POPULATION-GENETIC STUDIES OF IRIS, A MODEL GENUS FOR INVESTIGATING SPECIATION

by

ROBERT SCOTT CORNMAN

(Under the Direction of Michael Arnold)

ABSTRACT

Empirical population-genetic studies can be used to investigate a variety of processes related to speciation. I applied population-genetic approaches to understand how genetic divergence occurs at various scales in the genus *Iris*, which is a model system for investigating speciation in plants. I first examined the problem of recombinant hybrid speciation, the occurrence of which is theoretically difficult but has been empirically demonstrated in *Iris*. Hybrid speciation requires the stabilization of a recombinant genome in sympatry with parental genotypes. Stabilization can be facilitated by factors such as assortative mating, selfing, vegetative propagation, or localized gene flow. To examine the extent to which these factors contribute to actual hybrid zone dynamics, I compared the genetic structure of established plants in a Louisiana Iris hybrid population with their mating patterns and progeny genotypes. There was no evidence that variation in phenology contributed to assortative mating, nor was there evidence of limited pollen dispersal: individuals from outside the censused population constituted one-half of all outcross paternity. Nonetheless, after controlling for clonal structure, we observed a strong spatial localization of related genotypes demonstrated by autocorrelation and cluster analyses. Furthermore, these clusters were not the product of selfing because the

average inbreeding coefficient was near zero. I conclude that postzygotic selection limits successful sexual reproduction in this hybrid zone to matings between similar genotypes. Distinct hybrid lineages are therefore buffered from recombination while still in close geographic association with other recombinant or parental types, greatly facilitating hybrid speciation in this system.

The second application of population-genetic approaches to the study of speciation involved the development of nuclear and chloroplast markers for *I. missouriensis* to determine whether strong genetic differentiation is present in this potentially cryptic species complex. I did not find evidence of genetic differentiation indicative of multiple species, but I was able to infer patterns of demographic history associated with interglacial climate change. Populations in the southern Rocky Mountains are more closely related to coastal California than to the more biogeographically similar Sierra Nevada and Great Basin, suggesting the importance of long-distance dispersal in the history of the species.

INDEX WORDS: Iris, speciation, hybridization, paternity analysis, cluster analysis, transposon display, phylogeography.

POPULATION-GENETIC STUDIES OF IRIS, A MODEL GENUS FOR INVESTIGATING SPECIATION

by

ROBERT SCOTT CORNMAN

B.A., Rice University, 1992

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006

POPULATION-GENETIC STUDIES OF IRIS, A MODEL GENUS FOR INVESTIGATING SPECIATION

by

ROCERT SCOTT CORNMAN

Major Professor:	Michael Arnold
Committee:	Wyatt Anderson Shu-Mei Chang Rodney Mauricio John Wares

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2006

DEDICATION

To my family

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

The genus *Iris* has emerged as an important model for the study of speciation and related processes in plants. This is because of the number of key studies that have examined North American members of the rhizomatous beardless irises, subgenus *Limniris*. For example, one of the first and most clearly documented cases of allopolyploid speciation was described by Anderson (1936) based on detailed phenotypic analyses and chromosome counts. Anderson inferred that *Iris versicolor*, a phenotypically variable species of the northeastern United States, was an allopolyploid offspring of *I. virginica*, which occurs in the mid-Atlantic and southeastern United States, and *I. setosa*, which has a boreal distribution from Japan to Newfoundland (Anderson 1936). (Interestingly, the multivariate data set of flower morphology he collected to demonstrate the intermediacy but distinctiveness of *I. versicolor* relative to its putative parental species is one of the most frequently analyzed data sets in the history of botany, because it has become a reference for the field of cluster analysis.) Cytogenetic studies confirmed that *I. versicolor* has a chromosome count of approximately 2n = 108, whereas *I. virignica* and *I. setosa* have 2n = 72 and 2n = 36 chromosomes, respectively (Anderson 1936). Presently, little is known about the relative fitness of the allopolyploid and its parents although contact zones exist today and have likely been common in recent geologic history as climate change-induced migration has occurred. Furthermore, no population-genetic study has been performed to investigate the number of origins of *I. versicolor* or their direction and timing, nor species-wide genetic

diversity. Such studies would be critical for understanding the dynamics of polyloid speciation, a process that has contributed significantly to the overall diversity of angiosperms.

Iris was also an early model of rapid ecological speciation in the absence of strong postzygotic isolation. The Pacific Coast Irises (series *Californicae*) are a monophyletic group (ref) of approximately ten species first investigated by Lenz (1959), who found them to be highly interfertile but occupying distinct ecological niches which presumably maintained their isolation. In his landmark text, "Plant Speciation", Grant (1971) considered the *Californicae* "[t]he most complex syngameon which has been analyzed in detail (p. 55)." Disturbance, particularly human-mediated, has led to hybridization (Lenz 1959, Young 1996), but the strength and nature of ecological selection on hybrid phenotypes remains largely unknown. One difficulty that has hindered further research on this historically important group, as well as other iris (see below and subsequent chapters), is the relative paucity of genetic markers for phylogeny and population genetics (Young 1996, Wilson 2003).

Perhaps the best-known and most-studied irises are the Louisiana Irises (series *Hexagonae*), which have been investigated as a model for the creative role of diploid hybridization in adaptive introgression and speciation (Arnold 2006). Hybridization between *I. brevicaulis*, *I. fulva*, and *I. hexagona* has been investigated since the early 19th century (Small and Alexander 1931, Viosca 1935, Foster 1937, Riley 1938) and the occurrence of the three-way hybrid species *I. nelsonii* was postulated by Randolph (1966) and later confirmed by Arnold (1993). Additionally, certain complex fitness-related traits such as shade tolerance or salinity tolerance may have been transferred between species by introgression. This conclusion is based on novel ecological associations of otherwise phenotypically 'normal' individuals (e.g., Cruzan

and Arnold 1993), and has begun to be investigated experimentally with genetically mapped individuals (Martin et al. 2006).

While *Iris* has been a useful model for investigating different modes of speciation such as allopolyploidy, ecological radiation, and diploid hybridization, ultimately we wish to understand the genetic and ecological factors that promote cladogenesis in some lineages but not in others. Answering this question will require a detailed understanding of the genetics of speciation across a range of related taxa, and *Iris* may ultimately serve as a model for such a comparative approach as well. First steps towards this goal include increasing the number of genetic tools available and further clarifying the geographic structure of genetic variation within species, so that among species comparisons can be more fruitful.

This dissertation describes empirical population-genetic studies that investigate different scales of gene flow as it relates to speciation in *Iris*. Chapter Two describes a micro-scale study of the dynamics of diploid hybridization in a Louisiana Iris hybrid zone. The goal of this study was to determine whether processes suggested by theory (McCarthy et al. 1995, Buerkle et al. 2000) to be important for hybrid speciation are actually present in a contemporary hybrid zone. Chapter Three describes the development of genetic markers for investigating the population-genetics of *I. missouriensis*. Chapter Four describes the application of these markers to quantifying the continental scale of variation in this widespread species. The dynamics of population migration across complex landscapes generates isolation and remixing that can lead to proto-species, generation of epistatic novelty, or re-establishment of species cohesion (Hewitt and Ibrahim 2001). Thus, the first study examines the fine-scale resolution of a divergence process that is in part generated by, and strongly constrained by, processes examined in the latter studies.

CHAPTER 2

CONTRASTING GENETIC STRUCTURE OF ADULTS AND PROGENY IN A LOUISIANA IRIS HYBRID POPULATION[1]

[1] Cornman, R. S., J. M. Burke, R. A. Wesselingh, and M. L. Arnold. 2004. Evolution. 58:2669-2681. Reprinted here with the permission of publisher.

ABSTRACT

Studies of natural hybridization have suggested that this process may be a creative stimulus for adaptive evolution and speciation. An important step in this process is the establishment of fit recombinant genotypes that are buffered from subsequent recombination with unlike genotypes. We used molecular markers and a two-generation sampling strategy to infer the extent of recombination in a Louisiana Iris hybrid zone consisting predominantly of *Iris fulva*-type floral phenotypes. Genotypic diversity was fairly high, indicating that sexual reproduction is frequent relative to clonal reproduction. However, we observed strong spatial genetic structure even after controlling for clonality, which implies a low level of pollen and seed dispersal. We therefore used cluster analysis to explore the hypothesis that the *fulva*-type hybrids are an admixture of groups between which there has been limited recombination. Our results indicate that several such groups are present in the population and are strongly localized spatially. This spatial pattern is not attributable strictly to a lack of mating opportunities between dissimilar genotypes for two reasons: 1) relatedness of flowering pairs was uncorrelated with the degree of overlap in flowering, and 2) paternity analysis shows that pollen movement among the outcross fraction occurred over large distances, with roughly half of all paternity attributed to pollen flow from outside the population. We also found evidence of strong inbreeding depression, indicated by contrasting estimates of the rate of self-fertilization ($s = 0.48$) and the average inbreeding coefficient ($f = 0.007$) of *fulva*-type hybrids. We conclude that groups of similar hybrid genotypes can be buffered from recombination at small spatial scales relative to pollen flow, and selection against certain recombinant genotypes may be as or more important than clonal reproduction and inbreeding.

Keywords: Natural hybridization, paternity analysis, mating patterns, inbreeding depression, reproductive isolation, spatial autocorrelation

INTRODUCTION

Numerous investigators have examined the dynamics of natural hybridization in order to identify the nature of species boundaries. These studies show that species boundaries are eclectic phenomena that may have simple or complex genetic and ecological foundations (reviewed by Templeton 1981, Orr 2001, Howard et al. 2002). Furthermore, species boundaries may be semi-permeable, allowing the introgression of genetic material from one species to another (Key 1968, Harrison 1990). An extensive literature has developed to address the theoretical basis and empirical pattern of introgression in a diverse array of organisms (Endler 1977, Barton and Hewitt 1985, Harrison 1990, Arnold 1997). These studies seek to draw general conclusions regarding the evolutionary consequences of hybridization from detailed study of dispersal, selection, and mating patterns.

Such work has led to the hypothesis that selection may favor the introgression of certain traits, and can result in the transfer of complex adaptations and even the creation of ecologically distinct homoploid (i.e., non-polyploid) species (Anderson and Stebbins 1954, Lewontin and Birch 1966). Indeed, recent studies have documented the occurrence of natural hybrids in habitats outside the normal range of parental types (e.g., Cruzan and Arnold 1993) as well as greater fitness of hybrid genotypes in novel or extreme environments (e.g., Lexer et al. 2003). Molecular techniques have been used to verify species of hybrid origin (e.g., Rieseberg 1991, Arnold 1993, Wolfe et al. 1998), and the genetic architecture of adaptation via hybridization has been the topic of recent investigations (Rieseberg et al. 2003).

There are, however, theoretical difficulties with the establishment of novel recombinant lineages. The primary obstacle is the breakup of rare advantageous gene combinations by subsequent recombination with the more abundant parental types. Several authors have investigated models that address the likelihood of adaptive introgression and/or hybrid speciation (Grant 1971, Templeton 1981, McCarthy et al. 1995, Buerkle et al. 2000, Barton 2001), usually in the context of recombinational speciation (Grant 1971). These models demonstrate the feasibility of hybrid speciation under some combination of chromosomal or genic sterility, self-compatibility or asexual reproduction, and ecological isolation. More empirical studies of hybridization are needed to determine whether these conditions are frequently realized in nature, or indeed if empirically demonstrated cases conform to these predictions.

A number of complementary strategies can be employed to study the outcomes of hybridization. Mechanistic studies of prezygotic and postzygotic barriers have been used to infer the rates of initial and subsequent hybrid formation, and to identify conditions under which hybridization is most likely to occur (e.g., Grant and Grant 1996, Hodges et al. 1996, Campbell et al. 1998). Common garden studies of known or artificially produced hybrid classes have been used to evaluate the performance of early-generation hybrids in various environments (e.g., Emms and Arnold 1997, Wang et al. 1997, Burke et al. 1998a, Orians et al. 1999, Schweitzer et al. 2002). A third approach that has grown in utility over the past few decades is the use of molecular markers to infer population-genetic processes from the distribution of genotypes in natural populations (e.g., Arnold et al. 1987, Szymura and Barton 1991). Molecular markers are useful tools because they provide more lucid descriptions of the structure of hybrid zones, which are often admixtures of individuals with complex ancestry, and because they can be used to infer patterns of gene exchange, recombination, and selection. One goal of marker-based studies is to

use patterns of genetic structure at different life-history stages to predict how the genotypic constitution of hybrid zones may change over time due to these factors (e.g. Cruzan and Arnold 1994 and references therein). This cross-generational approach is made more powerful if phenotypic traits that affect mating are recorded, so that potential genetic correlations induced by mate choice can be identified and not erroneously attributed to postzygotic selection. Molecular markers can thus be used to infer population-level processes relevant to the likelihood of adaptive introgression or hybrid speciation.

Here we utilize this latter approach to investigate sources of genetic structure in a Louisiana Iris hybrid zone. The factors shaping the distribution of genotypes in adults and progeny have been examined previously in Louisiana Iris hybrids by Cruzan and Arnold (1994). The authors used species-specific nuclear and chloroplast markers for *I. fulva* and *I. brevicaulis* and a spatially restricted mixed-mating model to generate expected distributions of seed genotypes. By comparing this expected distribution with the observed distribution of progeny, they concluded that there was assortative mating amongst *I. fulva* cytotypes (i.e., *I. fulva* and "*I. fulva*-like" hybrids) resulting in progeny with more *fulva* markers than expected. This was not observed for *I. brevicaulis* cytotypes. Furthermore, seed progeny that had intermediate genotypes were aborted at higher rates. Thus, both assortative mating and post-fertilization selection were detected by this study. However, this assortative mating was with respect to fairly differentiated groups, *I. fulva* cytotypes versus *I. brevicaulis* cytotypes, because intermediate forms were lacking among adult plants. Due to this, the study lacked the resolution to determine how much gene flow occurred among genets of a given cytotype. It is this latter question that is most relevant to the stabilization of novel adapted lineages, because such matings are less constrained by prezygotic or postzygotic factors.

In the present study, we used molecular markers to compare the genetic structure of adult plants and their seed progeny produced during one reproductive season in a Louisiana Iris hybrid zone. We also recorded the dates of opening and closing of each flower in the population to determine whether overlapping phenology leads to genetic correlations among mating pairs. From these combined data, we then inferred the extent to which phenology, postzygotic selection, and/or dispersal direct hybridization outcomes. Specifically, we asked the following questions:

1. What is the genetic structure of the adult population in terms of inbreeding coefficients, linkage disequilibrium, and the spatial distribution of alleles?
2. Are co-flowering individuals genetically correlated?
3. What is the distribution of family-level outcrossing rates?
4. What is the genotypic distribution of progeny, and what is the distance of pollen movement?
5. Are the observed selfing rates and patterns of pollen flow consistent with the observed genetic structure of the population?

MATERIALS AND METHODS

Study Organism and Sampling

The Louisiana Irises have long served as a model system for studying hybridization in plants (Riley 1938, Anderson 1949, Randolph 1966, Bennett and Grace 1990, Arnold 2000). This North American species complex (section Hexagonae) consists of four accepted species: *Iris fulva*, *I. brevicaulis*, *I. hexagona*, and *I. nelsonii* (we follow Foster (1937) and Goldblatt (1990) in maintaining *I. hexagona* var. *giganticaerulea* at subspecific rank). *Iris fulva*, *I. brevicaulis*, and *I.*

hexagona have broad but largely distinct ranges in the central and eastern United States. These distributions overlap in southern Louisiana where hybridization is frequent. The fourth species, *I. nelsonii*, is itself a stabilized hybrid of the other three (Randolph 1966, Arnold 1993), and is a rare endemic of southern Louisiana. The species are distinctive in terms of pollination biology and ecophysiology, but hybrids often have high fitness under experimental conditions (Johnston et al. 2003) and in nature can be found in both parental and novel environments (Johnston et al. 2001). This system is therefore ideal for investigating the creative role of hybridization in producing new adaptations and/or new species.

The Louisiana Iris species are perennial, spring-flowering herbs that also reproduce vegetatively by budding of underground rhizomes. Because of this clonal habit, we shall use the term ramet to identify physically separate individual plants and the term genet to identify the collection of clonally derived ramets that are ultimately the same genetic individual. Two very distinct pollinator syndromes are present in the Louisiana Irises. The red-flowered *I. fulva* is predominantly visited by hummingbirds, whereas the blue-flowered *I. hexagona* and *I. brevicaulis* are visited primarily by bumblebees. While there is a small amount of overlap in pollinator visitation between the species, the efficacy of these visits in effecting pollination is unknown and the frequency of F_1 seed formation in nature is very low (Arnold et al. 1993, Hodges et al. 1996). Although the species are self-compatible, flowers are protandrous for one to two days after anthesis, after which the stigma surface is exposed and receptive (Cruzan and Arnold 1994).

The studied population borders Young's Coulee (a bayou) in Vermillion Parish, southern Louisiana. Although we have limited knowledge of the history of this population, Louisiana Iris hybrids have persisted in the area for decades, if not longer (Randolph et al. 1961). The

population consists of dense patches of ramets growing on a mudflat and adjacent shaded banks of the bayou. The site has been previously investigated (Arnold 1993) and nuclear markers for *I. fulva*, *I. brevicaulis*, and *I. hexagona* and chloroplast markers for *I. fulva* and *I. hexagona* were detected. Most plants have *I. fulva* cytotypes and are morphologically and genetically hybrid. There is also a single patch of ramets that is morphologically *I. hexagona* but within which *I. fulva* nuclear markers have also been detected (Arnold 1993). We therefore consider two broad *phenotypic* classes of iris to be present: *fulva*-type hybrids and *I. hexagona*. Mating between phenotypic classes is expected to be rare due to prezygotic barriers such as low pollinator overlap and pollen competition (reviewed in Arnold 2000). Although quantitative variation in floral traits exists among *fulva*-type hybrids in nature (R. Cornman unpublished data), we do not know to what extent, if any, this variation affects mating patterns within this phenotypic group. However, Wesselingh and Arnold (2000) have shown that hummingbirds do not discriminate against a broad range of hybrid phenotypes.

The Young's Coulee population is bordered by pasture, housing, and swamp and is separated from most other patches of ramets by 50 meters or more. In the spring of 1997, all ramets were spatially mapped to the nearest centimeter. The mapped area was approximately 60 by 25 meters in size and contained 176 *fulva*-type ramets and 21 *I. hexagona* ramets. A small number of ramets were located within 50 meters of the circumscribed population but could not be mapped for logistical reasons.

The study site was visited almost daily from March 30th to April 24th and the gender phase of each flower was recorded from anthesis until wilting. In a few instances of missing observations, the status of flowers was estimated by assuming that stigmas become receptive one day after anthesis. This assumption is conservative with respect to the exclusion of candidate

mating pairs. As is typical for these species, most ramets produced a single inflorescence; however, two ramets each produced two separate inflorescences. Reproductive output was relatively low (cf. Wesselingh and Arnold 2003), presumably due to drought conditions. We therefore collected three fruits from two additional *fulva*-type hybrids located near the mapped population in order to improve our estimate of the mating system and pollen flow. These plants were not included in any genetic structure analysis of adults.

The average number of flowers per inflorescence was 3.1 ± 0.9 (SD), the percentage of flowers setting fruit was 29% (23 of 80), and the average number of seeds per fruit was 19.0 ± 14.2 (SD). Fruits were allowed to mature naturally prior to collection and transported to the University of Georgia in June 1997. After fruits had dried in the laboratory, seeds were counted and then germinated in soil. Leaf tissue of germinated progeny was harvested for DNA extraction. Seeds that failed to germinate after six months were deemed inviable. Tissue was stored at –80° C prior to extraction and genotyping in 2001 and 2002.

DNA Extraction and Genotyping

Genomic DNA was extracted using a hexadecyltrimethyl ammonium bromide (CTAB) extraction buffer protocol (Doyle and Doyle 1987). All ramets and all progeny were genotyped at five microsatellite loci characterized by Burke and Arnold (1999). Polymerase chain reaction (PCR) was performed in a 20 µl volume containing 8 pmol each of a fluorescently labeled forward primer and an unlabeled reverse primer, 0.8 units of *Taq* polymerase in the supplied buffer (Promega Corp., Madison, WI), 2.0 mM magnesium chloride, 125 µM of each deoxyribonucleotide triphosphate, and approximately 500 ng of genomic DNA. Reactions were performed with a Perkin Elmer (Wellesley, MA) model 9700 thermocycler. The reaction products were separated on a 6% polyacrylamide gel using an ABI Prism 377 electrophoresis

unit (Applied Biosystems, Foster City, CA), and allele sizes were estimated using the GeneScan software package (Applied Biosystems, Foster City, CA).

Two of the five primer pairs, IB025 and IB141, had segregation patterns in progeny arrays indicating a high frequency of null alleles, and the allelism of bands could not be reliably ascertained. Because strong linkage disequilibrium exists in this population (see Results) and because these loci had low polymorphism and effectively dominant expression, they added negligible information and were discarded from further analysis. A third primer pair, IF073, amplified multiple loci with overlapping allele sizes, as indicated by non-Mendelian segregation and the presence of individuals with up to five alleles. Our attempts to redesign the primers to isolate a single locus were not successful. We therefore treated this locus as a dominant DNA fingerprint and used it to eliminate candidate males in the analysis of paternity; however, we removed this locus from analyses of genetic structure (see below).

