Neutral and Selective Processes Drive Population Differentiation for *Iris hexagona*

Jennafer A. P. Hamlin and Michael L. Arnold

From the Department of Genetics, University of Georgia, 120 East Green St., Davison Life Sciences Building, Athens, GA (Arnold) and the Department of Biology, Indiana University, Bloomington, IN 47405, USA (Hamlin).

Address correspondence to Jennafer A. P. Hamlin at the address above, or e-mail: jenhaml@iu.edu

Data deposited at Dryad: [http://dx.doi.org/doi:10.5061/dryad.4n15b](http://dx.doi.org/doi:10.5061/dryad.4n15b)

Received March 5, 2015; First decision April 21, 2015; Accepted June 11, 2015.

Corresponding editor: Christopher Smith

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 8 populations of *Iris hexagona*. Using a genotyping-by-sequencing approach, we generated a panel of 750 single nucleotide polymorphisms (SNPs) and used population genetic analyses to determine what may affect patterns of divergence across *I. hexagona* populations. Specifically, genetic differentiation was compared between populations at neutral and nonneutral SNPs and detected significant differences between the 2 types of markers. We then asked whether loci with the strongest degree of population genetic differentiation were also the loci with the strongest association to morphology or climate differences, allowing us to test if pollinators or climate drive population differentiation or some combination of both. We found 2 markers that were associated with morphology and 1 marker associated with 2 of the environmental variables, which were also identified in the outlier analysis. We then show that the SNPs putatively under selection were positively correlated with both geographic distance and phenotypic distance, albeit weakly to phenotypic distance. Moreover, neutral SNPs were only correlated with geographic distance and thus isolation-by-distance was observed for neutral SNPs. Our data suggest that both deterministic and neutral processes have contributed to the evolutionary trajectory of *I. hexagona* populations.

Subject areas: Population structure and phylogeography; Molecular adaptation and selection

Key words: environmental variation, local adaptation, Louisiana irises, morphological divergence, outlier SNPs, population genetics
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 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).

Genetic drift, which is also considered a neutral process and leads to random changes in allele frequencies, can render selection less efficient (Haag and Roze 2007). Furthermore, genetic drift is related to IBD in that the random change in population gene frequencies across both space and time is essentially an accumulation of local genetic differences under geographically restricted dispersal. If there is more differentiation in traits or gene frequencies between populations than expected 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). And one can evaluate evolutionary patterns at outlier vs. 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 nonmodel 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 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). Moreover, statistical association of alleles with a trait of interest can provide evidence of how species have adapted to their environment. For example, genome scans have been used for a wide variety of tree species to identify candidate gene regions associated with environmental variables or phenotypic differences (Eckert et al. 2010; Grivet et al. 2011; Tsumura et al. 2012; Alberto et al. 2013).

Many phenotypic and environmental 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). 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); however, demonstrating that floral divergence among conspecific wild populations is the outcome of variable selection can be difficult (Herrera and Bazaga 2008). 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 the strongest association to morphology or climate, allowing us to test if pollinators or climate drive population differentiation 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

---

**Figure 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.
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 1). The floral stem of *I. hexagona* attains heights of 1–2 m 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; Emmas and Arnold 2000; Van Zandt and Mopper 2004*). Previous work has suggested that *I. gigantea* could be considered 2 different species: *Iris gigantea* and *I. savannarum* located throughout Florida. This differentiation was based on variation in capsule (dried fruit) morphology (*Henderson 2002; Meerow et al. 2011*). Here, we predict that populations will be broadly morphologically structured, which will mirror the genetic structure identified. Furthermore, we seek a first approximation toward 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, Measurement of Phenotypic Traits, and Collection of Environmental Data**

During the spring of 2013, we collected samples throughout the range of *I. hexagona* (Figure 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 to 20. Individuals sequenced included only a subset of those that were collected from each population (Table 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, floral stem 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 was length of the tallest leaf and floral stem 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 3 floral traits (petal length, petal width, and nectar guide), all 3 floral units were measured and averaged for each genotype (*Iris* flowers are tripartite).

