Bowman J. L. and Meyerowitz, E. M. 1990. Early Flower Development in Arabidopsis. *The Plant Cell* **2**: 755-761

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

Taylor, S. J., Arnold, M. and Martin, N. H. 2009. The genetic architecture of reproductive isolation in Louisiana irises: hybrid fitness in nature. *Evolution* **63-10**: 2581-2594

Taylor, S. J., Willard R. W., Shaw J. P., Dobson M. C. and Martin N. H. 2011. Differential Response of The Homoploid Hybrid Species *Iris nelsonii* (Iridaceae) and Its Progenitors to Abiotic Habitat Conditions. *American Journal of Botany* **98(8)**: 1309-1316

Taylor, S. J., Rojas, L. D., Ho, S. W. and Martin, N. H. 2012 A. Genomic Collinearity and the Genetic Architecture of Floral Differences between the Homoploid Hybrid Species *Iris nelsonii* and One of Its Progenitors, *Iris hexagona*. *Heredity* 1-8

Taylor, S. J., AuBuchon K. J. and Martin, N. H. 2012 B. Identification of Floral Visitors of *Iris nelsonii*. *Southeastern Naturalist* **11(1)**: 141-144

Turelli M., Lipkowitz J. R. and Brandvain Y. 2013. On the Coyne and Orr-igin of species: effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on drosophila speciation. *Evolution* **12330**: 1-12

Viosca, P. 1935. The irises of southeastern Louisiana: a taxonomic and ecological interpretation. *Bull. Am. Iris Soc.* **57**: 3-56

Wang, H., McArthur, E. D., Sanderson, S. C., Graham, J. H. and Freeman, D. C. 1997. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentate*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* **51(1)**: 95-102

Wesselingh, R. A. and Arnold, M. L. 2000. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. *J. Evol. Biol.* **13**: 171-180