Prior to the availability of these microsatellite markers, the mapped ramets were genotyped at seven isozyme loci by Burke et al. (2000) in order to investigate clonal structure in this and another hybrid population. We have included the Young's Coulee isozyme data of Burke et al. (2000) in the present analysis of adult population genetic structure. However, we have not repeated analyses reported in that paper except where relevant to the present study.

Analysis of Genetic Structure

Our primary goal in this study was to describe spatial genetic structure and mating patterns within the *fulva*-type phenotypic class, because matings between phenotypic classes are expected to be rare. Thus, we performed the analyses described below on each of two data sets, the complete population and the subset of *fulva*-type hybrids.

Each locus was tested independently for deviations from Hardy-Weinberg equilibrium using the exact test of Guo and Thompson (1992) computed with GenePop (Raymond and Rousset 1995). Genotypic linkage disequilibrium was evaluated by calculating the common correlation coefficient of alleles with the program LinkDos (Garnier-Gere and Dillman 1992) assuming independence of each pairwise comparison. Statistical significance was tested using the χ^2 method of Weir (1979) implemented by LinkDos. The program GenAlEx (Peakall and Smouse 2001) was used to test for spatial autocorrelation of alleles at successive two-meter distance classes, with confidence intervals generated by 999 iterations of bootstrap resampling. We also calculated the residual spatial autocorrelation not attributable to clonal structure by including a single ramet per genet positioned at the center of the genet, i.e. the mean X and Y coordinates of all ramets.

Spatial autocorrelation is a statistically powerful measure of average spatial structure, but such summary measures may obscure biologically relevant detail. We therefore included additional exploratory analyses of spatial structure. Our first method was to quantify the degree to which particular genotypic classes were spatially restricted in their distribution relative to the population as a whole. For each genet, we determined the center as described above, from which we calculated the average radial distance of ramets. We then used a t-test to compare the average radial distance within clones versus between clones, with degrees of freedom adjusted for unequal variances using SPSS 11.0 (SPSS Inc., Chicago, IL). The radial dispersal of each single-locus genotype and each allele present in at least three genets was calculated analogously.

The second approach was to use the model-based clustering method of Pritchard et al. (2000), which parameterizes the degree to which individuals in a sample are admixtures of individuals derived from genetically differentiated source populations. In the context of a hybrid

zone, the clusters can be considered subsets of individuals within which there has been random mating, and individuals of mixed ancestry can be identified. The method is implemented by the program Structure (Pritchard et al. 2000) and does not require prior knowledge of the actual number of source populations or their gene frequencies. These are inferred by clustering samples into groups that minimize Hardy-Weinberg and linkage disequilibria, without reference to their spatial location in the population. The Structure simulation was run using the admixture model with correlated allele frequencies for 100,000 iterations after a burn-in of 10,000 iterations. The correlated allele frequencies option is preferable for investigating structure in less-differentiated populations (Pritchard et al. 2000), although there was no appreciable difference in our results if uncorrelated frequencies were assumed. For the most probable value(s) of K, we then mapped individuals according to their assigned group and characterized the extent to which the inferred groups are spatially clustered using spatial autocorrelation. The posterior probabilities of K were determined by Bayes' Rule using a uniform prior probability for all values of K, although these probabilities should be considered rough approximations (Pritchard et al. 2000).

An assumption of the Structure model is that the presumed source populations are approximately randomly mating. This assumption is not likely to hold strictly for these species (e.g., Cruzan et al., 1994), and this departure from the model will tend to inflate estimates of K, the number of randomly mating groups (Pritchard et al. 2000). The objective of this analysis is not to determine the exact number of 'true' clusters, however, but to approximate the degree to which recombination has occurred in a spatial context. A second difficulty is the presence of clonal genetic structure, which can generate linkage disequilibrium among ramets even if the underlying genets are in linkage equilibrium. We therefore performed the Structure analysis on genets only, but included all ramets in the spatial autocorrelation of group identity.

Effect of Phenology on Genetic Assortative Mating

Louisiana Iris species vary in their phenology (Cruzan and Arnold 1993), which is a potentially important factor determining hybridization dynamics. Furthermore, phenology varies among artificial *I. fulva* x *I. brevicaulis* hybrids and is heritable (A. Bouck unpublished data). It is therefore possible that in nature similar genotypes overlap in flowering to a greater extent than dissimilar genotypes, thereby retarding the rate of recombination and potentially serving as a mechanism preserving novel lineages.

To examine the effect of phenology on the genetic correlation of mating pairs, we first determined the relatedness, as defined by Queller and Goodnight (1989), of flowering individuals relative to the population as a whole. The program Relatedness 5.0 (http://www.gsoftnet.us/Gsoft.html) was used to calculate relatedness values. Within the cohort of flowering individuals, the effect of phenology on genetic assortative mating was assessed in two ways. First, we examined the genetic distance between individuals relative to the date of first flowering. This phenological trait has moderately high heritability under greenhouse conditions (A. Bouck unpublished data) and may be the most relevant to mating patterns in nature, because the first flower in *I. fulva* contributes disproportionately to seed set (Wesselingh and Arnold 2003). We calculated genetic distances (Φ_{PT}, Smouse and Peakall 1999) between all flowering individuals. These values were then regressed on the pairwise difference, in number of days, between the respective dates of first flower using a Mantel test. The second method weights pairwise comparisons by the total number of mating opportunities (i.e., pairwise fertilities), which we here define as the total number of distinct pairs of male-phase and hermaphrodite-phase flowers open on a given pair of ramets on a given day. Thus, the weight for each flowering pair *ij* is calculated as (the number of flowers open on *i* times the number of flowers with

receptive stigmas on *j*) + (the number of flowers open on *j* times the number of flowers with receptive stigmas on *i*), summed over all days of overlap. This weight is symmetric with respect to each member of a pair, so that a Mantel test can be used to correlate a matrix of weights and a matrix of genetic distance. It should be noted that the *per diem* component of the weight increases geometrically with flower number, and thus mating opportunities increase nonlinearly with greater phenological overlap. We therefore compute the Mantel regression after adding one to each weight and taking the natural logarithm. Overall, this method of correlation includes more information about potential mating pairs than does the regression of genetic distance on date of first flower. However, since flowers at different locations in the inflorescence are not equally likely to set seed (Wesselingh and Arnold, 2003), the method is not necessarily a more precise way to assess whether phenology creates genetic structure in progeny arrays. Rather, the methods are complementary and we present both for completeness. The program GenAlEx (Peakall and Smouse 2001) was used to calculate genetic distances and to perform the Mantel tests.

Analysis of Mating System and Progeny Genetic Structure

The total outcrossing rate of each fruit was calculated simply as the number of apparent outcross seeds divided by the total number of seeds. The minimum number of fathers contributing to each fruit was determined manually from the total number of unique outcross alleles detected. We did not incorporate a maximum-likelihood correction for the probability of undetected outcross events (see Ritland 2002) for three reasons. First, sample sizes were small for some fruits and pollen allele frequencies were expected to differ among families, conditions that limit the robustness of maximum likelihood estimation (Ritland 2002). Second, the rate of detectable outcrossing per family was generally very high or very low, and thus correcting for cryptic

outcrossing should not change mating system estimates appreciably. Third, available estimation methods do not allow mixture of dominant and codominant markers.

We used a fractional method of paternity assignment (Schoen and Stewart 1986, Devlin et al. 1988), which is particularly appropriate for quantifying population-level parameters such as pollen flow (Devlin et al. 1988). For each flower, candidate pollen donors were identified from their observed dates of flowering. Candidate males were then tested for compatibility with each mother-offspring pair at the two codominant microsatellite loci, using the software package Cervus (Marshall et al. 1998). The remaining candidate males were then examined manually at the dominant marker IF073. Paternally derived alleles detectable in progeny were used to exclude candidate males, whereas candidate males were not excluded if their visible IF073 alleles were absent in progeny because the candidate male might be heterozygous for a recessive (null) allele. Apparent outcrossed progeny were then assigned fractionally to unexcluded males, weighted by their respective mating opportunities (as defined above) and transmission probabilities at the codominant loci. We chose not to weight candidate males by distance from the maternal parent (cf. Adams et al. 1992), because we lack sufficient empirical data to construct a robust probability distribution for interpair distance.

RESULTS

Genetic Diversity

The allelic diversity of each microsatellite locus is shown in Table 2-1. At both codominant loci, allele frequencies among ramets and among genets departed significantly from Hardy-Weinberg expectation, but in opposite directions (Table 2-2). There was a significant excess of homozygotes at IB145, which may be partly attributable to null alleles, based on null frequency

estimation by Cervus (Table 2-1) and observed segregation of null alleles in three maternal individuals (data not shown). In contrast, null alleles appeared to be rare among *fulva*-type hybrids at IF061 and homozygotes were significantly less frequent than expected. All *I. hexagona* plants were null at IF061. (The presence of null alleles reflects the difficulty of isolating microsatellite loci that amplify across the three hybridizing species, and we have noted where this might bias our interpretation.) The isozyme loci also showed no consistent pattern with respect to deviations from Hardy-Weinberg equilibrium. Six of the seven polymorphic isozyme loci assayed by Burke et al. (2000) deviated significantly from Hardy-Weinberg equilibrium; four had excess homozygosity and two had excess heterozygosity. Overall, the average *f* across loci was 0.037 for the whole population and 0.007 for *fulva*-type ramets. In calculating the average *f* we excluded locus PGI-3 because the *I. hexagona* patch and the *fulva*-type hybrids were virtually fixed for alternate alleles, creating a Wahlund effect. These results suggest that there has not been a history of strong inbreeding in this population.

Genotypic linkage disequilibria among *fulva*-type genets, as measured by pairwise correlation coefficients, are shown in Table 2-3. Approximately half of all pairwise comparisons produced correlation coefficients significantly greater than zero. The correlation coefficients were similar when all *fulva*-type ramets were included (data not shown) but more comparisons were statistically significant, presumably because of the larger sample size. While there was significant linkage disequilibrium between the two codominant microsatellite loci, no disequilibrium was detected within highly selfed fruits, indicating that these loci are not tightly linked physically (data not shown). These results indicate limited recombination among *fulva*-type hybrids as a group, either due to mating patterns or postzygotic selection.

As would be expected, greater resolution of genets was achieved by combining the genetic information of the microsatellite loci with the previous isozyme analysis of Burke et al. (2000). We identified 106 distinct multilocus genotypes, with an average of 1.86 ± 1.83 (SD) ramets per genet and a maximum genet size of 11 ramets. In comparison, Burke et al. (2000) identified 46 genets with an average and maximum size of 4.00 ± 0.50 (SD) and 19 ramets, respectively. The *I. hexagona* patch consisted of 4 genets constituting a total of 21 ramets; the remaining 102 genets were *fulva*-type hybrids.

Spatial Genetic Structure

The *fulva*-type plants showed pronounced genetic structure at small spatial scales, which declined to zero at approximately five meters. Considering all *fulva*-type ramets, r = 0.250 (P = 0.001) at two meters and r = 0.057 (P = 0.001) at four meters. At the microsatellite loci, spatial autocorrelation was greater at IF061 (r = 0.571 at two meters, P = 0.001) than at IB145 (r = 0.276 at two meters, P = 0.001). These spatial patterns are in part due to the clonal habit of these plants (Burke et al. 2000). However, after excluding multiple ramets of each genet, the strength of autocorrelation was somewhat reduced but remained highly significant. For all microsatellite and isozyme loci combined, r = 0.188 (P = 0.001) at two meters and r = 0.048 (P = 0.001) at four meters (Fig. 2-1). For IF061, r = 0.476 (P = 0.001) at two meters and r = 0.114 (P = 0.001) at four meters; for IB145, r = 0.258 (P = 0.001) and r = 0.080 at four meters (P = 0.001).

The mean radius of clonal patches was less than 1.2 meters in all but one case, in which four ramets with the same multilocus genotype were found widely dispersed in the population (mean radial distance of 20.3 meters). This extreme outlier was excluded from the analysis. The average of the mean radial distances within clones was 0.56 meters and between clones was 11.79 meters, a highly significant difference by two-tailed t-test (P< 0.0001). When the effect of

clonality was removed by including only one ramet per genet at its centerpoint, single-locus genotypes and alleles were still frequently localized to very small areas. Of the 23 single-locus microsatellite genotypes present in three or more genets, 15 (65.2%) had a mean radial dispersal of < 2 meters (Fig. 2-2). Of the 16 alleles present in three or more genets, 6 (37.5%) had a mean radial dispersal of < 3 meters.

Table 2-4 shows the probability of the *fulva*-type genotype array for different values of K, the number of admixed groups, as calculated by Structure (Pritchard et al. 2000). The value of α, which reflects the proportion of individuals of mixed ancestry (Pritchard et al. 2000), and the posterior probability of K are also shown for each run (Table 2-4). The parameter α was consistently low (0.038 - 0.055), indicating that there is little intermixing between groups (Pritchard et al. 2000). The posterior probability of K was highest for K = 5. These results showed no appreciable change across multiple independent runs of the simulation. For each value of K, ramets were assigned to the group that formed the largest part of their inferred ancestry (see Methods) and their spatial locations marked accordingly. Figure 2-3 illustrates the location of ramets in each group for K = 5. Individual ramets assigned to a group based on a fraction of ancestry less than 0.8 are marked with an X in Figure 2-3. For the studied values of K, these clusters were strongly localized. For example, the spatial autocorrelation of group identity for K = 5 was 0.766 at 2 meters and 0.380 at 4 meters.

Genetic Correlations with Phenology

Overlapping phenology was not associated with increased genetic relatedness of potential mating pairs. Flowering ramets were not more related to each other on average than to the population as a whole (R = 0.1016 ± .0882 [SE], $P > 0.1$) based on the combined microsatellite and isozyme data. Moreover, among flowering individuals, the order of flowering did not produce genetic

correlations between potential mating pairs. The Mantel regression of pairwise difference in date of first flower on genetic distance was not significantly different from zero (r = 0.10, $P = 0.140$). The Mantel regression of total pairwise mating opportunities on genetic distance was also not significant (r = -0.22, $P = 1.000$). These regressions were also not significant when considering *fulva*-type hybrids alone (data not shown). The probability per flower that pollen drawn at random from co-flowering pollen donors would produce an apparent outcrossed zygote was generally high (mean = 75%).

Mating System

A total of 325 out of 494 seeds germinated and were successfully genotyped for at least one microsatellite locus (65.8%). The population mean apparent selfing rate was high (s = 0.677), but the distribution of selfing estimates was bimodal, with most fruits highly selfed or highly outcrossed. All but three seeds produced by *I. hexagona* fruits were compatible with selfing, but because all flowering *I. hexagona* had the same microsatellite genotypes, outcrossing among these genets could not have been detected. We can conclude only that matings between the two phenotypic classes is rare. Considering only the *fulva*-type plants, the mean apparent selfing rate was considerably lower (s = 0.478).

Not surprisingly, given the rate of selfing in this population, the total progeny array was significantly inbred (Table 2-2). Even among outcross progeny, there was a significant excess of homozygotes at IB145 ($f = 0.288$, $P < 0.0001$) but not IF061 ($f = -0.023$, $P = 0.214$). However, three fruits were heterozygous for a null allele at IB145 as evidenced by their segregation patterns, which inflates the IB145 inbreeding coefficient considerably. Excluding these fruits reduces f to 0.146 but the estimate remains significantly greater than zero ($P < 0.0001$).

Significant linkage disequilibrium between these loci ($P = 0.004$) was maintained in the outcross progeny array, which is probably due in part to the low number of fathers per fruit. By genetic exclusion, the minimum number of fathers was 1 to 3 per fruit, with a mean of 1.4.

Patterns of Paternity

The number of outcross progeny for which a single, unique candidate male was identified was 10 out of 105 (9.5%) detectable outcross events. All censused males were excluded for 57 (54.3%) of these detectable outcross events; these progeny were attributed to pollen flow from outside the population. The remaining 38 outcross events (36.2%) were assigned fractionally to all unexcluded candidates as described in the Methods.

The average number of seeds sired per flowering individual was 1.88 ± 1.76 (SE). Eighteen out of 24 flowering individuals were assigned outcross paternity (75%); this value is likely to be substantially inflated by the fractional method of paternity assignment (Devlin et al. 1988). The distance between mating pairs was relatively high, and a large fraction of outcross pollen was derived from outside the censused population (i.e., "Gene flow" distance class, Fig. 2-4). Although fractional paternity assignment might in principle distort the true distance of pollen movement, for example by overestimating the contribution of more distant candidate males, this bias is unlikely to strongly affect these results. This is due to the strong spatial structure of the population, a consequence of which is that unexcluded candidate males were generally in close proximity, in which case the weighted average of the group of pairwise distances was roughly the same as each individual distance. However, this spatial structure is itself a source of bias, because matings over shorter distances are less likely to be identified as outcrossing. Nonetheless, the linkage disequilibrium among marker loci and the strong correlation of selfing within fruits (c.f. Cruzan et al. 1994) argue that the available markers

provide a suitably accurate estimate of the mating system. Regardless of this uncertainty with respect to the frequency distribution of pollen movement, it is clear that a significant portion of pollen is transported over scales of tens of meters.

DISCUSSION

Comparison of Adult and Progeny Structure

In this study, we described patterns of genetic structure in a Louisiana Iris hybrid zone at adult and seed life-history stages. Overall, there was greater genotypic diversity among *fulva*-type hybrids at Young's Coulee than we anticipated based on previous work (Burke et al. 2000). The relatively large number of genotypes (102 multilocus genotypes among 176 mapped ramets) and small size of clones (1.73 ± 1.56 [SD] ramets per genet) among *fulva*-type plants indicate that recruitment from seed is frequent. Genetic structure was nonetheless very strong at short spatial distances (< 5 m). This structure was partially independent of clonality, as evidenced by 1) the significant spatial autocorrelation among genets and 2) the localization of many single-locus genotypes and alleles. These observations are not expected based on clonal propagation per se, as there is no reason that genets separated by considerable distances should not share alleles if the original seedlings were derived from outcross pollen flow. Instead, the data indicate low effective dispersal of both pollen and seed, probably in conjunction with selfing.

Although the spatial distribution of alleles of adults indicates a small genetic neighborhood, genotypic proportions show no evidence of historically strong inbreeding: the mean *f* across loci (excluding PGI3) among *fulva*-type hybrids is 0.007. Using Ritland's equilibrium-model estimation of inbreeding depression (Ritland 1990), we obtain an estimate of inbreeding depression between seedling and adult life-history stages that approaches unity (D =

0.987). While it is possible that the mating system observed in this reproductive season is atypical, studies of pure *I. fulva* have also shown a mixed-mating system with bimodal selfing rates (Cruzan et al. 1994). The population outcrossing rate estimated by Cruzan et al. (1994) was 0.47 – 0.67, depending on estimation method, values close to our own (t = 0.52). Thus, it appears that in this population, *fulva*-type hybrids have a mating system similar to that of the pure species. A second concern is whether an assumption of equilibrium between mating system, selection, and adult genetic structure is appropriate for a given population, especially for potentially long-lived organisms. Although theory indicates a fairly rapid approach to equilibrium under the mixed-mating model (Clegg 1980), our results indicate that the assumption of random mating among the outcross fraction is violated due to significant deviations from Hardy-Weinberg and linkage equilibrium among outcross progeny. We therefore treat the quantitative estimate of D with appropriate skepticism, but the qualitative result is not altered by this uncertainty. That is, homozygote progeny are not present in the adult population in proportion to the rate at which they are formed, by some considerable margin. Whether this is due to traditional mechanisms of inbreeding depression (e.g., exposure of deleterious recessive alleles or loss of heterotic interactions) or due to genetic factors peculiar to hybrid zones (such as Dobzhansky-Muller epistatic interactions) is unknown.

Not surprisingly, linkage disequilibrium was moderately strong among ramets in the population as a whole. We expected to find some level of linkage disequilibrium because it is an immediate consequence of the mixing of genetically differentiated groups, particularly in interspecific hybrid zones where the action of recombination may be opposed by natural selection against particular hybrid genotypes (Dobzhansky 1970). Linkage equilibrium is also

opposed by selfing and biparental inbreeding (Haldane 1949). However, it is somewhat surprising that even among the *fulva*-type hybrids there is a low rate of recombination.

When we used Structure to identify groups of *fulva*-type ramets in approximate Hardy-Weinberg and linkage equilibrium, the inferred groups were strikingly localized in space. For example, the red and orange groups represent large groups of genets between which there is a narrow border a few meters wide. This boundary appears at K = 2 and remains largely intact across a wide range of K (data not shown). Whether this is due to demographic factors such as limited seed dispersal or postzygotic selection is unknown, but the proximity of the groups and the fact that plants in each group flowered simultaneously (data not shown) suggest a role for the latter. The blue group is also noticeably confined, but it appears to have greater clonal structure than on average. For K = 5, the cluster consists of 8 genets and 24 ramets, including two clones of 5 and 11 ramets. Interestingly, no plants in this group flowered during the study year. Other groups identified by Structure show less pronounced spatial localization. We draw two general conclusions from these results: 1) recombination among *fulva*-type hybrids has occurred within roughly contiguous blocks over restricted spatial scales, and 2) recombination occurs infrequently between these groups. While these results were presaged by and are consistent with the other spatial analyses we performed, the Structure analysis integrates a greater amount of information and identifies specific genets that make up genotypic clusters or are of mixed ancestry.