For understanding population-level environmental differences, we downloaded raster-formatted climate variables (19 Bioclimatic variables) at 30-arc s from the Worldclim database in R and data for 4 land cover layers derived from satellite-borne remote sensors (NASA-MODIS/Terra data set). Land cover layers included the normalized difference vegetation index (NDVI; measure of vegetative greenness), the yearly SD of NDVI (NDVISTD), the QSCAT (measures reflected microwave radiation and provides a measure of soil roughness and wetness), and the percentage of tree cover (TREE) (*Wooten and Gibbs 2012*). Within R, we extracted values for all 23 variables for each collection site used in this study.

**DNA Extraction, SNP Discovery, and Genotyping**

Leaf material was collected from individuals sampled for phenotypic traits with 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 1 blank was included (i.e. control) for sequencing. DNA from each individual was separately digested using *EcoT221* 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 TASSLE (*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 taxon coverage at 0.1, and minimum minor allele frequency at 0.01. Essentially, 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 TASSLE 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

**Table 1. Collection information for *I. hexagona* populations**

<table>
<thead>
<tr>
<th>Population ID</th>
<th>Latitude</th>
<th>Longitude</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL_01</td>
<td>25.790</td>
<td>–81.100</td>
<td>12</td>
</tr>
<tr>
<td>FL_04</td>
<td>27.267</td>
<td>–82.121</td>
<td>12</td>
</tr>
<tr>
<td>FL_08</td>
<td>26.944</td>
<td>–81.449</td>
<td>12</td>
</tr>
<tr>
<td>FL_10</td>
<td>28.463</td>
<td>–82.054</td>
<td>12</td>
</tr>
<tr>
<td>FL_11</td>
<td>28.507</td>
<td>–82.124</td>
<td>13</td>
</tr>
<tr>
<td>FL_13</td>
<td>29.983</td>
<td>–81.675</td>
<td>11</td>
</tr>
<tr>
<td>LA_02</td>
<td>29.880</td>
<td>–91.784</td>
<td>11</td>
</tr>
<tr>
<td>LA_04</td>
<td>30.084</td>
<td>–90.449</td>
<td>12</td>
</tr>
</tbody>
</table>

n, number of individuals.
threshold of log10 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 Differentiation and Structure

We computed Fst 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 StrAmp (Pemberton et al. 2013). StrAmp uses the method proposed by Wright (1951) and updated by Weir and Cockerham (1984) in order to calculate Fst. We compared the overlap of confidence intervals for 2 datasets (the total dataset of 750 SNPs with the outlier dataset with 14 SNPs) to determine if pairwise Fst 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 (DA) on the retained principal components. DAPC consists of a 2-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 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 Differentiation

First, we tested whether population mean values for each trait differed significantly using an ANOVA. Then to visualize the level of morphological variation across populations, we conducted a PCA of the 5 morphological measurements and plotted the results using the function ggbiplot in R. Pearson’s correlations were also calculated for all 5 traits in order to determine which traits were significantly correlated.

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 ≤ 0.90 and then performed a PCA (Supplementary Table 1 online). We then determined the variables with the highest contributions on the first 2 principal components, which captured 91.3% of the variation in the climate data. The environmental variables with the highest loadings were then used in subsequent analyses.

Inference of selection is strengthened by multiple independent tests, where loci identified as significant in more than one test are considered robust candidates for being under selection (Bradbury et al. 2013). Thus, we tested for associations between the SNP genotypes and either the climate or morphology dataset using a general linear model implemented in TASSEL 3.0 (Bradbury et al. 2007). To reduce the confounding effects of population structure, we incorporated a Q-matrix of population membership estimates. Here, we tested for an association based on the first 2 principal components (see results below) for the morphology dataset. For the climate dataset, we tested for an association with climatic variables, which had the highest loading on the first 2 principal components (see results below). Associations were considered significant at a P value of 0.05 and using the positive FDR method. From this, we assessed if SNPs associated with either morphology or climate were the same as those loci with the strongest degree of population differentiation identified via BAYESCAN.