Inference of the true value of K, the number of groups, is based primarily on two factors, the posterior probability of the data set given various hypothesized values of K and the biological context in which the groups are interpreted (see Pritchard et al. 2000 for a discussion). Ideally, the calculated probability of the data given K has a clear maximum for some biologically

plausible value of K, which should be a small number for most scenarios of admixture and hybridization. In the present case, the posterior probability was highest for K = 5. However, violations of the clustering model are possible (e.g., inbreeding and null alleles) and these will tend to inflate the estimate of K. It is probably more useful to consider a range of plausible K and the commonalities within that range with respect to the spatial patterns of recombination and admixture. It should be noted that when there is little real structure in a sample, the program tends to identify groups approximately equal in size and the value of α is unstable within runs (Pritchard et al. 2000). In our analysis, neither of these phenomena was observed (data not shown), which provides additional support for the detected structure.

One caveat of this analysis is the limited resolution afforded by the available genetic markers. Although our results are based on a moderate number of loci, effectively eight given that the *fulva*-type genets are virtually fixed at PGI-3, this number of independent markers argues forcefully against free recombination at least in a large portion of the genome. A separate issue is the effect of low polymorphism at some loci on the power of our study to detect existing structure. While a more distinct spatial pattern of genetic structure may well be detected with the addition of more polymorphic markers, the reverse finding - that there is less spatial structure - seems highly improbable. Low polymorphism is inherently conservative with respect to the statistical analysis of genetic structure because it makes genetically distinct individuals appear similar.

Effect of Phenology on Genetically Correlated Mating

Patterns of adult genetic structure in hybrid zones reflect both the genetic composition of the population founders and subsequent patterns of effective mating. Whether mating patterns are due to prezygotic or postzygotic factors requires direct investigation of the processes that

determine mating pairs. In this study we examined the phenology of the flowering cohort to determine how variation in the pollen pool might structure mating pairs. Because the phenology of *I. fulva* x *I. brevicaulis* hybrids is heritable under greenhouse conditions (A. Bouck unpublished data), we hypothesized that coflowering individuals in hybrid zones may be more related to each other on average than to the population as a whole. If so, this would result in genetic correlations among mating pairs, thereby producing clusters of progeny with similar genotypes if seed dispersal is limited. Although differences in microhabitat variables such as soil moisture and shade are also expected to contribute to phenology, previous studies (Cruzan and Arnold 1993; Johnston et al. 2001) found significant associations between genotype and microhabitat in another Louisiana Iris hybrid zone. If generalizable, this covariation should tend to reinforce the genetic correlation of mating pairs due to phenology rather than dilute it. However, in this study we did not find any evidence that co-flowering individuals were genetically correlated at the population level. Yet average values of genetic relatedness or distance may be misleading – although on average coflowering individuals were not genetically correlated, coflowering *near neighbors* typically shared similar or identical microsatellite genotypes. Thus, if as suggested by studies of pollinator behavior (Wesselingh and Arnold 2000) matings are frequently between near neighbors, biparental inbreeding would be accordingly more common. It is also possible that other potential mechanisms of assortative mating not investigated in this study, such as quantitative variation in floral morphology among *fulva*-type hybrids, may be important in Louisiana Iris hybrid zones.

While phenology did have an effect on the genotypic distribution of progeny in an *I. fulva* x *I. brevicaulis* hybrid zone (Cruzan and Arnold 1994), the mean flowering time of these species differs by several weeks and – as has been observed in controlled crosses – there is segregating

variation for relative flowering time among their hybrids (A. Bouck unpublished data). Since the species *I. fulva* and *I. hexagona* have similar phenologies in nature (Arnold et al. 1993), it is possible that genetic variation for flowering time could only come from the *I. brevicaulis* genome. Thus, there may be insufficient segregating variation for phenology in this population to effect a correlation between genetic similarity and flowering time.

Gene Flow between I. hexagona and fulva-type Hybrids

The low level of pollen flow between *I. hexagona* and *fulva*-type hybrids (~1%) accords with our expectation based on previous work (Hodges et al. 1996). Furthermore, when the *I. hexagona* genets are included in the Structure analysis, only one *fulva*-type genet with a clearly recombinant genotype clusters with them, and only for some values of K (data not shown). This indicates that intermediate genotypes are rare among the adult population as well. Collectively, these observations justify the analysis of *fulva*-type hybrids as a distinct subpopulation. The rarity of phenotypic and genotypic intermediates is characteristic of Louisiana Iris hybrid zones, a pattern that appears to be stable based on multi-year and multi-site data (Cruzan and Arnold 1993, Cruzan and Arnold 1994). This probably reflects both the extensive barriers to F_1 seed formation in these species and some form of hybrid breakdown. Although experimental studies have documented high vegetative and sexual fitness of F_1's (Burke et al. 1998a, Emms and Arnold 1997), viability selection against intermediate genotypes in the F_2 generation has also been documented (Burke et al. 1998b).

The Distribution of Progeny Genotypes

Although the observed rate of outcrossing was insufficient to break down genotypic disequilibria at the microsatellite loci within the overall progeny array, paternity assignment indicated that pollen did move over substantial distances. Approximately 70% of assigned outcross paternity

was to individuals >20 meters from the maternal plant. Furthermore, a large fraction of these outcross progeny (54.3%) were fathered by individuals outside the censused population, as determined by genetic exclusion. The relatively high dispersal distance of pollen could be interpreted in two ways. One possibility is that because the probability of detecting outcross events between near neighbors is generally lower, the true frequency of pollen flow over short distances is higher than evidenced by these data and thus actual outcrossing rates are underestimated. This view is supported by observations of pollinator behavior in this system (Emms and Arnold 2000, Wesselingh and Arnold 2000), which show a strong bias toward near-neighbor transitions. However, the alternative explanation, that long-distance pollen flow is fairly common, is supported by the emerging generality of this finding in plant paternity analyses (Ennos 2001), and by evidence of introgression via pollen flow between disjunct populations of Louisiana Iris (Arnold et al. 1992).

The observed patterns of pollen movement and negligible inbreeding coefficients among adults are contrary to the strong spatial structure of genets, in which single-locus genotypes and individual alleles are frequently limited to small regions of the population. Our results also imply a low rate of seed disperal, which is somewhat surprising given that *I. fulva* and *I. hexagona* seeds have a corky, buoyant seed coat that is presumably adapted for water dispersal. It is possible that the patchiness of adult genotypes reflects spatially episodic recruitment of seedlings dominated by a few nearby maternal individuals. Alternatively, there may be undetected environmental variables that favor particular genotypes in different patches. These hypotheses are purely conjectural; the simplest explanation remains that most seeds are dispersed near the maternal individual despite the potential for long-distance transport by water.

A difficulty with our analyses that merits attention is the presence of null alleles at the microsatellite loci, which obscure the true genotype of individuals and may lead to false inferences regarding genetic structure and paternity. In our sample there was a high frequency of null alleles at IB145 among *fulva*-type individuals, and all *I. hexagona* individuals were null at IF061. However, the estimated null frequency at IB145 was much lower within the cohort of flowering individuals. Furthermore, *I. hexagona* are *a priori* not expected to mate frequently with *fulva*-type plants (e.g., Hodges et al. 1996) and are also quite distinct at IB145 and IF073 in this population (data not shown). Thus, we do not believe that null alleles significantly impact the paternity results. With respect to the spatial analyses, null alleles at IB145 can cause both positive and negative deviations from actual genetic distance, as defined by Smouse and Peakall (1999), for individual pairwise comparisons. It is therefore difficult to assess the cumulative effect of nulls on the spatial autocorrelation statistic at this locus, but the impact on spatial autocorrelation across all loci should be minimal. However, null alleles should tend to inflate the estimation of K in the Structure analysis by creating false deviations from Hardy-Weinberg equilibrium, and we caution that this bias may be present in our data.

Despite these caveats, the overall patterns of paternity and genetic structure among *fulva*-type hybrids indicate that neither outcrossed progeny nor homozygous selfed progeny are recruited at the frequency with which they are formed. Rather the contrast between adult and progeny structure suggests that progeny with genotypes similar to adults are the most successful. It bears pointing out that this conclusion is qualitatively unchanged if actual outcrossing rates differ somewhat from our estimates, either due to inherent ambiguities in estimation or to natural variation from year to year. Although we do not know the basis of postzygotic selection in this population, such mechanisms have been an area of active investigation in Louisiana Irises.

Marker-based studies of experimental hybrids have uncovered selection against certain hybrid genotypes in the F_2 (Burke et al. 1998b) and BC_1 generations (A. Bouck unpublished data), as indicated by segregation distortion of marker loci and negative epistatic interactions between loci. In addition to nuclear genetic incompatibilities, cytonuclear incompatibilities have been identified in these hybrid classes (Burke et al. 1998), which might contribute to the maintenance of linkage disequilibrium and heterozygosity in this hybrid zone.

These conclusions have important consequences for the establishment of recombinant lineages that are more fit than the parental types in some habitats. Grant (1971) and others have pointed out the Sisyphean dilemma of hybrid novelty, which is that adaptive gene combinations brought into existence by recombination will be undone by the same process. Our results suggest that partial reproductive isolation can occur over small spatial scales relative to pollen flow in Louisiana Iris hybrids. Thus, in this system, life-history characteristics such as selfing and asexual propagation do appear to contribute to the stabilization of hybrid lineages as predicted by theory (e.g., Grant 1971, McCarthy et al. 1995), but postzygotic selection against subsequent recombination is also significant. Future research will need to address a separate but equally important question, which is whether novel recombinant types can spread geographically and demographically from their points of origin.

ACKNOWLEDGEMENTS

We thank M. Bulger and J. Vogel for invaluable field assistance. We also thank A. Bouck and E. Kentner for useful discussions and technical advice on the analysis of the data and A. Bouck and R. Mauricio for valuable comments on the manuscript. The manuscript was further improved by the comments of L. Galloway, D. Howard, and an anonymous reviewer. This work was

supported by National Science Foundation Training Grant DBI 9602223 (RSC) and National Science Foundation Grants DEB-0074159 (MLA) and DEB-0345123 (MLA). RSC was also supported by a University of Georgia Presidential Fellowship.

Table 2-1. Allelic diversity of microsatellite loci. Expected heterozygosities (H_e) were calculated using Cervus (Marshall et al. 1998). Null frequency in the total population was estimated using GenePop (Raymond and Rousset 1995). Null frequency among flowering individuals was estimated manually from segregation patterns in fruits. Only the total number of alleles amplified by the IF073 primer pair is presented, because at least two overlapping loci are present and each allele is treated as a dominant locus for use in paternity analysis (see text).

	Total Population				*Flowering Individuals*			
Locus	Alleles	H_e	H_o	Null Frequency	Alleles	H_e	H_o	Null Frequency
IB145	10	0.849	0.741	0.101	9	0.840	0.920	0.042
IF061	9	0.810	0.894	0.026[1]	8	0.781	0.913	0.125[1]
IF073	7	-	-	-	7	-	-	-

[1]All *I. hexagona* genets are null at locus IF061.

Table 2-2. Deviations from Hardy-Weinberg equilibrium. Inbreeding coefficients calculated by GenePop (Raymond and Rousset 1995) using the method of Weir and Cockerham 1984.

	f				
Locus[1]	All genets	*fulva*-type genets	*fulva*-type ramets	All seedlings	Outcross seedlings
IB145	0.205**	0.215**	0.148**	0.366**	0.288**
IF061	-0.075**	-0.075**	-0.105**	0.216**	-0.023
F16	-0.091	-0.081	-0.070	-	-
FE-1	0.353**	0.389**	0.290**	-	-
6PGD-1	-0.240**	-0.254**	-0.260**	-	-
6PGD-2	0.238*	0.231*	0.164*	-	-
PGI-3	1.000**	1.000**	1.000**	-	-
SKDH	0.220**	0.167**	0.280**	-	-
TPI-1	-0.420**	-0.433**	-0.394**	-	-
Average[2]	0.024	0.020	0.007	0.291	0.133

[1]F16 = Fructose-1,6-diphosphate, FE = Fluorescent esterase, 6PGD = 6-phosphogluconate dehydrogenase, PGI = phosphoglucoisomerase, SKDH = shikimate dehydrogenase, TPI = triose-phosphate isomerase.
[2]Excluding PGI-3, see text.
*$P < 0.05$, **$P < 0.01$

Table 2-3. Linkage disequilibrium among *fulva*-type genets as measured by common correlation coefficients. Correlation coefficients were calculated by LinkDos (Raymond and Rousset 1995) and statistical significance was tested using the method of Weir 1979, assuming independence of each comparison.

	TPI-1	SKDH	PGI-3	6PGD-2	6PGD-1	FE-1	F16	IF061
IB145	0.108	0.065	0.086***	0.105*	0.066	0.168***	0.104	0.137***
IF061	0.156***	0.143**	0.101***	0.129*	0.063	0.148***	0.057	
F16	0.103	0.015	0.059	0.150	0.024	0.084		
FE-1	0.192*	0.107	0.225***	0.131*	0.011			
6PGD-1	0.247**	0.021	0.125	0.495***				
6PGD-2	0.224***	0.342***	0.065					
PGI-3	0.095	0.175**						
SKDH	0.130							

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

Table 2-4. Structure (Pritchard et al. 2000) simulation results for *fulva*-type genets for given values of K. The parameter α reflects the proportion of individuals of mixed ancestry. Parameter estimates are obtained by Markov-chain Monte Carlo simulation with 100,000 runs after a 10,000 run burn-in. The posterior probability of K is calculated using Baye's rule with a uniform prior probability for all K. See text for details.

Number of Groups (K)	ln(Probability of K)	α	Posterior Probability of K
2	-1376.0	0.055	0.000
3	-1324.4	0.047	0.000
4	-1289.0	0.041	0.000
5	-1264.0	0.041	0.953
6	-1267.0	0.040	0.047
7	-1299.6	0.040	0.000
8	-1383.3	0.039	0.000
9	-1429.8	0.038	0.000
10	-1614.0	0.039	0.000

Figure 2-1. Spatial autocorrelation across all microsatellite and isozyme loci among *fulva*-type genets, as calculated by GenAlEx (Peakall and Smouse 2001). In the computation of this statistic, each genet is represented only once and its spatial coordinates are defined by the average x- and y-coordinates of all ramets of the genet. The solid, bold line represents the autocorrelation coefficient and the dashed lines represent the bootstrapped 95% confidence interval.

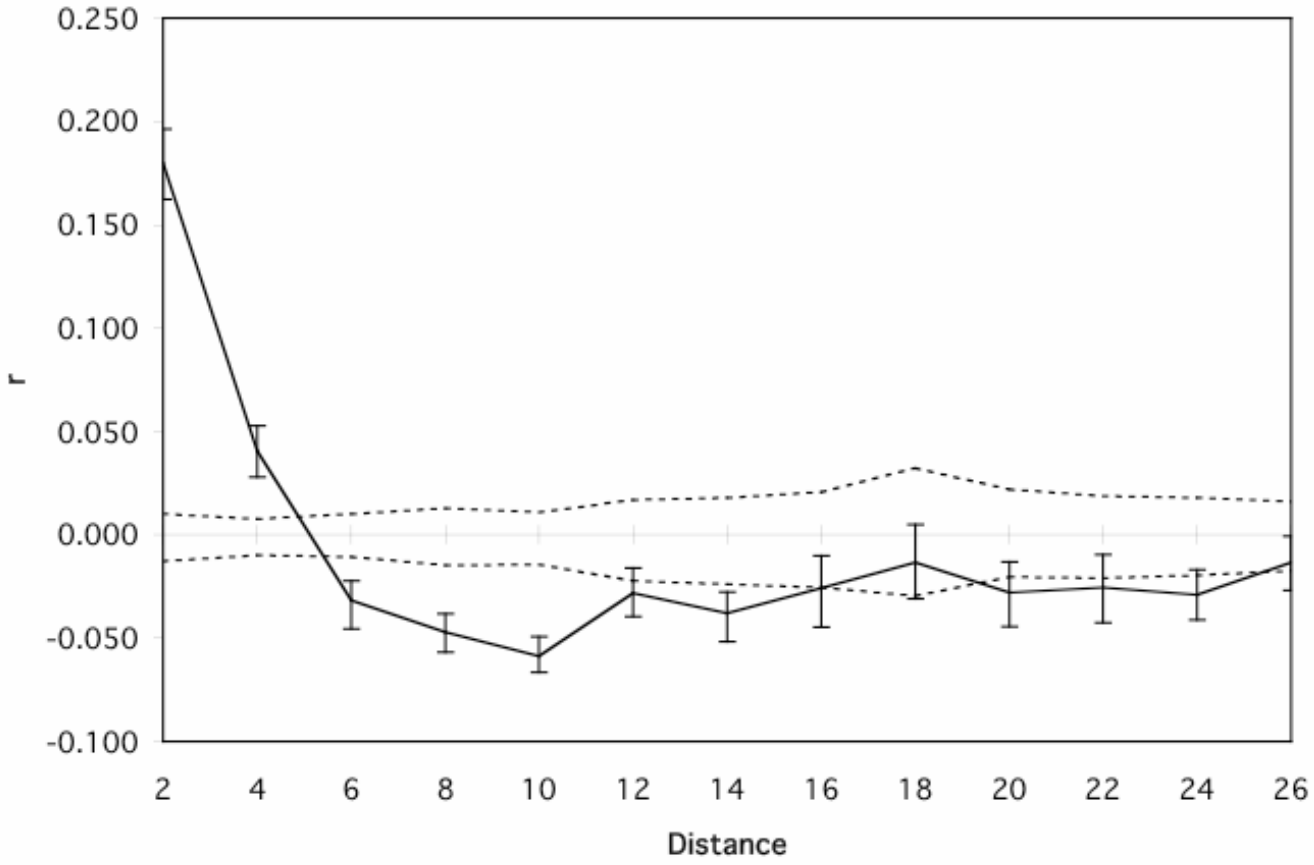

0.250
0.200
0.150
0.100
0.050
0.000
-0.050
-0.100
r
2
4
6
8
10
12
14
16
18
20
22
24
26
Distance

Figure 2-2. Average radial dispersal of all single-locus genotypes of the microsatellite loci IB145 and IF061 that are present in three or more individuals. Average radial dispersal is the distance in meters from the center of each genet to the center of all genets sharing the same single-locus genotype.

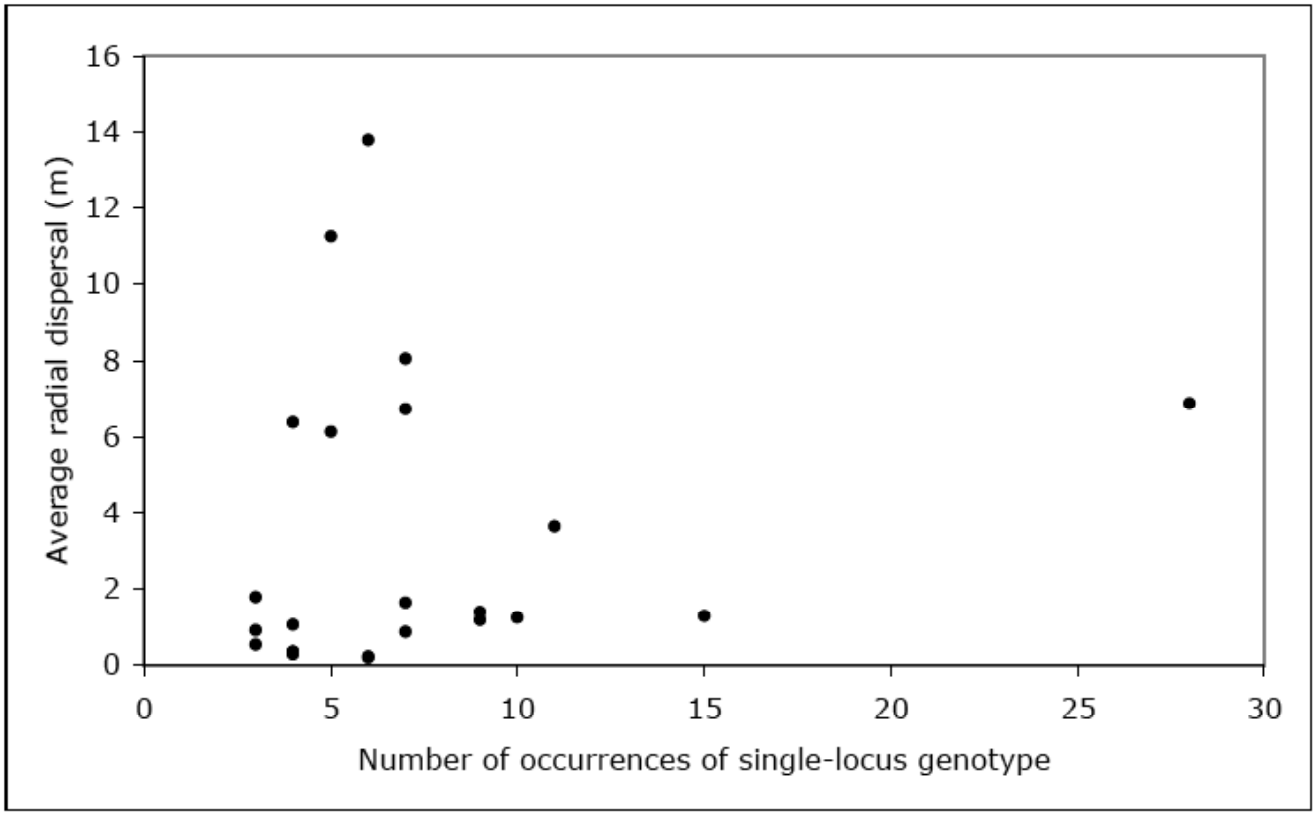

Average radial dispersal (m)
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
Number of occurrences of single-locus genotype

Figure 2-3. Map of ramets by group identity as determined by Structure (Pritchard et al. 2000) for K = 5. Each color represents a specific group. Solid squares represent individuals with an estimated percent ancestry of ≥ 80% from the specified group. Squares with an X derive the greatest fraction of their ancestry from the specified group but the estimated percentage is < 80%; these individuals are considered to be of mixed ancestry. *I. hexagona* ramets are represented by open black triangles. Axes are in meters.