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 with 1000 permutations, 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 ≤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 10−8 (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 Fst value higher than expected when the FDR was set at 0.01. These loci were inferred to be under diversifying selection (Table 2). We found a significant effect from these 14 outlier loci on the range of estimates of Fst (Tables 3 and 5). We then plotted the frequency of the 14 outlier SNPs to determine the pattern of allele frequencies across populations (Supplementary Figure 1 online). For some of the outlier SNPs (n = 3), we see a pattern where Florida populations are fixed for one allele and the Louisiana populations fixed for the alternate allele. However, most of the outlier SNPs exhibit a pattern of variability within Florida populations.

Population Differentiation and Structure

Using a panel of 750 SNPs and 92 individuals, we estimated pairwise Fst to range from 0.085 to 0.392 when we used all 750 SNPs; this
From the DAPC analyses, the BIC reached its minimum value at $K = 5$ suggesting 5 genetic clusters (Supplementary Figure 2 online). The 2 Louisiana populations (LA_02 and LA_04) were assigned as a single population. The 2 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 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 1). Finally, 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 and FL_11 from the other Florida populations (Figure 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: $P = 0.00222$, stem height: $P = 1.41e-11$, petal length: $P = 3.7e-06$, petal width: $P = 5e-06$, and nectar guide: $P = 2.03e-08$). 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 2b). The first 3 principal components explained 88.2% of the total variation (Figure 2b). Floral trait measurements were the major loadings of PC1 (petal length $= -0.523$, petal width $= -0.4752$, and nectar guide $= -0.4762$). Leaf height was the major loading of PC2 ($-0.710$). Leaf height was only correlated with floral stem height (Table 6). All other trait comparisons were significantly correlated with each other (Table 6). For the first 2 principal components, we identified 26 SNPs with the strongest association to morphology principal component 1 (Supplementary Table 2a online) and 25 SNPs with the strongest association to morphology principal component 2 (Supplementary Table 2b online) as candidates potentially underlying trait differences between populations. All candidate SNPs had a $-\log_{10}(P)$ value $> 3.0$ (Supplementary Figure 4a,b online). Of note, SNP 349 was associated with principal component 1 and SNP 294 was associated with principal component 2, which were both identified as exhibiting very strong evidence of being under selection from the BAYESCAN analysis.

### A positive $\alpha$ value indicates diversifying selection.

Prob, posterior Bayes probability.

### Table 3. Pairwise $F_{st}$ values for all populations using all 750 SNPs

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Prob</th>
<th>Alpha</th>
<th>$F_{st}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.983</td>
<td>1.333</td>
<td>0.62981</td>
</tr>
<tr>
<td>71</td>
<td>0.9974</td>
<td>1.8353</td>
<td>0.72171</td>
</tr>
<tr>
<td>117</td>
<td>0.9962</td>
<td>1.8104</td>
<td>0.71708</td>
</tr>
<tr>
<td>172</td>
<td>0.9958</td>
<td>1.4317</td>
<td>0.65004</td>
</tr>
<tr>
<td>184</td>
<td>0.9964</td>
<td>1.5854</td>
<td>0.67854</td>
</tr>
<tr>
<td>294</td>
<td>0.9986</td>
<td>1.6781</td>
<td>0.69651</td>
</tr>
<tr>
<td>330</td>
<td>0.9788</td>
<td>1.4377</td>
<td>0.64931</td>
</tr>
<tr>
<td>349</td>
<td>0.9798</td>
<td>1.4638</td>
<td>0.65427</td>
</tr>
<tr>
<td>381</td>
<td>0.9842</td>
<td>1.5478</td>
<td>0.67065</td>
</tr>
<tr>
<td>480</td>
<td>0.998</td>
<td>1.8041</td>
<td>0.71655</td>
</tr>
<tr>
<td>575</td>
<td>0.9862</td>
<td>1.504</td>
<td>0.66269</td>
</tr>
<tr>
<td>605</td>
<td>0.99</td>
<td>1.2707</td>
<td>0.61786</td>
</tr>
<tr>
<td>647</td>
<td>0.9994</td>
<td>1.7354</td>
<td>0.7074</td>
</tr>
<tr>
<td>733</td>
<td>0.9838</td>
<td>1.323</td>
<td>0.62773</td>
</tr>
</tbody>
</table>

Shown in bold are values with nonoverlapping confidence intervals between the total dataset of 750 SNPs (Table 3) and the 14 putatively nonneutral SNPs (Table 5) and thus are significantly different.