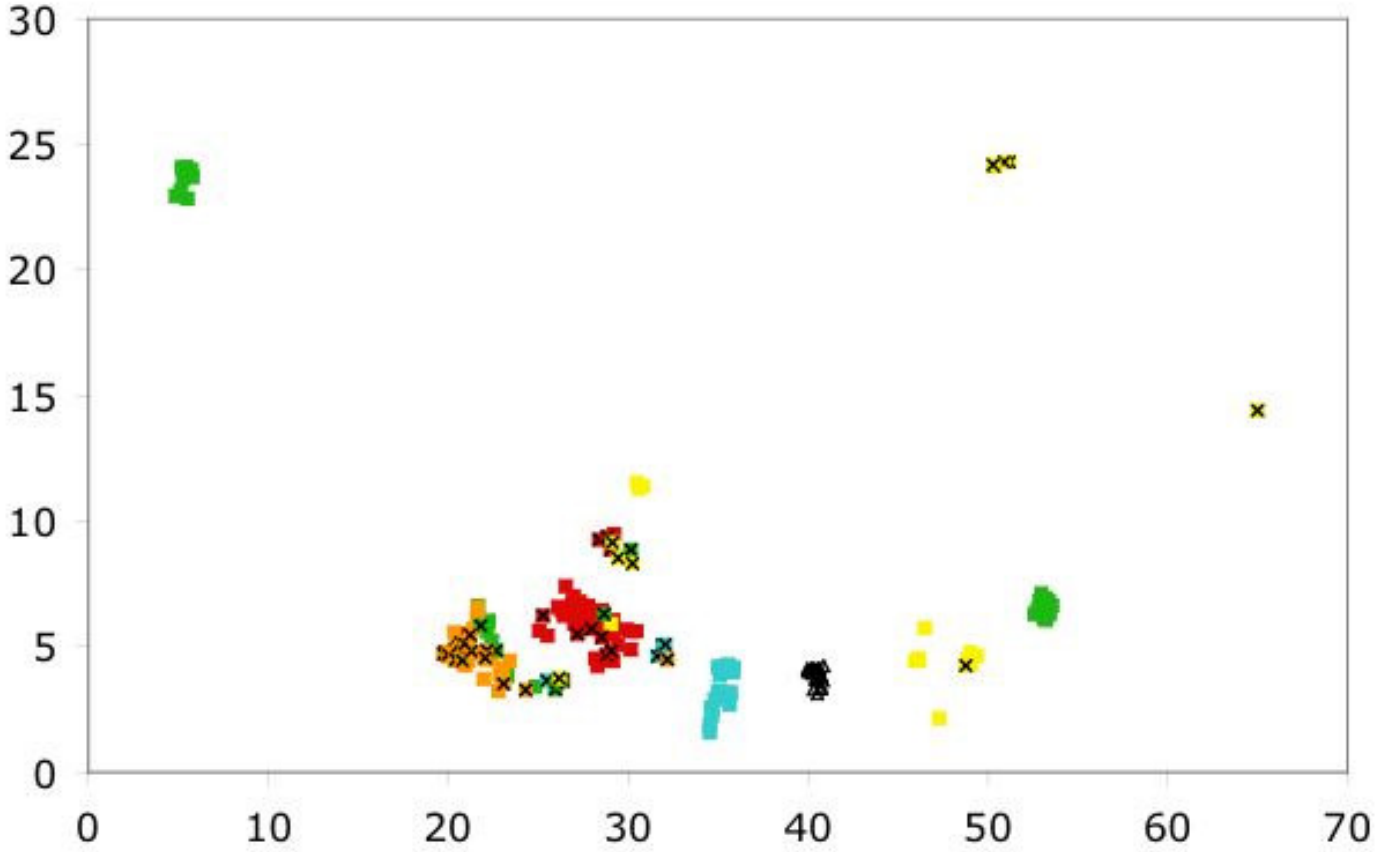
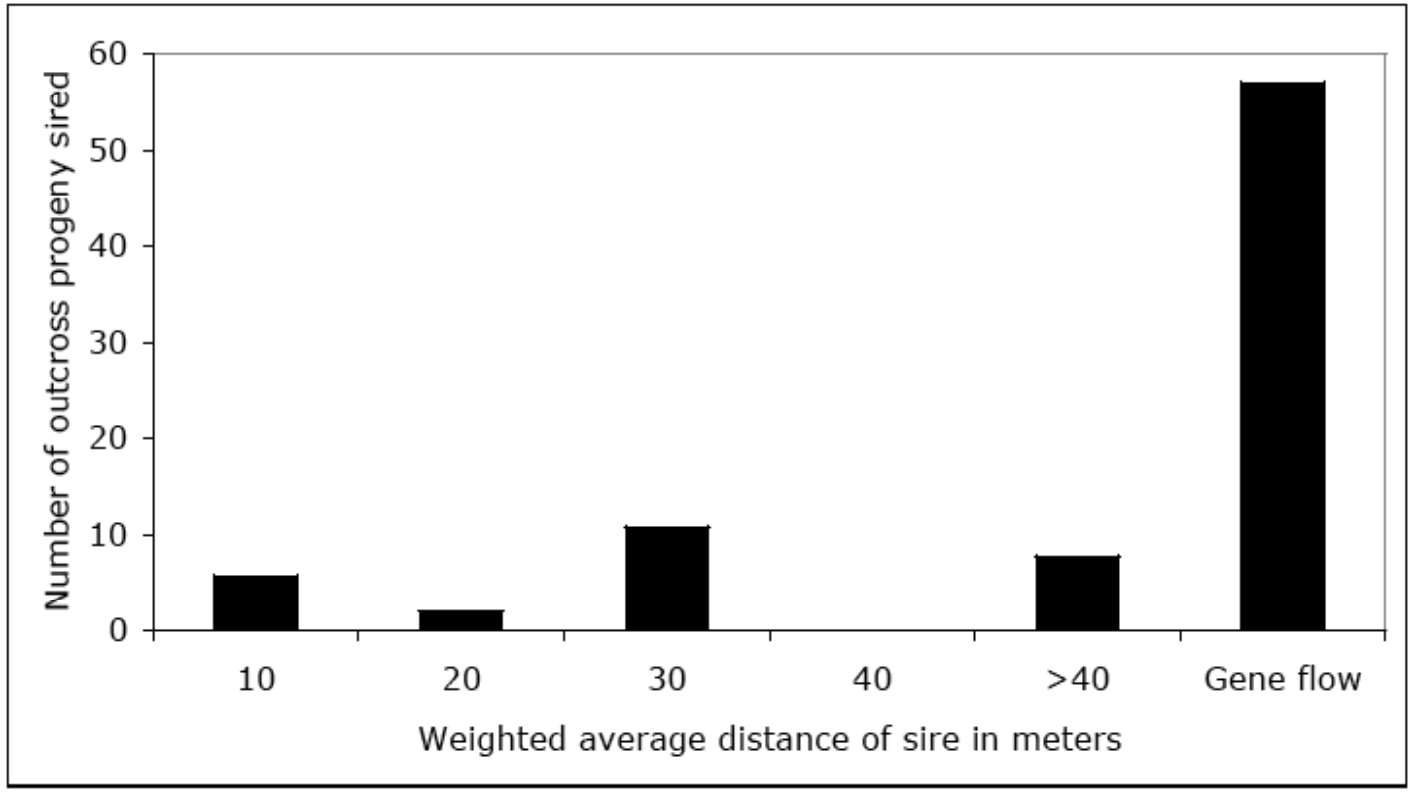

30
25
20
15
10
5
0
0
10
20
30
40
50
60
70

Figure 2-4. Frequency of detected outcross pollen movement by distance class. Pollen movement is based on fractional paternity assignment as described in the Methods. Therefore, assignment of individual outcross events is made fractionally to each distance class according to the weights of each candidate father in that distance class.


Number of outcross progeny sired
60
50
40
30
20
10
0
10
20
30
40
>40
Gene flow
Weighted average distance of sire in meters

CHAPTER 3

DEVELOPMENT OF TRANSPOSON-SSAP MARKERS FOR POPULATION-GENETIC ANALYSIS OF *IRIS MISSOURIENSIS*[2]

[2] Cornman, R. S. and M. L. Arnold. To be submitted to *Molecular Ecology Notes.*

ABSTRACT

The widespread application of technologies for isolating informative and cost-effective genetic markers has led to a boon in population-genetic studies of non-model organisms. Here we describe the development of transposon sequence-specific amplified polymorphisms (transposon-SSAPs) in the herbaceous perennial *Iris missouriensis* (Iridaceae). Transposon-SSAPs are dominant fragment-size markers similar to amplified fragment-length polymorphisms (AFLPs), but are strictly nuclear and can be more robust for large genomes or that contain high numbers of repetitive elements. Our cloning strategy identified two groups of long-terminal repeat (LTR) retroelements of the iris retroelement (IRRE) type originally defined by Kentner et al. (2003). One group was a divergent but recognizable form of the IRRE1A class, whereas the other group was a novel IRRE sequence we termed IMRE1 (*Iris missouriensis* retroelement 1). Primers homologous to conserved regions of these elements generated repeatable, polymorphic, and informative markers for population-genetic analysis. We also developed a phylogeny of representative species of subgenus *Limniris* based on three chloroplast regions to investigate the distribution and evolutionary tempo of these two elements. *I. missouriensis* occupied the basal position of the resulting phylogeny, which was well supported by bootstrap resampling. Sequences from both groups were widespread in *Limniris*, but there was clear evidence that both groups have amplified within *I. missouriensis* since its divergence from other *Limniris*. The geographic structure of transposon-SSAP polymorphism was strikingly different, most likely due to a high mutation rate from IMRE1-based markers, indicating that these elements are significant contributors to genetic variation and possibly mutational load.

INTRODUCTION

The advent of genetic markers that are relatively easy to isolate and frequently polymorphic has led to an explosion of population-genetic studies of non-model organisms. Anonymous DNA length polymorphisms such as randomly amplified polymorphic DNA (RAPDs, Williams et al. 1990) or amplified fragment-length polymorphisms (AFLPs, Vos et al. 1995) are frequently used because they can be highly polymorphic and require no prior sequence information on the taxon of interest. While these markers are generally dominant and therefore less informative per locus than codominant loci, this weakness is offset by the potentially large number of loci that can be generated, allowing genome-wide scans of diversity.

Despite the utility and popularity of AFLPs, it has become apparent that this approach can be less reliable for large genomes (Fay et al. 2004). Furthermore, AFLP markers may concentrate in centromeric regions, which are rich in repetitive sequence, limiting the utility of these markers for some applications (Vuylsteke et al. 1999). Another potential difficulty of AFLP markers is that nuclear and organellar polymorphisms are indistinguishable unless mode of inheritance can be inferred from pedigrees. For these reasons, transposon-based sequence-specific amplified polymorphisms (transposon-SSAPs or 'transposon display'; van der Broeck et al. 1998) are a useful alternative to AFLPs. The transposon-SSAP methodology is a derivative of the standard AFLP protocol (Vos et al. 1995), in which known sequences are ligated to anonymous restriction sites in the genome. Unlike AFLPs, however, transposon-SSAP uses only one adapter primer. The other primer in the PCR anneals to a transposon sequence that has been previously characterized. Each amplification product therefore concatenates a common region at the 5' or 3' end of the transposon and the unique 'host' genomic sequence in which each transposon copy is embedded. Because most mobile elements are absent from organellar

genomes (Kumar and Bennetzen 1999), transposon-SSAPs reflect only nuclear variation. Individual transposon-SSAP loci can also be sequence-characterized for the unique host sequence and potentially converted to co-dominant loci, as is true of other dominant DNA markers (Paran and Michelmore 1993). However, the chromosomal distribution of transposon families can be strongly skewed: some are known to be concentrated in centromeric regions (Kumar and Bennetzen 1999) and are therefore less useful for marker applications such as genetic mapping. Although transposon-SSAP requires knowledge of specific tranpsoson sequences, the rapid expansion of genomic databases has enabled the isolation of useful sequences for a wide range of taxa.

In this paper, we describe the development of transposon-SSAP markers for the *Iris* species *I. missouriensis* (*Liliopsida: Iridaceae)*. Although no prior genomic information was available for this species, related species have relatively large genomes (diploid nuclear contents of ~20 pg) and a high abundance of repetitive elements (Kentner et al. 2003). Furthermore, both the standard AFLP protocol (Vos et al. 1996) and a modified protocol (using the six-cutter *Hin*dIII and appropriate adapters in place of *Mse*I) produced few bands and minimal polymorphism in *I. missouriensis* (R.S.C. unpublished data). We therefore used the cloning approach developed by Kentner et al (2003) to isolate and characterize retroelement sequences present in *I. missouriensis*. The Iris retroelement families IRRE1 and IRRE2 have been shown to be abundant, polymorphic, and dispersed across linkage groups in related species (Kentner et al. 2003, Bouck et al. 2005), indicating their utility for population-genetic analysis of *I. missouriensis*. The goals of our study were to 1) identify IRRE transposon-SSAP primers suitable for population-genetic studies in this species and possibly related taxa; 2) assess the

reproducibility of multilocus genotypes; 3) and investigate the phylogenetic distribution of transposable elements identified by our screen.

MATERIALS AND METHODS

Samples of *I. missouriensis* were collected by R.S.C. in the summer of 2002 near Marshall Pass, Saguache County, Colorado. *I. missouriensis* is the only iris native to Colorado and its identity was confirmed by chloroplast haplotype analysis (Chapter Three). Genomic DNA was extracted using a hexadecyltrimethyl-ammonium bromide (CTAB) and chloroform extraction protocol (Doyle and Doyle 1987) and pooled from four individuals collected near Marshall Pass (Saguache County), Colorado for use in the cloning procedure. Additional sampling of *I. missouriensis* for population-genetic analysis was performed as described in Chapter Three.

A PCR-based cloning strategy was used to obtain sequences of two *gypsy*-type long-terminal repeat (LTR) retrotransposon families that are specific to *Iris* (Kentner et al. 2003). Degenerate primers developed by Kentner et al. (2003) were used to amplify a defined region of IRRE families spanning approximately 450 bp from the 3' end of the integrase gene through approximately the first 300 bp of the 3' LTR. PCR products were cloned using the TOPO cloning vector (Invitrogen) and seventeen cloned fragments were subsequently isolated and sequenced. Sequences were aligned using Sequencher 4.2 (Gene Codes Corp.). Subfamilies were identified based on percent similarity within and among the clusters of a neighbor-joining dendrogram. These sequences were further compared with the IRRE family and subfamily designations of Kentner et al. (2003).

Regions of sequence conservation within each subfamily were identified by eye and candidate primers were evaluated for annealing temperature, GC content, and secondary

structure using IDT Inc.'s online software (www.idtdna.com/SciTools/SciTools.aspx). To identify transposon-SSAP loci, the following protocol was observed. Approximately 100 ng of genomic DNA was digested with 2.4 units of EcoR1 (Promega) in the supplied buffer and with 0.5 µg of bovine serum albumen (BSA) in a total volume of 25 µL. This mixture was incubated at 37° C for three hours and then at 65° C for 15 minutes to inactivate the enzyme. Ligation reactions were performed in the same microcentrifuge tube by adding 50 nanomoles of each adapter primer, 1 unit of T4 DNA ligase, 0.5 µg of BSA, and reaction buffer in a total volume of 40 µL. Ligation reactions were performed at room temperature for three hours or at 4° C overnight. The product of these reactions was then diluted four-fold and 3 µL was used as the template in the pre-amplification PCR, along with 10 picomoles of the standard EcoR1 pre-amplification primer (Vos et al. 1996), 10 picomoles of transposon primer, 125 µM of each dinucleotide triphosphate, 2.0 mM magnesium chloride, 0.7 units *Taq* polymerase, and reaction buffer in a 20 µL volume. Reactions were performed in a Perkin Elmer 9700 96-well thermocycler (Applied Biosystems) using a 'touchdown' cycling sequence: 10 cycles were performed with the annealing temperature declining stepwise from 60° C to 52° C, followed by 25 cycles at 56° C. Denaturing and extension temperatures were 94° C and 72° C, respectively. The selective amplification conditions were identical except that 5 µL of 10x-diluted pre-amplification product was used as the template and the annealing temperature regressed from 64° C to 56° C. Transposon selective primers were 5'-labeled with the fluorescent dye 6-FAM and the resulting product was separated with a 6% polyacrylamide gel with the GeneScan Rox-500 size standard (Applied Biosystems). Gels were run in an ABI 377 electrophoresis unit (Applied Biosystems) and gel images were analyzed with GeneScan software (Applied Biosystems).

As with other marker systems, the identification of loci and the assessment of polymorphism and reproducibility require broad sampling of individuals within populations or species (Petit et al. 2005). For each primer set tested, we screened seven to eight individuals each from 25 populations of *I. missouriensis* across its range. While including reference individuals is a desirable practice for fragment analysis, the high level of polymorphism observed in this study and large number of individuals genotyped precluded the use of reference individuals on each gel. To define loci, fragments of approximately equal size were sorted into bins spanning two base pairs by rounding size estimates to the nearest base pair (Genescan size estimates are continuous rather than discrete). For a fragment to be included in the binning procedure, a minimum threshold intensity of 500 relative flourescence units (RFUs) was required, and fragments not present in multiple individuals in any population were excluded. These preliminary bins were then examined for consistency to determine whether bands sized within one base pair fall into the same or adjacent bins. Problematic bins were re-defined and these loci were then re-scored at a lower threshold of 300 RFUs. After repeating this process at a final threshold of 150 RFUs, any locus that showed a high rate of inconsistency among replicates was excluded from analysis.

The complement of the Dice coefficient (Dice 1945) was used to represent the genetic distance between multilocus genotypes. This coefficient is a preferred measure of genetic similarity for dominant markers (Duarte et al. 1999) that considers only shared and unshared 'present' phenotypes. Other measures such as the Pearson correlation are less desirable if there is significant genetic structure, as there will be many more loci present in the species as a whole than are present in any given population. As a result, individuals will necessarily share a high number of 'absent' phenotypes, biasing the genetic-distance estimate downward.

The matrix of mean genetic distance within and among populations was determined for each primer set tested. The geographic structure of transposon-SSAP was analyzed by AMOVA using GenAlEx 6 (Peakall and Smouse 2005), which implements the genetic distance measure of Huff et al. (1993). The Mantel correlation of geographic and Dice genetic distance, which estimates the contribution of isolation by distance to genetic structure, was also determined with GenAlEx 6. The reproducibility of multilocus genotypes was assessed by determining the Dice genetic distance between replicates. This reproducibility measure incorporates all component stages of transposon-SSAP, i.e. restriction digest and adapter ligation, pre-amplification, selective amplification, and fragment analysis.

In order to determine the phylogenetic pattern of diversification of IRRE subfamilies, a phylogeny of *I. missouriensis* and related taxa is necessary. The genus is divided into several subgenera, the largest and geographically most widespread of which is the subgenus *Limniris*, which may be paraphyletic (Wilson 2004). Almost all North American species of iris, including *I. missouriensis*, are in *Limniris*, within which species are grouped into series (Dykes 1913, Mathew 1989). We therefore obtained sequence for species from representative series of *Limniris* to provide a phylogenetic context for the distribution of IRRE elements in *Limniris*. This phylogeny includes only representative members of series likely to be closely related to *I. missouriensis* and is not intended to represent relationships within the genus as a whole, which is very large and poorly resolved (Wilson 2004).

Leaf tissue samples of iris species for phylogenetic analysis were obtained from a variety of sources. *Iris missouriensis* samples were collected by R.S.C. as previously described. *Iris fulva* (series *Hexagonae*) was obtained from a clonally propagated specimen collected in southern Louisiana by M.L.A. The University of California Botanical Garden in Berkeley,

California provided the following species: *I. setosa* (series *Tripetalae*), *I. tenax* (series *Californicae*), *I. forrestii* and *I. sanguinea* (series *Sibericae*), and *I. pseudacorus* (series *Longipetalae*). *Iris cristata*, which is not in *Limniris* and serves as an outgroup, was collected by R.S.C. on the campus of the University of Georgia, Athens, Georgia.

DNA sequence was obtained from GenBank accessions or by sequencing purified PCR fragments for each of the following chloroplast regions: the *trnL* intron, the *trnL-trnF* intergenic spacer, and the *trnK* intron/*matK*. Primers for sequencing were obtained from Taberlet et al. (1991) and Wilson (2004). PCR and sequencing protocols were identical to those described in Chapter Three. Sequences were concatenated and subsequently aligned using Sequencher 4.2 (Gene Codes Corp.). The length of the combined alignment was 2,714 bp after coding indels of 1-6 bp as a single nucleotide substitution and indels of greater than 6 bp as 2 nucleotide substitutions. GenBank accession numbers for all sequences are given in Table 3-1. The phylogeny was estimated using the maximum parsimony algorithm implemented in DAMBE (Xia 2001) with 10,000 bootstrap-resampled replicates. Organelles were not separated from nuclei prior to DNA extraction, but there was no indication that any obtained sequence was a nuclear pseudogene: tree topologies for individual regions were broadly consistent with the combined data set.

To examine the presence and diversity of IRRE subfamilies along each branch of the phylogeny, a PCR-based assay was developed which utilized the "LTRSCREEN" forward primer of Kentner et al. (2003) paired with each transposon primer examined in the present study (Table 3-2). Each sample was also tested with both LTRSCREEN primers as a positive control and every primer was tested unpaired as a negative control for single-primer amplification products. The PCR product was run in a 1% agarose gel and visualized with ethidium bromide

staining. Presence of an appropriately sized band is positive evidence for the presence of at least a partially intact integrase gene and 3' LTR (Kentner et al. 2003), but is not negative evidence for the presence of solo LTRs, which can be a substantial portion of the total LTR sequence (e.g. Vicient et al. 2003, Vitte and Panaud 2003, but see Kentner 2003). Presence of a band was considered evidence that the given subfamily was present in the stem lineage of that species, which carries the implicit assumption that presence is not due to either homoplasy or introgression. In addition to the species used to construct the phylogeny, other iris species were screened for this assay: *I. macrosiphon*, *I. hartwegii*, *I. purdyi*, *I. douglasiana*, and *I. innominata* of series *Californicae*, *I. versicolor* of series *Sibericae*, and *I. brevicaulis* of series *Hexagonae*. The University of California Botanical Garden provided leaf tissue for the *Californicae* species and *I. versicolor*. *I. brevicaulis* leaf tissue was obtained from a clonally propagated specimen collected in southern Louisiana by M.L.A.

RESULTS

Neighbor-joining analysis of 17 sequenced and aligned *I. missouriensis* clones identified two clearly distinct groups of two and fifteen sequences each (Figure 3-1). The average pairwise nucleotide similarity is 0.91 and 0.83 within each group, respectively, and 0.45 between groups; few regions of clear homology can be detected between the groups. The smaller group shows sequence homology to the IRRE1 family: clone 'MISS-16' has 74% sequence identity to *I. fulva* clone 'FULVA-24' (Figure 3-1), a member of the IRRE1A subfamily of Kentner et al. (2003), and has the same LTR start sequence and nearly identical polypurine tract (PPT) sequence as that subfamily (Figure 3-2). The larger group is less than 50% similar to LTR sequences identified by Kentner et al. (2003) and has a different LTR start sequence and more divergent PPT. Because

sequence information was not available for the coding regions of these clones, the family membership (IRRE1 or IRRE2) of this group could not be ascertained. For the present paper, we will refer to this larger group as IMRE1, for *Iris missouriensis* retroelement 1.

Primers used in this study are described in Table 3-2. For transposon-SSAP primers, the selective bases used, average number of scored fragments per individual, total number of scored loci, the average genetic distances between replicates, and the proportion of variation among populations (Φ_{PT}) are given in Table 3-3. Based on transposon-SSAP, the IMRE1 element appears to be highly abundant. Furthermore, loci were highly polymorphic and showed excellent repeatability. The IRRE1A element was much less abundant and produced slightly less reliable transposon-SSAP loci: genetic distances between replicates averaged 0.055 for the IRRE1A family versus 0.088 for the IMRE1 element (Table 3-3); these error estimates are more conservative than Pearson correlations of replicates. This lower reproducibility may be attributable to the fact that the pre-amplifcation primer B3 is uncoupled from the B1 selective primer for at least some element copies and some individuals (see below). That is, for these IRRE1A elements, some transposon-SSAP loci were effectively pre-amplified whereas others were not, which likely contributes to the much greater variation in signal intensity (results not shown) observed for this family than for IMRE1 markers.

To confirm that transposon-SSAP alleles were in fact amplifying the expected retroelement sequences, a single IRRE1A marker allele was isolated from a 3% agarose gel and sequenced. The sequenced allele was 258 bp in size and had high sequence similarity to the consensus of the two clones (0.97), but clean sequence could only be obtained for two regions of 72 and 110 bp each. The poor sequence quality for some regions is probably due to the fact that the allele was not cloned prior to sequencing. Sequence aligment demonstrated that the

sequenced allele was not a 5' LTR and adjacent 'host' sequence, but rather a 3' LTR and adjacent interior sequence between the integrase gene and LTR. Sequence analysis further demonstrated that this allele resulted from a derived *Eco*R1 restriction site arising within a particular transposon copy via two nucleotide substitutions (GCCTTC GAATTC). Thus, the sequenced allele verified that primers correctly amplify the target sequence and that some transposon-SSAP loci were not insertional polymorphisms but instead internal sequence variants. This class of mutation is expected, but probably does not contribute substantially to the total polymorphism because the sequenced allele was rare and restricted in its geographic distribution.

Transposon-SSAP for the two groups of retroelements showed substantially different population structure as estimated by AMOVA. The proportion of variation within populations was 76% ($\Phi_{PT} = 0.24$) for the IMRE1 element and 39% ($\Phi_{PT} = 0.61$) for the IRRE1A element. Furthermore, the Mantel correlation of genetic distance and geographic distance was lower for the IMRE1 element than for the IRRE1A element ($R^2 = 0.05$ vs. $R^2 = 0.13$, both significantly different from 0 at $P < 0.001$). Since the same individuals were sampled for each element, such a large difference in Φ_{PT} must reflect differences in mutation rate or else some bias in the identification or scoring of loci (see Discussion).

The maximum-parsimony phylogeny of representative *Limniris* species is shown in Figure 3-3 with the inferred origin of transposon subfamilies mapped onto the branches. *I. missouriensis* occupies a basal position in this phylogeny and is most closely related to the Louisiana Irises (series *Hexagonae*). A maximum likelihood tree did not differ in topology (not shown).

Table 3-4 shows the transposon primers that produced an amplification product when paired with the primer "LTRSCREEN forward" (Kentner et al. 2003). No amplification occurred

for any single primer, eliminating the possibility of a false signal arising from adjacent transposon fragments. Sequences from the IRRE1A group are present in all irises tested, whereas the IRRE1A subfamily isolated from *I. missouriensis*, which contains the (overlapping) sequences B2 and B3, is present only in that species. The interpretation that this IRRE1A subfamily arose in *I. missouriensis* is supported by two additional observations. All tested individuals from two of 31 *I. missouriensis* populations screened failed to produce a PCR product for these primers, indicating the absence of intact elements from their genomes and thus a species-level polymorphism for this subfamily. Secondly, some populations that tested positive for the PCR assay produced no transposon-SSAP bands for the B2 primer (Figure 3-4), indicating a low number of such elements in the genome (only 3 of the 16 possible two-base selective primers were tested, however).

The IMRE1 element was more widely distributed than the IRRE1A element: all four primers tested positive in *I. setosa* and *I. hartwegii*, and three out of four primers were positive in series *Sibericae* (*I. sanguinea* and *I. forrestii*) and several species of series *Californicae*. The pattern of results suggests that this element obtained high abundance in *I. missouriensis* after divergence from these other species, however. Given the many copies of the A2 sequence present in the *I. missouriensis* genome as evidenced by transposon-SSAP, it is not parsimonious to conclude that all copies of this sequence have diverged or been lost in derived species such as *I. pseudacorus* and multiple species of series *Californicae* while being retained intact in *I. missouriensis*. More probably, the subfamily was present in the stem lineage of *Limniris* at low copy number and has since greatly amplified in *I. missouriensis*.

Iris cristata tested positive only for the overlapping B0/B1 primers and was the only species testing negative for all IMRE1 primers and the IRRE1A primer B4. These results further support the use of *I. cristata* as an outgroup to the *Limniris*.

DISCUSSION

This study has demonstrated that IRRE transposon-SSAP markers are informative markers for population-genetic analysis of *I. missouriensis* and potentially other related species as well. While the two subfamilies identified in our screen differed substantially in number of loci and levels of polymorphism, cluster analysis showed concordant patterns of relationships among populations to the extent that they could be resolved (results not shown, see Chapter Three). This indicates that both marker systems reflect similar population histories within the framework of very different mutational histories. For phylogeographic analyses such as that presented in Chapter Three, the IRRE1A subfamily is the most appropriate choice for transposon-SSAP because of its greater among-poulation genetic structure. For other applications, such as analysis of fine-scale genetic structure or genetic mapping, the IMRE1 subfamily would be preferrable. To our knowledge, this is the first study to evaluate transposon-SSAP marker variation among population samples obtained throughout the species range. Previous studies have predominantly used germplasm accessions of domesticated plant species and their wild relatives.

This study provides evidence for a substantial amplification of IRRE elements in *I. missouriensis* since its divergence from related taxa. This evidence includes a high number of the IMRE1 elements compared to related taxa, and the gain of a unique subgroup of the IRRE1A subfamily that is defined by the presence of the B2 and B3 sequences and is absent in related taxa. The recent amplification of the latter subfamily is further indicated by the phylogeographic

concordance of populations that lack this sequence or exhibit negligible transposon-SSAP bands with these primers. Figure 3-4 shows an intraspecific neighbor-joining dendrogram of genetic distance between *I. missouriensis* populations based on B1 primer transposon-SSAP (see Chapter Three). The average number of B2 transposon-SSAP bands present in each population has been mapped onto this dendrogram, which shows clear increases in the number of B2 bands along some branches independently of the number of B1 bands, most notably along the branch denoted with an asterisk in Figure 3-4. This comparison is valid because the two quantities are not significantly correlated ($r = -0.35$, $P = 0.086$), indicating that most elements containing the B1 sequence lack the presumably derived B2 sequence.

The remarkable difference in geographic genetic-structure statistics (Φ_{ST} and the Mantel regression) between elements requires either differential mutation during the recent history of these populations or else substantial random error due to the inclusion of uninformative PCR fragments. The latter seems unlikely since the element that is least structured and most polymorphic, characteristics that would suggest low repeatability, produces highly repeatable transposon-SSAP. It seems more likely that the mutation rate of the IMRE1 subfamily is substantially higher than the IRRE1A subfamily, which may reflect ongoing transposition of the IMRE1 element or other modes of mutation such as microsatellite indels, non-microsatellite indels, or generation of new alleles by nonhomologous recombination. In fact, nonhomologous recombination between retrotransposon LTRs has been invoked to explain the high frequency of solo LTRs found in some studies (e.g. Vitte and Panaud 2003, Ma and Bennetzen 2004, but see Kentner et al. 2003, Garcia-Martinez and Martinez-Izquierdo 2003). The IMRE1 element contains di- and tri-nucleotide repeats and regions with high rates of indel formation (alignment not shown), but so too do the IRRE1A elements characterized by Kentner et al. (2003) from the

Louisiana Irises. While an elevated mutation rate would explain the greater polymorphism of IMRE1 transposon-SSAP loci, it would not explain the minimal geographic structure of the polymorphism unless homoplasious mutations are common and local genetic drift is minimal.

Given the evidence for rapid amplification of both subfamilies in the *I. missouriensis* lineage and for rapid divergence of IMRE1 transposon-SSAP markers among individuals and populations, the apparent contribution of IRRE elements to genetic diversity in this species is dramatic. Indeed, the species known to exhibit extensive phenotypic variation for vegetative and floral characters (Foster 1937, Davison 1957); it seems plausbile that some of this variation arises from IRRE-related mutations. It is not surprising, then, that the evolutionary dynamics of transposable elements have become a topic of considerable interest as the potential magnitude of their impact has been revealed by marker-based studies such as the present investigation as well as comparative genomic approaches (e.g. SanMiguel and Bennetzen 1999, Vitte and Panaud 2005). Theoretical work has focused on the effects of selection and mating system on the number and frequency of transposable element insertions (Charlesworth and Charlesworth 1983, Wright and Schoen 1999, Morgan 2001). The mating system has been identified as an important trait for the accumulation of transposable elements (Wright and Schoen 1999) because outcrossing shields the genome from the deleterious effect of recessive loss-of-function mutations likely to be associated with element insertion or nonhomologous recombination. *I. missouriensis* is an ideal system for evaluating these theoretical predictions because it has characteristics that suggest a range of realized mating systems such as (presumed) self-compatibility, a diverse pollinator assemblage (Lyon 1973), and substantial variation in population size and clonal structure (R.S.C. pers. obs.). Furthermore, if the species has a polyploid history as is suggested

by cytogenetic studies of *Limniris* (Simonet 1934, Foster 1937), large-scale deletions may be relatively benign, which would further permit variation associated with transposable elements.

ACKNOWLEDGMENTS

We gratefully acknowledge the technical assistance of E. Kentner in the cloning procedure. This work was supported by a grant from the American Iris Society (R.S.C.), by National Science Foundation Grant DBI 9602223 (R.S.C.), and by National Science Foundation Grant DEB-0345123. R.S.C. was also supported by a University of Georgia Presidential Fellowship.

Table 3-1. Chloroplast Sequences Used for Phylogenetic Reconstruction.

	DNA Region		
Species	*trnL* intron	*trnL-trnF* intergenic spacer	*trnK* intron/*matK*
I. setosa	*not yet submitted*	*not yet submitted*	AY596644.1*
I. tenax	*not yet submitted*	*not yet submitted*	AY596647.1*
I. cristata	*not yet submitted*	*not yet submitted*	AY596639.1*
I. missouriensis	*not yet submitted*	*not yet submitted*	*not yet submitted*
I. sanguinea	*not yet submitted*	*not yet submitted*	AY596646.1*
I. forrestii	*not yet submitted*	*not yet submitted*	AY596645.1*
I. pseudacorus	*not yet submitted*	*not yet submitted*	*not yet submitted*
I. fulva	*not yet submitted*	*not yet submitted*	*not yet submitted*

*From Wilson (2004)

Table 3-2. Characteristics of Transposon-SSAP Candidate Primers. Position is given as the 3' to 5' positions of the transposon consensus sequence to which the primer is homologous.

Transposon Subfamily	Primer Code	Position of Primer Relative to LTR Start	Sequence (5' to 3')	Tm at 50mM NaCL (Celsius)
IMRE1	A2	43->22	TATTCAAGCCACAAATACGGAA	54.0
	A3	55->34	GAATACTGCACGTATTCAAGCC	55.3
	A4	62->39	TATACGAGAATACTGCACGTATTC	53.6
	A5	74->52	TACAGCAAGCAATATACGAGAAT	53.6
IRRE1A	B0	17->1	AGGAACGGGGAGTGACA	55.7
	B1	24->5	GTATATCAGGAACGGGGAGT	54.1
	B2	71->49	CTAGTCTACTAGTCTCTACCTCG	53.3
	B3	82->59	CTCACCCATATCTAGTCTACTAGT	53.6
	B4	141->121	YAAACCAAGCCATGCTAAACY	53.7

Table 3-3. Characteristics of Transposon-SSAP Loci for Primers Tested in *I. missouriensis.* All transposon-SSAP primers were paired with the standard EcoR1 pre-amplification primer or EcoR1 selective amplification with designated selective bases added (Vos et al. 1995). PCR conditions are described in Materials and Methods.

Transposon Subfamily	Transposon Pre-amplification Primer	Transposon Selective Primer	EcoR1 Selective Primer	Mean Number of Presence Alleles per Individual	Total Number of Scored Loci	Mean Genetic Distance between Replicate Inidviduals	Mean Genetic Distance within Populations	Proportion of Variation Among Populations (PhiPT)
IMRE1	A3	A2	+ACTA	12.4	75	0.055	0.505	0.24
			+AGTG	9.4	56			
			+CTAT	12.1	66			
IRRE1A	B3	B2	+AC	3.3	39	0.088	*	*
			+AG	3.4	45			
			+CT	4.1	39			
		B1	+A	10.5	87		0.296	0.39

*This primer combination did not produce bands in some *I. missouriensis* populations, which causes these statistics to be undefined.

Table 3-4. Results of PCR assay for the presence of candidate transposon-SSAP primer sequence. X indicates presence of the sequence in all individuals tested, P indicates presence in some but not all individuals tested, and - indicates absence in all individuals tested.

		Primer paired with LTRSCREEN Forward								
		IMRE1 Subfamily				IRRE1A Subfamily				
Series of Subgenus *Limniris*	Species	A2	A3	A4	A5	B0	B1	B2	B3	B4
Sibericae	*I. sanguinea*	X	X	X	-	X	X	-	-	-
	I. versicolor	X	X	X	-	X	X	-	-	X
	I. forrestii	X	X	X	-	X	X	-	-	X
Tripetalae	*I. setosa*	X	X	X	X	X	X	-	-	-
Hexagonae	*I. brevicaulis*	X	-	X	-	X	X	-	-	X
	I. fulva	X	-	X	-	X	X	-	-	X
Evansia	*I. cristata*	-	-	-	-	X	X	-	-	-
Laevigatae	*I. pseudacorus*	X	-	-	-	X	X	-	-	X
Californicae	*I. hartwegii*	X	X	X	X	X	X	-	-	X
	I. purdyi	X	-	X	X	X	X	-	-	X
	I. macrosiphon	-	-	X	-	X	X	-	-	X
	I. innominata	-	-	-	-	X	X	-	-	X
	I. tenax	X	X	X	-	X	X	-	-	X
	I. douglasiana	P	P	P	-	X	X	-	-	P
Longipetala	*I. missouriensis*	X	X	X	X	X	X	P	P	X

Figure 3-1. An unrooted neighbor-joining dendrogram of sequenced clones from *I. missouriensis* and representative IRRE1A elements from *I. fulva* characterized by Kentner et al. (2003). Bootstrap support for major nodes and subfamily designations applied in the text are shown.

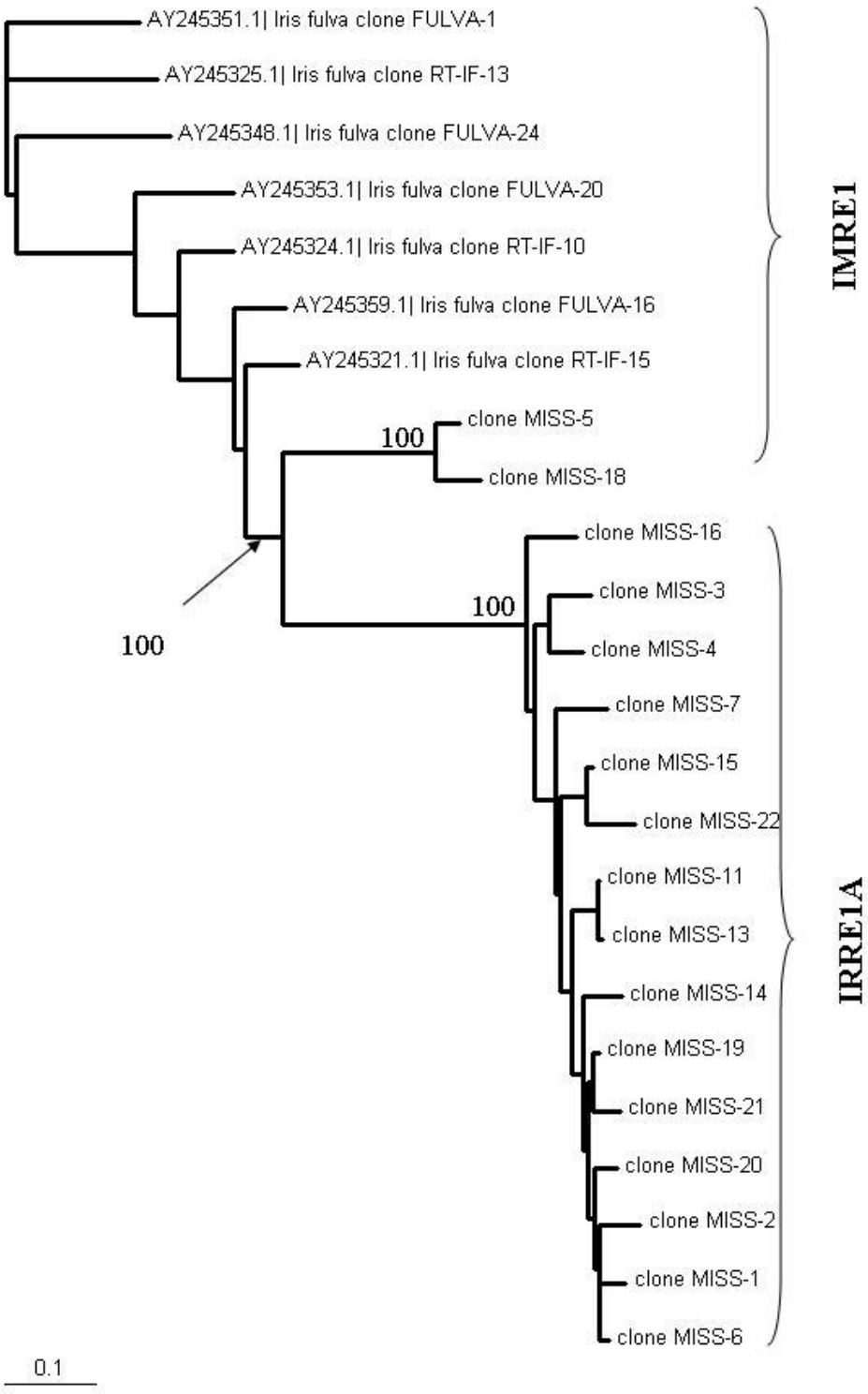

AY245351.1| Iris fulva clone FULVA-1
AY245325.1| Iris fulva clone RT-IF-13
AY245348.1| Iris fulva clone FULVA-24
AY245353.1| Iris fulva clone FULVA-20
AY245324.1| Iris fulva clone RT-IF-10
AY245359.1| Iris fulva clone FULVA-16
AY245321.1| Iris fulva clone RT-IF-15
clone MISS-5
clone MISS-18
clone MISS-16
clone MISS-3
clone MISS-4
clone MISS-7
clone MISS-15
clone MISS-22
clone MISS-11
clone MISS-13
clone MISS-14
clone MISS-19
clone MISS-21
clone MISS-20
clone MISS-2
clone MISS-1
clone MISS-6
100
100
100
IMRE1
IRRE1A
0.1

Figure 3-2. Schematic of sequence characteristics and primer locations for each of the two groups of IRRE sequence isolated from *I. missouriensis*. The polypurine tract (PPT) and LTR start are functional sequences and are well conserved among the two subfamilies as well as the elements identified by Kentner et al. (2003). Sequence characters in bold represent base substitutions relative to the consensus sequence of the IRRE1A subfamily of Kentner et al. (2003).