### Table 4. Pairwise $F_{st}$ values for all populations using 736 putatively neutral SNPs

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Prob</th>
<th>Alpha</th>
<th>$F_{st}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.983</td>
<td>1.333</td>
<td>0.62981</td>
</tr>
<tr>
<td>71</td>
<td>0.9974</td>
<td>1.8353</td>
<td>0.72171</td>
</tr>
<tr>
<td>117</td>
<td>0.9962</td>
<td>1.8104</td>
<td>0.71708</td>
</tr>
<tr>
<td>172</td>
<td>0.9958</td>
<td>1.4317</td>
<td>0.65004</td>
</tr>
<tr>
<td>184</td>
<td>0.9964</td>
<td>1.5854</td>
<td>0.67854</td>
</tr>
<tr>
<td>294</td>
<td>0.9986</td>
<td>1.6781</td>
<td>0.69651</td>
</tr>
<tr>
<td>330</td>
<td>0.9788</td>
<td>1.4377</td>
<td>0.64931</td>
</tr>
<tr>
<td>349</td>
<td>0.9798</td>
<td>1.4638</td>
<td>0.65427</td>
</tr>
<tr>
<td>381</td>
<td>0.9842</td>
<td>1.5478</td>
<td>0.67065</td>
</tr>
<tr>
<td>480</td>
<td>0.998</td>
<td>1.8041</td>
<td>0.71655</td>
</tr>
<tr>
<td>575</td>
<td>0.9862</td>
<td>1.504</td>
<td>0.66269</td>
</tr>
<tr>
<td>605</td>
<td>0.99</td>
<td>1.2707</td>
<td>0.61786</td>
</tr>
<tr>
<td>647</td>
<td>0.9994</td>
<td>1.7354</td>
<td>0.7074</td>
</tr>
<tr>
<td>733</td>
<td>0.9838</td>
<td>1.323</td>
<td>0.62773</td>
</tr>
</tbody>
</table>

Shown in bold are values with nonoverlapping confidence intervals between the total dataset of 750 SNPs (Table 3) and the 14 putatively nonneutral SNPs (Table 5) and thus are significantly different.

### Table 5. Pairwise $F_{st}$ values for all populations using 14 putatively nonneutral SNPs

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>FL_01</th>
<th>FL_04</th>
<th>FL_08</th>
<th>FL_10</th>
<th>FL_11</th>
<th>FL_13</th>
<th>LA_02</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.658</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>0.510</td>
<td>0.639</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>0.592</td>
<td>0.691</td>
<td>0.622</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>0.574</td>
<td>0.627</td>
<td>0.543</td>
<td>0.137</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>184</td>
<td>0.399</td>
<td>0.310</td>
<td>0.397</td>
<td>0.356</td>
<td>0.301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>330</td>
<td>0.789</td>
<td>0.615</td>
<td>0.781</td>
<td>0.766</td>
<td>0.678</td>
<td>0.320</td>
<td></td>
</tr>
<tr>
<td>349</td>
<td>0.829</td>
<td>0.641</td>
<td>0.829</td>
<td>0.809</td>
<td>0.735</td>
<td>0.373</td>
<td>0.098</td>
</tr>
</tbody>
</table>

Shown in bold are values with nonoverlapping confidence intervals between the total dataset of 750 SNPs (Table 3) and the 14 putatively nonneutral SNPs (Table 5) and thus are significantly different.