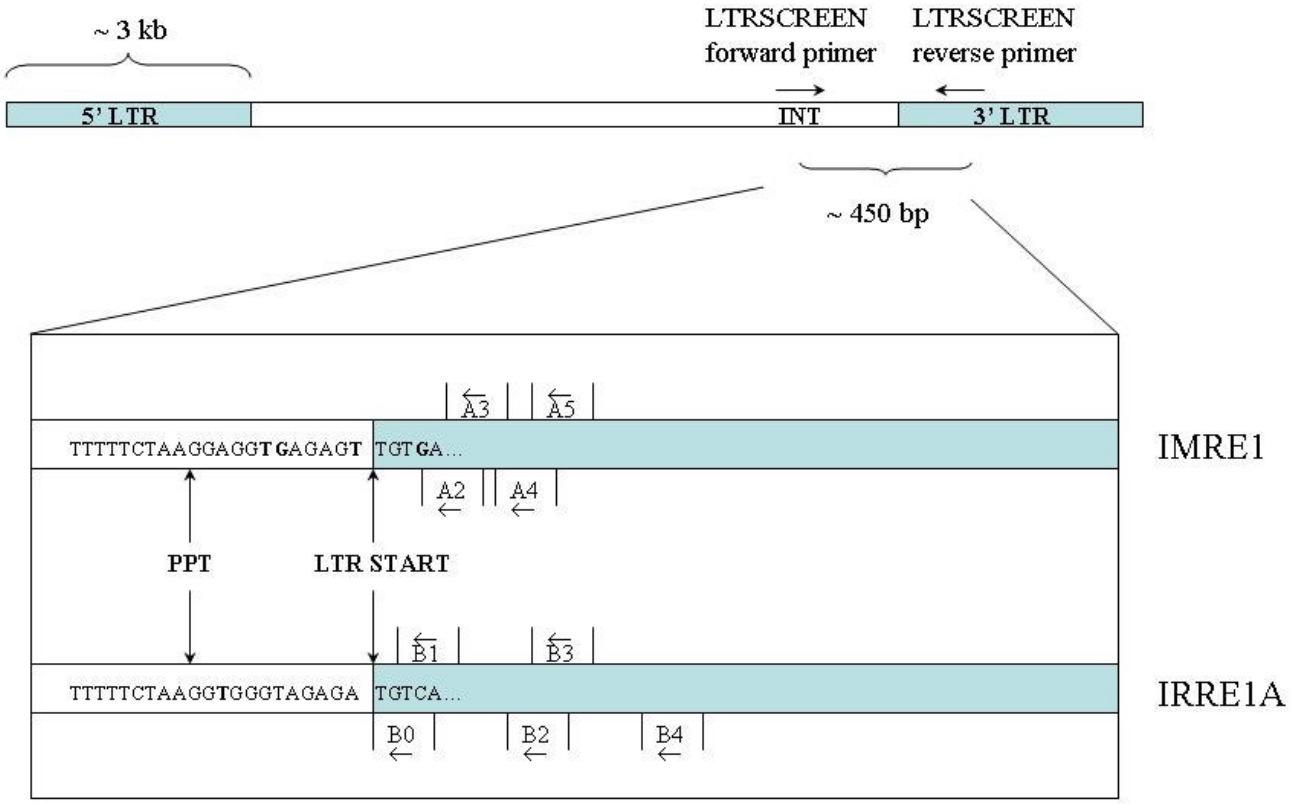

~ 3 kb
LTRSCREEN
forward primer
LTRSCREEN
reverse primer
5' LTR
INT
3' LTR
~ 450 bp
A3
A5
TTTTTCTAAGGAGGTGAGAGT
TGTGA...
IMRE1
A2
A4
PPT
LTR START
B1
B3
TTTTTCTAAGGTGGGTAGAGA
TGTCA...
IRRE1A
B0
B2
B4

Figure 3-3. The single most parsimonious tree of *I. missouriensis* and related species based on three chloroplast regions: *trnK* intron/*matK*, *trnL* intron, and *trnL – trnF* intergenic spacer. The total length of the tree is 243 steps and it is rooted with the species *I. cristata*, which is not in *Limniris*. Bootstrap support from 10,000 resampled data sets is shown for each node. Inferred events in the origination and amplification of IRRE subfamilies are mapped onto the phylogeny (see text and Table 3-4).

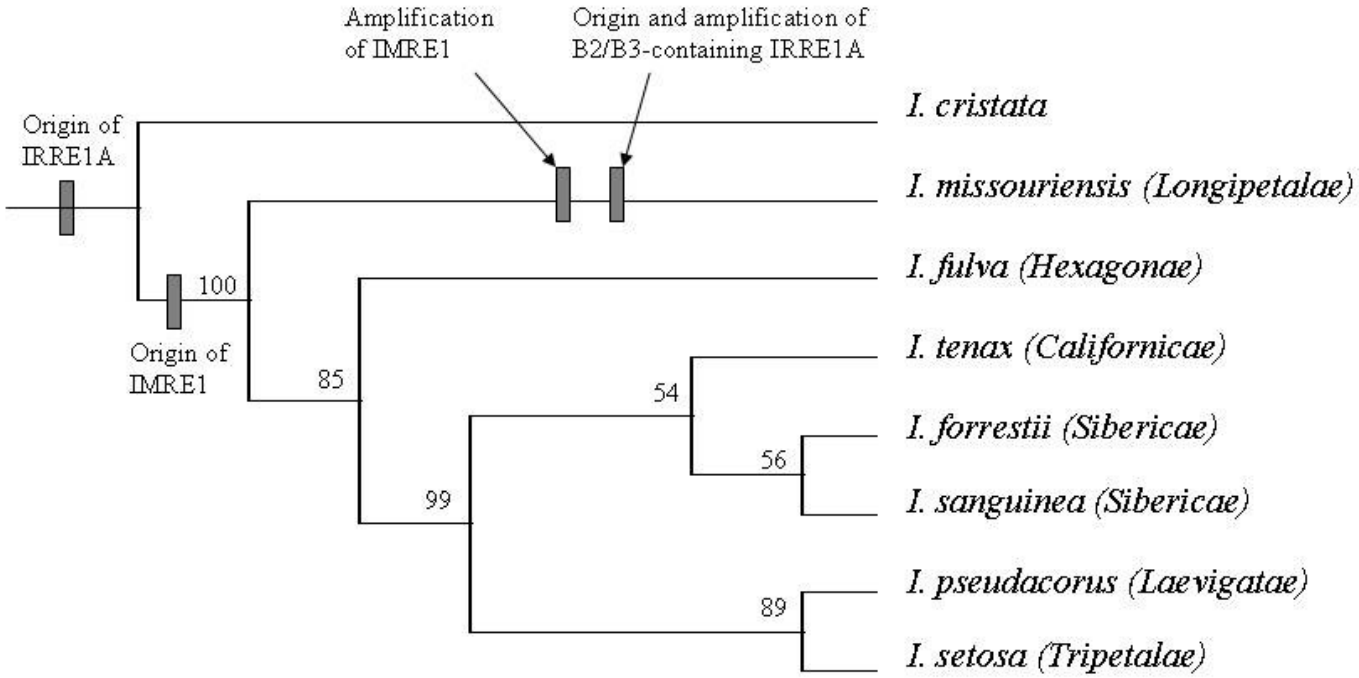
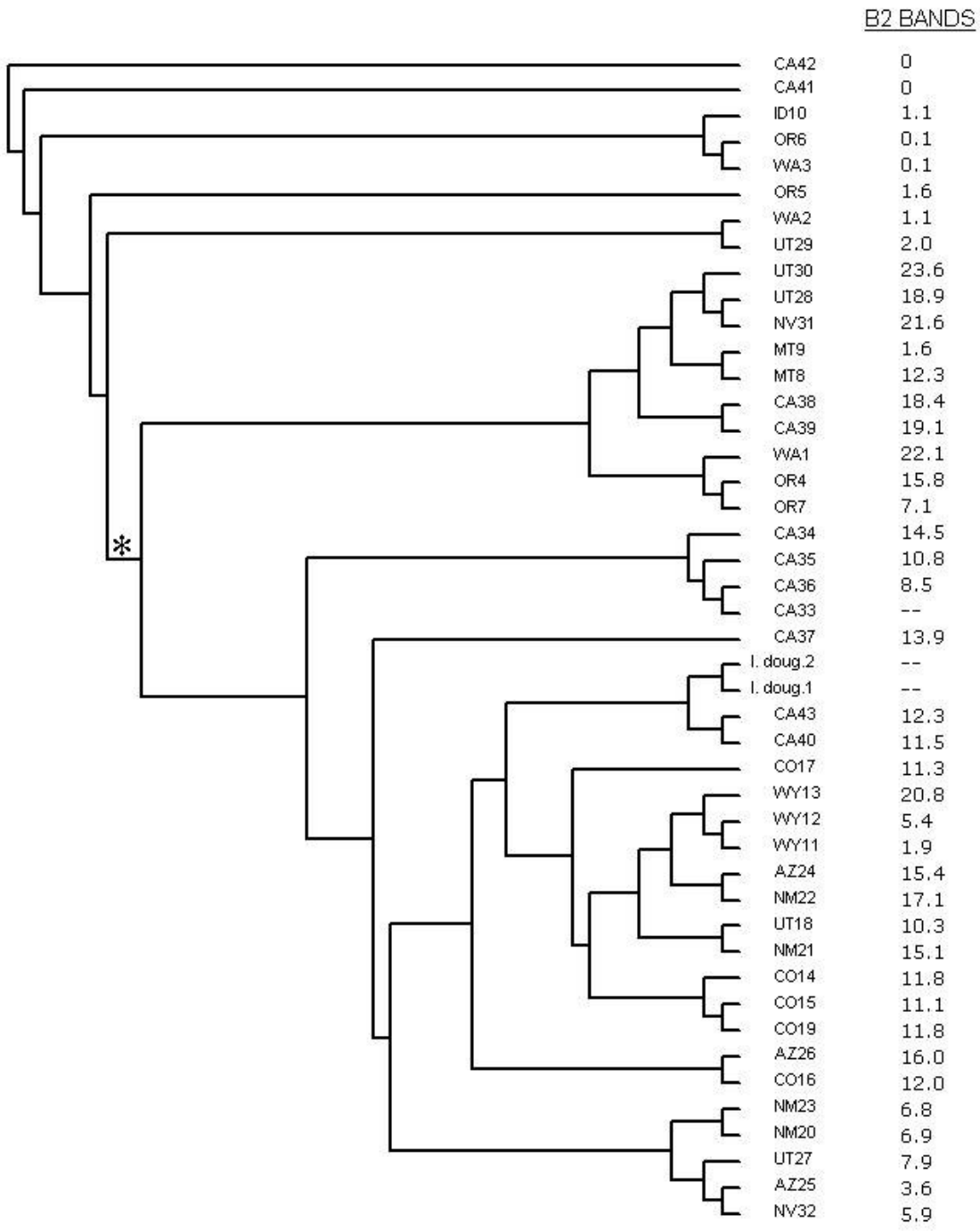

Amplification of IMRE1
Origin and amplification of B2/B3-containing IRRE1A
Origin of IRRE1A
100
Origin of IMRE1
85
54
56
99
89
I. cristata
I. missouriensis (Longipetalae)
I. fulva (Hexagonae)
I. tenax (Californicae)
I. forrestii (Sibericae)
I. sanguinea (Sibericae)
I. pseudacorus (Laevigatae)
I. setosa (Tripetalae)

Figure 3-4. Neighbor-joining dendrogram of Dice genetic distance matrix based on B1 selective primer transposon-SSAP. The population-mean number of bands produced by the B2 primer within the size range 130 to 425 bp is given next to each population. See text and Chapter Three for additional explanation and population locations. The asterisk indicates the node referred to in the text; -- indicates populations for which data are not available; “I. doug.” indicates *I. douglasiana* (see Chapter Three).


B2 BANDS
CA42 0
CA41 0
ID10 1.1
OR6 0.1
WA3 0.1
OR5 1.6
WA2 1.1
UT29 2.0
UT30 23.6
UT28 18.9
NV31 21.6
MT9 1.6
MT8 12.3
CA38 18.4
CA39 19.1
WA1 22.1
OR4 15.8
OR7 7.1
CA34 14.5
CA35 10.8
CA36 8.5
CA33 --
CA37 13.9
I. doug.2 --
I. doug.1 --
CA43 12.3
CA40 11.5
CO17 11.3
WY13 20.8
WY12 5.4
WY11 1.9
AZ24 15.4
NM22 17.1
UT18 10.3
NM21 15.1
CO14 11.8
CO15 11.1
CO19 11.8
AZ26 16.0
CO16 12.0
NM23 6.8
NM20 6.9
UT27 7.9
AZ25 3.6
NV32 5.9

CHAPTER 4

PHYLOGEOGRAPHY OF *IRIS MISSOURIENSIS* BASED ON NUCLEAR TRANSPOSON-SSAP AND CHLOROPLAST DNA SEQUENCE[3]

[3] Cornman, R. S. and M. L. Arnold. To be submitted to *Molecular Ecology*.

ABSTRACT

We used nuclear and chloroplast markers to investigate the geographic structure of genetic variation in *Iris missouriensis*. The first objective of the study was to determine whether genetic data were consistent with a single widespread species or whether previously described morphological variants were genetically distinct. The second objective was to evaluate support for phylogeographic hypotheses of relatedness due to paleoclimatic events associated with glacial cycling. Forty-three populations sampled throughout the range of the species were genotyped for 87 transposon sequence-specific amplified polymorphisms (transposon-SSAPs). We applied analysis of molecular variance (AMOVA) and principal components analysis (PCA) to the resulting distance matrix and constructed a neighbor-joining dendrogram. There was no evidence supporting the species status of the subspecies *I. missouriensis* var. *longipetala*. However, populations were strongly differentiated ($\Phi_{PT} = 0.61$) and distinct geographical structure was evident. Almost fifty percent of the variation could be partitioned among three axes by PCA, which was driven by the genetic distinctiveness of the southern Rocky Mountains, the Sierra Nevada, and a trans-Colorado Plateau group. These clusters were also evident in the neighbor-joining dendrogram, which we were able to root with two populations that lack a particular transposon sequence. Based on this rooting, we infer that the southern Rocky Mountains were colonized by individuals from coastal California rather than the more biogeographically similar Sierra Nevada or Great Basin provinces. The northern part of the range appears to have been colonized by several distinct lineages and there is an overall reduction in genetic diversity in that region. Chloroplast variation was also consistent with a recent colonization of the southern

Rocky Mountains and with multiple sources of re-colonization of the northern range. However, there was little overall concordance in nuclear and chloroplast genetic structure, suggesting that dispersal via pollen is largely independent of dispersal by seed. The highlands of Utah were an extreme example of this phenomenon, exhibiting a uniform chloroplast haplotype but divergent nuclear affinities. This study demonstrates the utility of transposon-SSAP markers for phylogeographic analysis and is, to our knowledge, the first to examine transposon marker diversity in population samples throughout a species range.

INTRODUCTION

Species diversity arises from an interplay between cladogenic processes and processes that maintain the connectivity of gene pools over species ranges. The connectivity of populations depends on such factors as the distribution of suitable habitat, the mating system, dispersal mechansisms and metapopulation dynamics, and the extent of shared selection regimes (Levin 2000, Rieseberg and Burke 2001). The history of migration of populations across a dynamic landscape also profoundly shapes population connectivity, i.e. the history of population expansion, contraction, fragmentation, and extinction (Avise 2000). These histories can produce correlations of allele frequencies among populations that are demographically more connected than average. Unfortunately for population biologists, correlations also occur that are random with respect to lineage history and reflect accidental outcomes of lineage sorting. For this reason, various statistical approaches have been developed to analyze genetic variation at geographic scales so as to infer patterns of connectivity and their underlying causes (Avise 2000). A major goal of

such analyses is to determine the relative importance of demographic history in generating evolutionary independence and ultimately cladogenesis.

Phylogeographic assessments of species-level genetic variation have become a common practice for animal taxa, and are becoming so for plant species despite greater difficulties in isolating organellar molecular markers. Studies have focused on organellar rather than nuclear loci because recombination and a four-fold greater effective population size greatly reduce the informativeness of individual nuclear loci. However, as pointed out by several authors (e.g. Hare 2001, Zhang and Hewitt 2003), the collective utility of nuclear loci is great because they represent numerous quasi-independent estimates of relatedness across the genome. Their interpretation is therefore less prejudiced by the vagaries of lineage sorting of individual loci.

As phylogeographic studies of plants and animals have accumulated, comparative analyses have shown that in spite of the idiosyncratic histories of species, concordant breaks in the genetic continuity of nuclear or organellar diversity are frequently found across species and can be associated with external factors such as vicariance events, topography, or climate history (Avise 2000). Thus, phylogeographic studies inform, and are informed by, the work of biogeographers and paleoclimatologists investigating the past distribution and stability of biomes. For example, phylogeographic studies can provide a large amount of independent evidence for hypothesized climate change (Hewitt 2000, Arbogast and Kenagy 2001, Kadereit et al. 2004) and are not limited to those organisms predisposed to fossilization, e.g. in the form of pollen in lake beds or branches in pack-rat middens. On the other hand, dating of such fossils is far more accurate than molecular-clock estimates over these times scales.

The current continental flora has been shaped by glacial cycles over the past 100,000 years, which have produced dramatic variation in temperature and moisture regimes far from the ice sheets themselves (COHMAP Members 1988, Williams et al. 1998, Thomspon and Anderson 2000). Glacial cycling is a consequence of the gradual cooling of global climate since the Tertiary, particularly since the emergence of the Panamanian Isthmus approximately three million years ago, resulting in increased responsiveness of climate to Milankovitch cycles of solar radiation (Williams et al. 1998). The effects of these oscillations on flora and fauna have been most extensively studied in Europe, eastern North America, and Japan. Less is known about phylogeographic patterns in western North America and even fewer studies have been conducted on the flora of other temperate zones.

Reviews by Soltis et al. (1997), Hewitt (2000), Brunsfeld et al. (2001), Hampe and Petit (2005), and Swenson and Howard (2005) have summarized the observed and expected patterns in western North American phylogeography, a region with complex topography and a diversity of biomes. Furthermore, explicit hypotheses have been developed (e.g. Brunsfeld et al. 2001) that predict patterns of genetic variation based on current biome associations and which are predicated on an imperfect but increasingly detailed (e.g. Thompson and Anderson 2000) understanding of the biogeography and climate of the Quaternary. Phylogeographic concordance is necessarily specific to particular biomes and the extent of concordance is expected to co-vary with life history traits such as dispersal and mating system.

In this study, we investigate the geographic structure of genetic variation in an herbaceous perennial, *Iris missouriensis*. This species is of interest for several reasons.

The species range encompasses the whole of the western contiguous United States with extensions into northern Mexico and the Canadian Rockies, so that multiple distinct phylogeographic hypotheses can be tested. Within this wide range, populations tend to be disjunct with presumably low rates of pollen and seed exchange. On the other hand, the degree of habitat connectivity – the species prefers seasonally wet meadows, seeps, and creek banks – has undoubtedly changed with climatic oscillations. It therefore seems likely *a priori* that demographic history would be an important determinant of genetic structure relative to a simple isolation-by-distance model of genetic relatedness.

Another motivation for investigating the phylogeography of *I. missouriensis* is to initiate comparisons with other North American members of the genus *Iris* subgenus *Limniris*, which have been important systems for the study of speciation and hybridization. For example, the section *Californicae* has been investigated by a number of authors (Lenz 1959, Young 1996, Wilson 2003) as an example of a rapidly diverged, species-rich clade arising in a relatively small geographic region and without strong reproductive barriers. The section *Hexagonae* presents an even more extreme example of phenotypic divergence, with respect to ecophysiology and pollination syndrome, despite minimal differentiation of nuclear and plastid genomes (Arnold 2006). In contrast, *I. missouriensis* is the only currently accepted species of the section *Longipetalae*, although the area of its range is greater than that of the *Californicae* and *Hexagonae* combined. Early treatments of the taxon identified *I. longipetala* (coastal California) and *I. montana* (north and east of the continental divide) as separate species (Dykes 1913), but these entities have since been reduced to subspecific status due to a lack of consistent morphological differences (Foster 1937, Species Group of the British Iris Society 1997).

A major objective of the present study is to determine whether the genetic structure of nuclear and organellar markers is in fact consistent with a single widespread species. A resolution of the species status of the taxon *I. missouriensis* ssp. *longipetala* is of particular concern because it has suffered a substantial decline due to urbanization in the San Francisco Bay region (L. Weeth pers. comm.) and therefore may be an appropriate conservation target. More broadly, we wish to gain insight into the processes that have led to ecological divergence in some geographic regions and *Iris* lineages but not in others, and the contribution of demographic history to these processes.

Study system

Iris missouriensis Nuttall (Iridaceae, Liliopsida), "Western Blue Flag", is a rhizomatous herb occurring in western North America from the highlands of northern Mexico to the southern Canadian Rockies. It prefers spring-wet, summer-dry montane habitats such as meadows, seeps, and creek banks where light competition is minimized (Davidson 1957). The species is facultatively clonal and presumed to be self-compatible because this is the condition of related taxa (Cornman et al. 2004, Zink and Wheelwright1997). It produces one to several open, insect-pollinated flowers with a nectar reward (Lyon 1973, Metcalf 1978). Seeds are produced in a capsule which is held upright on a woody stalk until the next growing season. Seeds are presumably dispersed by gravity primarily, but are sufficiently small that they may occasionally be transported longer distances on wet fur, plumage, and the like. The species is perhaps most frequently associated with aspen (*Populus tremuloides*), which unfortunately is not a significant component of the microfossil or macrofossil records. Co-distributed vegetation that is present in these records includes *Pinus* species such as Lodgepole Pine, Limber Pine, and Ponderosa

Pine, *Abies*, *Picea*, and *Salix*. The species can therefore be considered a component of the "cool conifer forest" biome of Thompson and Anderson (2000).

A number of authors have remarked on the high degree of phenotypic variability in this species (Dykes 1913, Foster 1937, Davidson 1957, Metcalf 1978) and early authors (e.g. Dykes 1913 and references therein) identified multiple species within the '*missouriensis* complex'. Current treatments (Species Group of the British Iris Society 1997) recognize only a single widespread species and the subspecies *I. missouriensis* ssp. *longipetala*, which occurs in the foothills of the Coast Range in central California and has a generally larger stature. This difference appears to be largely environmental rather than genetic (Metcalf 1978). Chromosome number appears to be variable in *I. missouriensis* (Simonet 1934, Foster 1937), but the extent of polyploid or aneuploid variation has not yet been determined.

Phylogeographic hypotheses for I. missouriensis

Dynamics at the northern and southern margins.. In Western North America, the Cordilleran ice sheet of the last glacial maximum extended into the northern parts of Montana, Idaho, and Washington, including the Puget Sound area (which was exposed continental shelf) (Williams et al. 1998). The Cascade Range was also glaciated except along the lower eastern slope, whereas mountain glaciation in the northern Rocky Mountains was more limited and alpine habitats moved downslope and were more connected. The climate in the intervening basins and valleys was arid and cold, resulting in an expanding steppe biome. As a result, mesic forest species were eliminated except for refugial bands in the northern Rocky Mountains, and possibly also along the lower eastern slopes of the Cascades or on large islands such as Vancouver or the Queen

Charlotte Islands (Brunsfeld et al. 2001). *Iris missouriensis* would have been able to persist in these refugia as well as warmer and wetter regions to the south (see below), but should have been excluded from the steppe vegetation of lowland Washington and Idaho. We therefore expect to find reduced nuclear and chloroplast genetic diversity in populations from this region resulting from recolonization bottlenecks. We also expect that if recolonization occurred from a single refugial source, populations in this region will cluster with the source populations. Alternatively, recolonization from multiple sources would be revealed by a pattern of greater than average variation among populations.

While a number of studies have investigated the genetic consequences of paleoclimate on the northern ranges of temperate species of the Northern Hemisphere, Hampe and Petit (2005) argue that the southern portions of species' ranges are understudied despite being of considerable relevance to conservation genetics and evolution. Southern edges may be reservoirs of genetic diversity under a stable southern-edge model or may show a loss of within-population diversity parallel to that of northern populations under an expansion-and-contraction model (Hampe and Petit 2005). In this study, we investigate genetic diversity within southern populations to determine which alternative is best supported for *I. missouriensis*.

Discontinuity between northern and southern Rocky Mountains. During glacial cycles, the alpine and montane zones migrated elevationally as well as latitudinally, experiencing alternating periods of connectivity and isolation. The arid, low-elevation Wyoming Basin separates the northern and southern Rocky Mountains and is a presumed barrier to dispersal between the regions because it does not appear to have supported

alpine flora during glacial maxima (DeChaine and Martin 2005). The result is a phylogeographic break between the two regions such as that identified by DeChaine and Martin (2005) in *Sedum lanceolatum*. Since *I. missouriensis* is primarily a montane plant with extensions into spring-wet components of more arid habitats, the Wyoming Basin has likely been only an intermittent dispersal barrier. We therefore expect that discontinuity between the two regions should be evident in *I. missouriensis* but weaker than is observed in *S. lanceolatum*.

Connectivity of the Sierra Nevada, southern (pluvial) Great Basin, and southern Rocky Mountains. During the last glacial period, the presence of the Cordilleran ice sheet is believed to have split the jet stream over western North America, causing one component to move over the southwestern United States and the other component to pass over southern Canada (COHMAP Members 1988). As a result, the southwestern U.S. experienced much greater precipitation, which, coupled with cooler temperatures, supported a great expansion of coniferous woodland from the southeastern foothills of the Sierra Nevada across the Great Basin to the southern Rocky Mountains. The Sierra Nevada themselves were drier than at present (Thompson and Anderson 2000). Reduced evaporation in the Great Basin supported the development of large pluvial lakes such as the massive Lake Bonneville, which further contributed to the development of more mesic habitat by generating lake-effect precipitation. As a result, habitat connectivity between the Sierra Nevada and the southern Rocky Mountains was much greater from approximately 18,000 yr bp until sometime prior to 6,000 yr bp (Thompson and Anderson 2000). We therefore predict that the southern Rocky Mountains will cluster more closely with southern Great Basin and Sierra Nevada populations than with other

western parts of the range. However, this prediction is confounded somewhat by the smaller distances between the former than the latter, such that positive evidence should be interpreted cautiously.

Connectivity of the Basin and Range province of the southern Rocky Mountains. Based on the same paleoclimatic considerations outlined above, the Basin and Range province of the southern Rocky Mountains is expected to have supported largely continuous tracts of conifer woodland during recent paleoclimatic history. The period of forest expansion persisted longer than in the Great Basin, extending until at least 6,000 yr bp (Thompson and Anderson 2000). It is therefore predicted that populations currently isolated by large tracts of inhospitable terrain will nonetheless show a continuous pattern of genetic variation with an isolation-by-distance signature that is weaker than what is expected across the Great Basin.

MATERIALS AND METHODS

Population sampling

Sites were identified from published floras, online herbarium databases, and internet searches. Samples were collected from 56 populations throughout the range; population locations are shown in Figure 4-1 and population codes are given in Table 4-1. *I. missouriensis* is the only iris throughout much of its range, except in California where it is parapatric with *I. douglasiana* near the Pacific Coast and *I. hartwegii* east of the Sierra Nevada. Numerous characteristics of leaf, rhizome, and fruit distinguish *I. missouriensis* from these parapatric taxa, such that field identification is unambiguous. Furthermore, tissue samples of all of these taxa were obtained from the University of California

Botanic Garden for DNA sequencing to confirm the species identity of field-collected samples.

At least eight samples were collected per site. Samples were taken haphazardly from individuals dispersed throughout the site to avoid sampling clones. The size of some populations was such that no more than eight samples could be collected with reasonable confidence that they were from separate genetic individuals. Samples consisted of approximately five cm of green leaf tissue and were either flash-frozen in liquid nitrogen or stored in dessicating silica granules. A hexadecyltrimethylammonium bromide (CTAB) extraction protocol was used to extract genomic DNA (Doyle and Doyle 1987).

Nuclear marker development and analysis

In order to assess genome-wide patterns of differentiation, we developed sequence-specific amplified polymorphism (SSAP) markers based on a family of long-terminal repeat (LTR) retrotransposons. This technique is commonly called 'transposon display' (van der Broeck et al. 1998) and development of these markers for *I. missouriensis* is described in Chapter Two. These markers have been shown to be repeatable, have Mendelian inheritance, and, importantly, cover the breadth of the genome (Chapter Two; Bouck et al. 2005). AFLP markers (Vos et al. 1995) were not used because in our experience they have low repeatability, low diversity, and fail to show Mendelian segregation (Chapter Two; A. Bouck, pers. comm.) in *Iris*. This is presumably due to the very large number of repetitive elements in the genome (Kentner et al. 2003), which for many four- or six-cutter restriction enzymes would produce multiple nonorthologous bands of similar size.

Laboratory methods for obtaining transposon-SSAP data were as described in Chapter Two. For this study, we utilized primers that annealed to LTR sequence of a cloned retroelement belonging to the IRRE1A subfamily described by Kentner et al. (2003). This element was chosen because it gave useful patterns of polymorphism during initial screens, that is, a moderate number of bands that had clear geographic structure. Samples were initially genotyped using a selective primer placed approximately 70 bp from the start of the LTR so as to avoid scoring small-sized amplification products; this primer is referred to as B2 in Chapter Two. All populations were genotyped using three two-base selective primers, however some populations produced no or very weak and unscorable PCR product with these primers. Since genetic distances between individuals that produce no amplification product are undefined, this data set was not useful for most analyses presented here. Instead, samples were genotyped again using the B1 primer (see Chapter 2) and a one-base selective primer (+A), which produced scorable bands in all individuals. The number of transposon-SSAP loci with the presence phenotype using the B2 primers was nonetheless useful for establishing the likely polarity of relationships identified with the B1 primer set, i.e. we assumed that high fragment number is a derived state based on results described in Chapter Two.

Analysis of transposon-SSAP polymorphism was premised on the Dice similarity coefficient for binary characters (Dice 1945), which is defined as (2*a)/(2*a + b), where a is the number of loci scored as present in both individuals and b is the number of loci scored as present in one individual but absent in the other. The complement of this coefficient is a preferred measure of genetic distance for dominant loci and has been

shown by simulation studies to perform well compared with other possible measures of genetic distance (Duarte et al. 1999).

The following genetic distances were calculated: the distance between all pairs of sampled individuals, the mean pairwise distance between individuals within a population, and the mean pairwise distance between individuals in each population pair. Distances were calculated using routines written in Visual Basic for Excel. Since no public software was available for bootstrap resampling of Dice coefficients, additional routines were written to automate data resampling for this purpose.

Principal components analysis of the Dice distance matrix was performed with GenAlEx v. 6 (Peakall and Smouse 2005) using the standardized distance option. GenAlEx was used to obtain Mantel correlations between (arcsine square-root transformed) genetic distance and (ln transformed) geographic distance. GenAlEx was also used to perform analysis of molecular variance (AMOVA), but because this analysis requires a Euclidean distance rather than a distance coefficient, the program implements the method of Huff et al. (1993) to calculate genetic distance. Unlike the Dice method, this method considers shared presence and shared absence to be equally informative. We expect that estimates of within-population genetic distance and pairwise population genetic distance will have wide variances for samples of eight individuals, especially for dominant markers, and thus advocate a conservative interpretation of these statistics. However, our primary goal is understanding the overall genetic structure of the species, which is considerably more sensitive to total sample size than to individuals per population (Petit et al. 2005), and our sampling scheme should provide robust estimates for this purpose.

A dendrogram of populations was constructed using the neighbor-joining method as implemented in Phylip (Felsenstein 2004). It bears emphasizing that the nodes of intraspecific trees represent only clusters of genetic similarity and do not imply a specific biological event, i.e. they do not represent a hypothesized cladogenic event as is usually assumed for interspecific trees. As emphasized by Smouse (1998), alternative representations of the data such as AMOVA may be more useful than bifurcating dendrograms for addressing specific questions.

Chloroplast marker development and analysis

In order to estimate the genetic structuring of populations via seed dispersal, we investigated the distribution of maternally inherited (~99%; Cruzan et al. 1993) chloroplast haplotypes. To identify haplotypes, we screened three regions of the chloroplast genome, the intron of the tRNA gene *trnL*, the spacer between *trnL* and the tRNA gene *trnF*, and the protein-coding gene *matK*. Since these regions show moderate interspecific variation despite the low overall level of diversity in the *Iris* chloroplast genome (e.g. Wilson 2003, Wilson 2004), they were appropriate choices for exploratory analysis. Polymorphism was not observed in the *trnL* intron in these samples during initial screens and this region was not further investigated.

We PCR-amplified organellar target regions from gross cellular DNA extracts using published primers (primers "e" and "f" of Taberlet et al. 1991, primer 7i of Wilson 2004, and primer 5 of Steele and Vilgalys 1994) and purified the PCR product for sequencing using a guanidine thiocyanide method (Promega "Wizard" system for PCR product). PCR product was pooled from four separate reactions for each sample to minimize the contribution of errors attributable to *Taq* polymerase in the sequenced

template. We performed chain-termination sequencing using the BigDye 3.1 chemistry (Applied Biosystems) and submitted the precipitated product to the Integrated Biotech Laboratory of the University of Georgia for separation on an ABI 3700 capillary device (Applied Biosystems). All polymorphic sites were sequenced in both directions.

Given the potentially low rate of seed dispersal and the haploid and uniparental nature (in most angiosperms) of the chloroplast genome, prior studies have typically found strong structuring of species-level variation among populations with low within-population variation (Petit et al. 2005). In fact, animal mitochondrial phylogeography is often predicated on sampling one individual per site for similar reasons (Avise 2000). Given these considerations, as well as the highly disjunct distribution of *I. missouriensis*, only 13 of 52 populations were surveyed for chloroplast haplotype in multiple individuals. In each of these 13 populations, chosen by their large size and/or proximity to populations with a different haplotype, 7 additional samples were sequenced. No variation was found in 12 of the 13 populations. The thirteenth population (WY13 in Table 1) contained five copies of the most common haplotype, and three singleton haplotypes. Although two of the singleton haplotypes were novel, one of the three singletons was also found at one other location (CO16) and reflects either an independent origin (homoplasy) or past seed flow. All of the four haplotypes found at WY13 were connected by single mutations, but the occurrence of the singleton haplotype L created a closed loop in the species-wide network, i.e. it requires homoplasy. While the WY13 population demonstrates that the assumption of monomorphism for chloroplast haplotype within populations is not strictly valid, as would be expected, the data well support the use of this assumption as an approximation, given that there is no indication that

haplotype sharing among populations occurs at an appreciable frequency. Thus, single samples per population should provide an accurate estimation of the geographic structure of maternal lineages.

Chloroplast haplotypes of *I. missouriensis* and *I .missouriensis* ssp. *longipetala* obtained from the University of California Botanic Garden (UC in Table 4-1) were completely within the range of variation in wild-collected samples. Furthermore, samples of *I. douglasiana* and *I. hartwegii* had chloroplast haplotypes that were nearly identical to each other but differed from the closest *I. missouriensis* haplotype by approximately 16 substitutions and 2 indels of 1 and 5 bp in the trnL-trnF spacer alone (>4% divergence, alignment not shown). Thus, the possibility that any specimens used in this study were misidentified, unlikely given the amount of morphological difference between *I. missouriensis*, *I. douglasiana*, and *I. hartwegii*, was nonetheless carefully excluded.

We performed nested-clade analysis of variance (NCA; Templeton 1998) to statistically evaluate whether the spatial distributions of haplotypes and phylogenetically independent higher-order clades are nonrandom across the landscape. This analysis was implemented with the program GeoDis (Posada et al. 2000). The single occurrence of the homoplasious haplotype L was removed from this analysis.

RESULTS

Patterns of nuclear diversity

The nuclear diversity of populations, presented as the mean pairwise genetic distance, is shown in Table 4-1. The highest genetic diversities were found in Colorado and California in the middle latitudes of the range. These regions are also the regions of

highest present-day abundance (Davidson 1957, R. S. C., pers. obs.). The lowest diversities were found in the northern part of the range, i.e. roughly north of 42° N latitude (equivalent to the northern borders of California, Nevada, and the panhandle of Utah). Seven of the eight least diverse populations were in the northern part of the range and all populations above this latitude were in the bottom 50% of diversity. Populations on the southern margin had average levels of diversity, and over the whole range there is no significant correlation between latitude and diversity.

The neighbor-joining dendrogram of the population-mean, pairwise Dice similarity matrix is shown in Figure 4-2. The percentage support from 100 bootstrapped data sets is shown for each node with support >50%. The tree is rooted with the populations CA41 and CA42, which by PCR assay do not contain intact copies of the LTR retrotransposon subfamily defined by the B2 sequence (see Methods above and Chapter Two) that is present in all other samples. The number of B2 bands amplified is given next to each population.

Bootstrap support for most nodes is moderate to low, as expected given that reciprocal monophyly between conspecific populations should rarely occur for nuclear genes. Population CA37 of the Sierra Nevada appears to be iconoclastic in its relationships and its inclusion in the dendrogram substantially lowers bootstrap values throughout the tree (results not shown). CA37 is a clear outlier in the principal components analysis as well and its position may reflect either gene flow or lineage sorting.

Several geographically concordant (with important exceptions, see below) clusters are evident in the tree, including a southern Rocky Mountain cluster, a Trans-

Colorado Plateau cluster, and a Sierra Nevada/southwestern California cluster. Such groupings represent greater connectivity within than between these geographic regions. Nevertheless, branch lengths for populations within these clusters are typically much longer than branches between these clusters, indicating that populations within regions remain distinct despite the effects of shared history or ongoing gene flow. This strong nuclear differentiation is clearly evidenced by AMOVA ($\Phi_{PT} = 0.62$, $P < 0.001$). This level of population structuring is high for a mixed-mating, widely dispersed monocot (cf. Hamrick and Godt 1996), reflecting low gene flow among disjunct populations, perhaps in conjunction with inbreeding (which has not been quantified in this species). Geographic regions defined using the neighbor-joining tree as a guide (Fig. 4-2) explained only 21.6% of the overall genetic structure, again indicating that the majority of divergence is among populations within regions, rather than among regions.

The Mantel correlation of genetic and geographic distance was significant ($R^2 = 0.13$, $P < 0.001$), indicating support for an isolation-by-distance effect of gene flow; however the strength of this correlation was relatively weak and is in accord with the AMOVA results. PCA partitioned approximately half of the total nuclear variation along three axes (Axes 1 and 2 are shown in Fig. 4-3). The first axis constitutes ~27% of the total variation and clearly separates the southern Rocky Mountains (including the Pacific Coastal populations CA40 and CA43, and the *I. douglasiana* populations *I. doug.*1 and *I. doug.*2) from the rest of the range. The second axis constitutes ~12% of the variation and is driven by the distinctiveness of the Trans-Colorado Plateau group, particularly with respect to more northerly populations. A third axis (not shown) represents ~9% of the variation and is driven by the differentiation of the Sierra Nevada cluster.

Variation in the abundance of the B2-containing IRRE1A subfamily provides information on the polarity of the neighbor-joining clusters. Since the B2 element is present in numerous copies (as determined by transposon-SSAP) in all other populations and is also present in the more derived species *I. douglasiana* (Chapter Two), the assumption that absence of B2 loci is ancestral is well-supported. Based on this rooting, as well as mapping the change in the number of B2 bands, we can infer that the direction of migration was from the western portion of the range to the Rocky Mountains based on its interior position (Fig. 4-3). Surprisingly, this colonization appears to derive from ancestors more similar to coastal California populations than Sierra Nevada populations: populations CA40 and CA43, as well as the *I. douglasiana* samples, lie within the otherwise contiguous southern Rocky Mountain cluster.

Geographic structure of chloroplast haplotypes

A total of 9 polymorphic sites were observed in the concatenated chloroplast sequence (887 bp aligned), seven in the *trnL-trnF* spacer and two in the partial *matK* gene. One of the polymorphic sites in the spacer was a three-state indel associated with a mononucleotide repeat. This mononucleotide repeat shows potentially homoplasic variation among *Iris* species based on an alignment of GenBank sequence (data not shown) and the probability of homoplasy for this length mutation is intrinsically higher than for nucleotide substitutions due to replication slippage. However, the geographic continuity and lack of within-population polymorphism for this character indicate that homoplasy at this site is not significant in the data set. A single minimum-spanning network of eleven haplotypes could readily be determined requiring a single parallel

mutation (Fig. 4-4A). All haplotypes but one were separated by a single step; the two-step haplotype D required the inference of an unsampled intermediate C.

NCA of haplotype distributions was significantly nonrandom, as were the distribution of all three one-step clades (Table 4-2). Clades 1-1 and 1-3 are significantly restricted in distribution relative to the total sampling range, whereas clade 1-2 had a wider range than expected by chance. Comparisons of the distribution of interior versus tip clades using the interpretive key (version 2.4) of Posada and Templeton (2005) were inconclusive with respect to whether range expansion of some haplotypes or isolation of other haplotypes contributed significantly to the observed nonrandom distribution.

Haplotype A (Fig. 4-4B) is the most geographically dispersed, most abundant, and most interior haplotype in the network and is connected to five tip haplotypes. This combination of frequency and network position strongly suggests that haplotype A is an old and perhaps ancestral haplotype. The southern Rocky Mountains, a region of high sampling density and natural abundance, contains only haplotype A, and four low-frequency tip clades of A which presumably have risen to high frequency at local scales but have not dispersed geographically. A similar pattern of uniformity is observed in the upland regions of Utah north of the Colorado River. Haplotype D (clade 1-1) was the only haplotype sampled in this region and may have been the dominant haplotype for a relatively long period of time. This is suggested by the fact that it is two steps from haplotype A and the inferred intermediate, haplotype C, was not recovered.

Both Clades 1-2 and 1-3 are present in the northern part of the range (Fig. 4-4B). While the low nuclear diversity in this region is consistent with extinction/recolonization dynamics, the presence of two cpDNA clades indicates multiple sources of immigrants.