---

**Tables 2, 3, 4, 5**

- **Table 2.** The 14 outliers identified using BAYESCAN from the total dataset of 750 SNPs
- **Table 3.** Pairwise $F_{st}$ values for all populations using all 750 SNPs
- **Table 4.** Pairwise $F_{st}$ values for all populations using 736 putatively neutral SNPs
- **Table 5.** Pairwise $F_{st}$ values for all populations using 14 putatively nonneutral SNPs

---

**Supplementary Tables and Figures**

- **Supplementary Table 2a**
- **Supplementary Figure 2a**
- **Supplementary Figure 2b**
- **Supplementary Figure 4a,b**
We identified population-level differences for environmental variation (Figure 3). 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). The environmental variable that contributed the most to principal component 1 was NDVI or the normalized difference of vegetative index (0.479). The variable that contributed the most to principal component 2 was mean temperature of the driest quarter (−0.500). Finally, the variable that contributed the most to principal component 3 was TREE (−0.779), which is a measure of percent tree cover. For mean temperature of driest quarter, we identified 35 SNPs with the strongest association (Supplementary Table 3a online) and 35 SNPs with the strongest association to NDVI (Supplementary Table 3b online) as candidates potentially underlying environmental differences between populations. All candidate SNPs had a −log10 (P) value > 4.0 (Supplementary Figure 5a,b online).

However, the R^2 values for environment variables are at least an order magnitude less than those calculated for morphology. Of note, SNP 733 was associated with both NDVI and mean temperature of driest quarter and was identified as exhibiting very strong evidence of being under selection from the BAYESCAN analysis. Patterns of allele frequencies for loci that were identified via multiple analyses were plotted (Supplementary Figure 1 online).

Distance Matrices and Mantel Tests
There is an overall relationship between pairwise neutral genetic distance and geographic distance, suggesting that IBD is causing the neutral, genetic structure (Supplementary Figure 3a online). This is further confirmed when we test for IBD only in the Florida populations with neutral genetic distance (Mantel test, R = 0.772; P = 0.004, Table 6. Correlations among the 5 morphological traits measured

<table>
<thead>
<tr>
<th></th>
<th>Petal. length</th>
<th>Petal. width</th>
<th>Nectar. guide</th>
<th>Leaf. height</th>
<th>Stem. height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petal. length</td>
<td>—</td>
<td>1.49e-10</td>
<td>1.33e-9</td>
<td>0.039</td>
<td>6.75e-5</td>
</tr>
<tr>
<td>Petal. width</td>
<td>0.6380</td>
<td>—</td>
<td>1.41e-9</td>
<td>0.985</td>
<td>2.55e-3</td>
</tr>
<tr>
<td>Nectar. guide</td>
<td>0.6113</td>
<td>0.6107</td>
<td>—</td>
<td>0.732</td>
<td>1.01e-3</td>
</tr>
<tr>
<td>Leaf. height</td>
<td>0.2294</td>
<td>0.0019</td>
<td>0.0385</td>
<td>—</td>
<td>5.418e-12</td>
</tr>
<tr>
<td>Stem. height</td>
<td>0.4279</td>
<td>0.3307</td>
<td>0.3588</td>
<td>0.497</td>
<td>—</td>
</tr>
</tbody>
</table>

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 2. (a) Samples are assigned to their genetic cluster by discriminant analysis of PCs. The bar graph inset displays the eigenvalues of the 4 principal components in relative magnitude and illustrates the variation explained by the 4 PCs. (b) Principal component analysis of 5 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 Figure 2a. Individuals from different sample sites which are considered a single genetic cluster are differentiated by open and closed circles.

Figure 3. PCA 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.
data not shown). We did find a positive association between non-neutral genetic distance and geographic distance (Supplementary Figure 3b online), suggesting that these loci are the strongest contributors to geographic structure, which may reflect regions involved in local adaptation. The regression of mean floral traits on genetic distance between populations based on neutral loci is not greater than expected by chance (Supplementary Figure 3c online). In contrast, we detected a moderate and weak correlation between mean floral traits and the genetic distance between populations based on nonneutral SNPs (Supplementary Figure 3d online). While there is not much of a difference in significance between neutral and non-neutral loci on the regression of mean floral traits, the direction is consistent with our prediction.

**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 Fst 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 pairwise values of Fst; however, not all values were significantly different from the estimate of Fst 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. We estimate that 3.33% of the SNPs (n = 25) were significantly associated with principal component 1 or floral shape. Furthermore, one of these same SNPs was also identified in the BAYESCAN analysis and is likely to contribute to flower size differences between populations. Moreover, populations exhibited distinct environmental differences when examining the uncorrelated environmental variables. One SNP (733) was associated with 2 environmental variables and identified in the BAYESCAN analysis.

The SNPs putatively under selection were positively correlated with both geographic distance and phenotypic distance, albeit weakly to phenotypic distance. Selection may be acting on the genomic regions surrounding the putatively selected SNPs. Neutral SNPs were only correlated with geographic distance and thus IBD was observed for 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 United States 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 Loci under Selection**

Despite the spatial genetic structuring of *I. hexagona* populations revealed by the analyses based on all loci, 14 loci exhibited strong nonneutral signatures and were presumably selected or linked to selected regions of the genome. Additionally, when we examined patterns of allele frequencies for the 14 SNPs, we detected extreme allele frequency differences between different geographic regions (Florida vs. 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 a lack of diversity 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 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. For example, Eckert et al. (2010) found loblolly pine outlier loci to be associated with aridity or temperature. For *I. hexagona*, we identified a strong signature of diversifying selection at one SNP (733), as well as an association with 2 environmental variables that may be important in driving adaptation in natural populations of *I. hexagona*. Using a similar analysis design, Bradbury et al. (2013) identified 2 strong candidates for diversifying selection in natural populations of a forest tree. The strength of this study was the inclusion of molecular markers with homology to known genes, which allowed Bradbury et al. (2013) to elucidate the genes and ecological processes that may be involved in adaptation. However, in our current study, it is not possible to ascertain the potential function of this marker as we are only using molecular markers without homology to known genes.

Furthermore for *I. hexagona*, the finding of an association at outlier loci and variation in floral traits provides evidence that selection may be acting on phenotypic divergence within this species. And 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 specific 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).

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 IBD and local adaptation. 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.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxford-journals.org/.

Funding
American Iris Society; National Institute of Health (award number T32GM007103); and National Science Foundation (grants DEB-0949479/0949424) (collaborative grant with NH Martin of Texas State University) and DEB-1049757. M.L.A. was supported by a grant from the Chinese Academy of Sciences, Kunming Institute ofZoology (Kunning, PRC).

Acknowledgments
J.A.P.H. would like to acknowledge the many governmental agencies that provided valuable information regarding natural populations. J.A.P.H. would also like to thank members of the Wares and Sweigart lab for fruitful discussions regarding this manuscript.

References
Bouck A, Wessler SR, Arnold ML. 2007. QTL analysis of floral traits in Loui-ator 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).

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 IBD and local adaptation. 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.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxford-journals.org/.

Funding
American Iris Society; National Institute of Health (award number T32GM007103); and National Science Foundation (grants DEB-0949479/0949424) (collaborative grant with NH Martin of Texas State University) and DEB-1049757. M.L.A. was supported by a grant from the Chinese Academy of Sciences, Kunming Institute ofZoology (Kunning, PRC).

Acknowledgments
J.A.P.H. would like to acknowledge the many governmental agencies that provided valuable information regarding natural populations. J.A.P.H. would also like to thank members of the Wares and Sweigart lab for fruitful discussions regarding this manuscript.

References
Bouck A, Wessler SR, Arnold ML. 2007. QTL analysis of floral traits in Loui-ator 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).

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 IBD and local adaptation. 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.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxford-journals.org/.

Funding
American Iris Society; National Institute of Health (award number T32GM007103); and National Science Foundation (grants DEB-0949479/0949424) (collaborative grant with NH Martin of Texas State University) and DEB-1049757. M.L.A. was supported by a grant from the Chinese Academy of Sciences, Kunming Institute ofZoology (Kunning, PRC).

Acknowledgments
J.A.P.H. would like to acknowledge the many governmental agencies that provided valuable information regarding natural populations. J.A.P.H. would also like to thank members of the Wares and Sweigart lab for fruitful discussions regarding this manuscript.

References
Bouck A, Wessler SR, Arnold ML. 2007. QTL analysis of floral traits in Loui-ator 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).

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 IBD and local adaptation. 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.