Collectively, populations in California show the most haplotype diversity, with clades 1-2 and 1-3 equally represented. NCA with respect to geographic distance showed no significant structuring of either clade at the scale of the entire state, although B is the predominant haplotype at the western margin and A the predominant haplotype of the southern and eastern margins. The lack of nesting clade geographic structure plus the presence of tip clades of both 1-2 and 1-3 suggest that these clades have coexisted in California for some time.

Nuclear differentiation, as assessed by AMOVA, was almost completely independent of chloroplast differentiation. When populations were assigned to one of three groups according to which one-step clade the population haplotypes belonged, this hierarchical organization explained only 2% ($P < 0.001$) of the nuclear variation. This result occurred despite the fact that the entire southern Rocky Mountains forms a distinct clade based on nuclear variation and contains only chloroplast clade 1-2. In contrast, and reflecting the AMOVA results, the Sierra Nevada show nuclear marker clustering despite the presence of multiple haplotype clades and the Utah highlands show striking nuclear differentiation despite the presence of only a single haplotype.

DISCUSSION

Iris missouriensis presently occurs in populations that are often highly disjunct, such as the sky islands of the Great Basin and southwestern United States and seasonally wet meadows of the Columbia plateau, and is apparently in low abundance throughout these portions of its range (R.S.C. pers. obs., Davidson 1957). These conditions prohibit frequent pollen or seed dispersal between contemporary populations. We therefore

expected that substantial among-population differentiation would be present due to drift. However, given the paleoclimatic history of habitat connectivity, long coalescence of nuclear genes, and slow evolution of chloroplast DNA, we nonetheless expected that hierarchical patterns of regional variation would be informative with respect to demographic history. While the regional component of AMOVA is in fact statistically significant, as is the Mantel test of isolation-by-distance, the proportion of variation explained by these statistics is relatively small (21% and 13%, respectively). Most of the total variation is among populations within regions. The simplest explanation for the generally high variation between nearby populations is genetic drift acting on a highly polymorphic ancestral lineage. This is reasonable given the self-compatibility and vegetative spread of irises. Even so, most populations have retained a surprising level of genetic diversity (Table 4-1).

While the geographic component of the overall variation at nuclear loci is not high, neighbor-joining cluster analysis and PCA identified three large geographic regions that are genetically cohesive and a number of smaller groupings. The three large groupings are the southern Rocky Mountains, a trans-Colorado Plateau group, and the Sierra Nevada. Differentiation of the southern Rocky Mountain cluster explains the greatest proportion of the variation and is concordant with the cpDNA analysis; this is in contrast to the Trans-Colorado group, which includes all three one-step cpDNA clades. While this latter group is well supported by bootstrap values for the neighbor-joining tree, the PCA illustrates that these populations are much less clustered than the larger southern Rocky Mountain group. Early taxonomic treatments identified separate species in the southwestern U.S. (*I. arizonica* and *I. pelogonus*) that were later considered synonymous

with *I. missouriensis* (Dykes 1913). While we lack sufficient morphological data for specimens from collected sites to evaluate any congruity between the observed genetic cluster and the described *arizonica* or *pelogonus* phenotype, the geographic pattern is suggestively similar. A common garden of samples from within this cluster as well as elsewhere would be necessary to further explore the genetic basis of any purported morphological variety.

For the various phylogeographic questions described in the introduction, the weight of evidence discriminating between alternate hypotheses varies for each question. While in many cases we can evaluate which hypothesis is most favored by the data, it is not presently possible to formulate alternatives in such a way as to model the probabilities of each, which is the goal of "statistical phylogeography" as advocated by Knowles and Maddison (2002). This is a weakness of dominant fragment analysis, which is not amenable to coalescent modeling. Our approach here has been to provide an initial survey of range-wide patterns of variation and a qualitative assessment of how such variation accords with proposed phylogeographic hypotheses. We believe such an approach is essential for subsequent statistical tests of phylogeography, because they guide the scale of sampling and marker resolution required for powerful tests of alternatives.

Phylogeographic hypotheses

Dynamics at the northern and southern margins. Mean within-population genetic distance, a proxy for nuclear diversity, is highest in the center of the species distribution (the Rocky Mountains of Colorado and in central California) indicating that these two regions have maintained larger effective population sizes than other geographic areas.

The northern part of the range had lower diversity than the southern part of the range, indicating that the loss of diversity associated with paleoclimatic oscillations was greater in the cold arid conditions of the northern Great Basin than in the southern limits of *I. missouriensis*. We interpret this as evidence that populations at the southern edge persisted in high-elevation refugia during unfavorable climates whereas northern-edge populations were largely extinguished. Furthermore, there is no evidence of a single refugial source for recolonization of the northern range: multiple cpDNA haplotypes and multiple nuclear clusters are present. Two of the three nuclear clusters are consistent with expansion along the eastern slope of the Cascades/western Columbia Plateau from a source south of the glaciated Cascades (Groups 1 and 3 in Fig. 4-2). There is a rather dramatic increase in B2 band number in Group 3 from south to north, further evidence of a south-to-north polarity of migration. Group 2, on the eastern side of the Great Basin, is more closely related to Group 3 than to the trans-Colorado Plateau or Southern Rocky Mountain group, indicating that these latter regions were not significant contributors of migrants northward during the current interglacial cycle. More likely, populations in the northern pluvial lake region connected eastern and western lineages that subsequently moved northward as well as upward in elevation (e.g., UT28 and NV31).

Discontinuity between northern and southern Rocky Mountains. Some previous work in plants (DeChaine and Martin 2005) has shown genetic discontinuities across the arid Wyoming Basin separating the northern and southern Rocky Mountains. The present study did not identify a strong break across the Wyoming Basin, although populations WY11 and WY12 north of the Wyoming Basin are more closely related to each other than any southern Rocky Mountain populations and are clearly outside of the main

southern Rocky Mountain cluster based on PCA analysis. It is also clear that populations in the northern Rocky Mountains do not form a cohesive cluster with respect to either nuclear loci or cpDNA haplotype. Since *I. missouriensis* is a montane plant capable of persisting in mesic microhabitats of the plains, it is not surprising that the discontinuity is small between the two regions of the Rocky Mountains. Interestingly, the southern region is much more uniform than the northern region, suggesting that habitat in the latter region was less connected during paleoclimatic change.

Connectivity of the Sierra Nevada, southern (pluvial) Great Basin, and southern Rocky Mountains. There is minimal evidence that gene flow occurred between the Sierra Nevada and the southern Rocky Mountains during the last interglacial. The axis that explained the largest component of variation (~27%) clearly separates the southern Rocky Mountains from the rest of the range (Fig. 4-3). While two populations of the Sierra Nevada (CA33, CA36) are in the group that is maximally distant from the southern Rocky Mountains, three populations (CA34, CA35, and CA37) are intermediate along this axis, perhaps due to gene flow. However, the neighbor-joining branch length between the southern Rocky Mountain and the Sierra Nevada clusters is rather high, and Great Basin populations do not fall out as intermediates. While the geographic distance between these regions is large, similar distances in the southern Rocky Mountains have not resulted in equivalent genetic structure (see below).

Connectivity of the Basin and Range province of the southern Rocky Mountains. Perhaps the most remarkable conclusion we can draw from the present data set is the occurrence of a dispersal event in which the southern Rocky Mountains were colonized by founders genetically similar to individuals present today in coastal California. This

dispersal event can be inferred from the nesting of coastal Californian populations of *I. missouriensis* as well as the west coast endemic *I. douglasiana* in an otherwise monophyletic southern Rocky Mountain clade. This pattern of genetic similarity was also observed in transposon-SSAP results (not shown) for a different LTR retrotransposon subfamily (IMRE1, see Chapter Two), which further supports this inference. The putative source of the southern Rocky Mountain colonization is counter to expectation based on biogeography. Typically, the eastern Sierra Nevada flora shows high affinity with the Great Basin and Cascade floras and little affinity with central coastal California (e.g. Patterson and Givnish 2003, Beardsley et al. 2004, Kimball et al. 2004). The inclusion of *I. douglasiana* in the cluster indicates that some of the present-day diversity of B2 sequences predates the divergence of these taxa and a subset of *I. missouriensis* has retained an ancestral genotypic pattern. The close spatial proximity of these California populations argues against an accident of lineage sorting and instead indicates that the 'common ancestor' genotype has persisted in this geographic region. Hybridization between the species is not a plausible explanation of this pattern because it would require extensive introgression between the species and hybrids have not been reported in the wild. Additionally, chloroplast haplotypes differ by almost five percent sequence divergence at the *trnL-trnF* intergenic spacer (see Materials and Methods), indicating the large divergence between these taxa.

A recent colonization of the southern Rocky Mountains by A-haplotype founders is supported by the fixation of chloroplast clade 1-2 in this region. Loss of genealogically deep haplotype diversity is expected with founder events (Nei et al. 1975) and has been observed in some plant phylogeographic studies (Petit et al. 2001) in which colonization

of an unoccupied habitat and subsequent rapid population growth results in regions of haplotype uniformity with abrupt borders. Reduced nuclear genetic structure in the southern Rocky Mountains ($\Phi_{PT} = 0.46$ [$P < 0.001$] versus $\Phi_{PT} = 0.62$ overall) also argues for a recent range expansion and thus greater connectivity. However, it need not be literally true that seed traversed the entire distance in a single physical event; rapid colonization of intermediate areas followed by extinction of stepping-stone populations following climate change would produce similar genetic relationships. Indeed, we located few populations within the entire Great Basin region, limiting our ability to infer intermediate steps in the southern Rocky Mountain colonization. The close relationship between populations OR6 and WA3 in the Northwest clearly suggests a recent long-distance dispersal, an important indicator of the potential for rapid movement across intervening poor habitat.

Additional phylogeographic patterns. The Colorado River drainage as it passes through the sandy strata of Utah and Arizona seems to have been a barrier for cpDNA but only weakly for nuclear DNA. Additional sampling in this region, coupled with comparative analyses across multiple taxa, would be desirable to confirm the presence of a phylogeographic break. There is evidence of vicariance and isolation attributable to the Colorado River from some animal systems such as *Microtus* but not others (Conroy and Cook 2000). Surprisingly, the highlands of Utah show no genetic continuity in *I. missouriensis* at nuclear loci despite the uniformity of cpDNA haplotype and apparent habitat continuity.

Conservation implications

It is clear that the three populations of *I. missouriensis* ssp. *longipetala* (CA40, CA41, and CA42) sampled for transposon-SSAP are not notably distinct in their nuclear genomes from other *I. missouriensis* populations nor are they monophyletic. The latter observation may reflect a difficulty of taxonomy rather than biological reality since distinguishing morphological characters are few and prone to plasticity (e.g., leaf size). In contrast, the seven *I. missouriensis* ssp. *longipetala* populations for which chloroplast haplotype was assayed (which includes four additional samples provided by the University of California Botanic Garden) demonstrate a distinct bias: 6 of 7 had a haplotype of clade 1-3. Furthermore, chloroplast haplotype diversity was elevated generally within the '*longipetala*' range (in the vicinity of San Francisco Bay and the northern and southern continuations of the Coast Ranges), with four haplotypes detected. This suggests a long history of *I. missouriensis* ssp. *longipetala* and *I. missouriensis* in general in this region yet a possible deviation from cytoplasmic equilibrium. Given that most historical *longipetala* populatons have been extirpated by urbanization based on collection data of the University of California Jepson Herbarium (2004), informative nuclear and chloroplast diversity has probably been lost in this region, as has plant biodiversity generally. It should be pointed out that genetic divergence related to local adaptation need not be detectable by bulk marker analysis, and that the ecological properties of the species in the Mediterranean climates of California are undoubtedly distinct from elsewhere in the range. It would be unfortunate if the weediness of *I. missouriensis* in heavily grazed rangelands of the Eastern Sierran slope, which has resulted in the placement of the taxon on the noxious weed list of that state, were to

prevent conservation management of *I. missouriensis* ssp. *longipetala* in its few remaining habitats.

ACKNOWLEDGMENTS

We gratefully acknowledge the following individuals for providing advice and assistance in locating I. missouriensis populations: L. Weeth, N. Henderson, R. Richards, E. Pardini, and D. Trapnell. Population WA1 was kindly provided by P. de la Chapelle, and additional material was provided by the University of California Botanic Garden as stated in the text and under the guidance of H. Forbes. This work was supported by a grant from the American Iris Society (R.S.C.), National Science Foundation Training Grant DBI 960223 (R.S.C.), and National Science Foundation Grant DEB-0345123 (M.L.A.). R.S.C. was also supported by a University of Georgia Presidential Fellowship.

Table 4-1. Sample locations and estimated within-population genetic diversity. See Figure 4-1 for population locations. The symbol -- represents locations for which chloroplast but not nuclear data were collected.

Population Code	County	State	Mean Pairwise Genetic Distance Within Population	Rank of Within-Population Genetic Distance
WA1	Island	Washington	0.21	11
WA2	Kittitas	Washington	0.33	22
WA3	Spokane	Washington	0.02	1
OR4	Wasco	Oregon	0.07	5
OR5	Deschutes	Oregon	0.33	21
OR6	Baker	Oregon	0.06	3
OR7	Lake	Oregon	0.09	6
MT8	Granite	Montana	0.03	2
MT9	Silver Bow	Montana	0.24	15
ID10	Butte	Idaho	0.24	16
WY11	Park	Wyoming	0.06	4
WY12	Crook	Wyoming	0.18	8
WY13	Albany	Wyoming	0.23	13
CO14	Park	Colorado	0.38	31
CO15	Huerfano	Colorado	0.38	32
CO16	Delta	Colorado	0.51	43
CO17	San Miguel	Colorado	0.44	38
UT18	San Juan	Utah	0.47	40
CO19	Hinsdale	Colorado	0.36	27
NM20	Cibola	New Mexico	0.34	25
NM21	Bernalillo	New Mexico	0.41	33
NM22	Lincoln	New Mexico	0.48	41
NM23	Grant	New Mexico	0.13	7
AZ24	Cochise	Arizona	0.43	36
AZ25	Greenlee	Arizona	0.27	18
AZ26	Coconino	Arizona	0.43	34
UT27	Kane	Utah	0.50	42
UT28	Garfield	Utah	0.21	10
UT29	Wasatch	Utah	0.44	37
UT30	Uintah	Utah	0.37	30
NV31	Elko	Nevada	0.30	19
NV32	Lander	Nevada	0.23	14
CA33	Mono	California	0.31	20
CA34	Mono	California	0.43	35
CA35	Tuolumne	California	0.27	17
CA36	Alpine	California	0.20	9
CA37	Ventura	California	0.34	24
CA37	Inyo	California	0.37	29
CA38	San Diego	California	0.22	12

CA40	Marin	California	0.36	28
CA41	Marin	California	0.35	26
CA42	Mendocino	California	0.34	23
CA43	Glenn	California	0.45	39
CA-a	Alpine	California	--	--
CA-b	Mono	California	--	--
CA-c	Mono	California	--	--
CA-d	Mono	California	--	--
CBG1	Marin	California	--	--
CBG2	San Francisco	California	--	--
CBG3	Marin	California	--	--
CBG4	San Mateo	California	--	--
CBG5	Siskiyou	California	--	--
CO-a	Pitkin	Colorado	--	--
CO-b	Saguache	Colorado	--	--
CO-c	Gunnison	Colorado	--	--
WY-a	Albany	Wyoming	--	--

Table 4-2. Patterns of Nonrandom Distribution of Chloroplast Haplotypes Determined by Nested Clade Analysis. Dc = the average distance of samples from the geographic center of the clade. Dn = the average distance of samples from the geographic center of the nesting clade. Letters next to Dc and Dn probabilities indicate significantly large (L) or significantly small (S) distances relative to expectation. NS = not significant at alpha = 0.025 for each hypothesis, i.e. a two-tailed alpha = 0.05.

Clade	Permutational Contingency Probability	Interior Clade	Tip Clade	Dc	Probability	Dn	Probability
Clade 1-2	0.024	A		664.2	NS	665.7	NS
			G	0	NS	922.3	NS
			H	0	NS	803.6	NS
			I	58.9	NS	487.3	NS
			J	58.1	NS	671.1	NS
			K	0	NS	657.7	NS
		Interior-Tip Distance		630.8	NS	NS5.8	NS
Clade 1-3	0.000	B		392.7	NS	395.2	NS
			E	0	NS	734.1	NS
			F	13.9	NS	388.1	NS
		Interior-Tip Distance		389.9	0.005 L	NS269.7	NS
Total Cladogram	0.000	Clade 1-2		689.8	NS	695.6	NS
			Clade 1-1	144.1	0.001 S	158.8	0.000 S
			Clade 1-3	203.9	NS	837.9	NS
		Interior-Tip Distance		529.8	0.000 L	355.7	0.000 L

Figure 4-1. Map of sampling populations described in Table 4-1. Filled circles represent populations for which both transposon-SSAP and cpDNA data were collected, whereas unfilled circles represent populations for which only cpDNA data were collected.

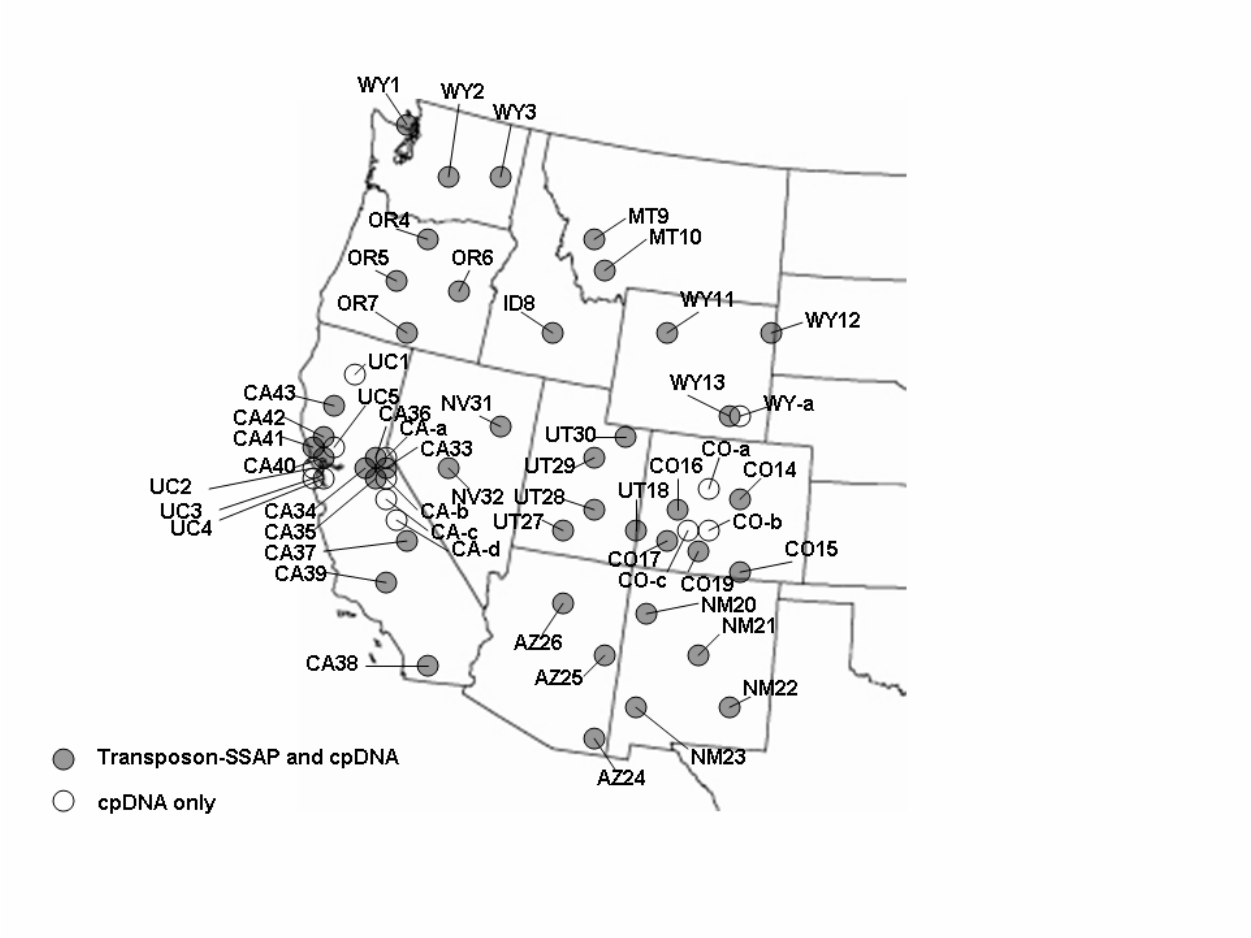

WY1
WY2
WY3
OR4
MT9
MT10
OR5
OR6
OR7
ID8
WY11
WY12
UC1
WY13
WY-a
CA43
UC5
CA36
NV31
CA42
CA-a
CA41
CA33
UT30
CO-a
CA40
UT29
CO16
CO14
UC2
NV32
UT28
UT18
UC3
CA34
CA-b
UT27
CO-b
UC4
CA35
CA-c
CA37
CA-d
CO17
CO15
CA39
CO-c
CO19
NM20
NM21
AZ26
CA38
AZ25
NM22
NM23
AZ24
Transposon-SSAP and cpDNA
cpDNA only

Figure 4-2. A neighbor-joining dendrogram of the population genetic-distance matrix. Bootstrap support from 100 resampled data sets is presented for nodes with values ≥ 50. Labelled groups are discussed in the text and were used to define regions for AMOVA (excluding southern California). The mean number of B2 bands is given for each population (see text for details). I. doug. = *I. douglasiana.*

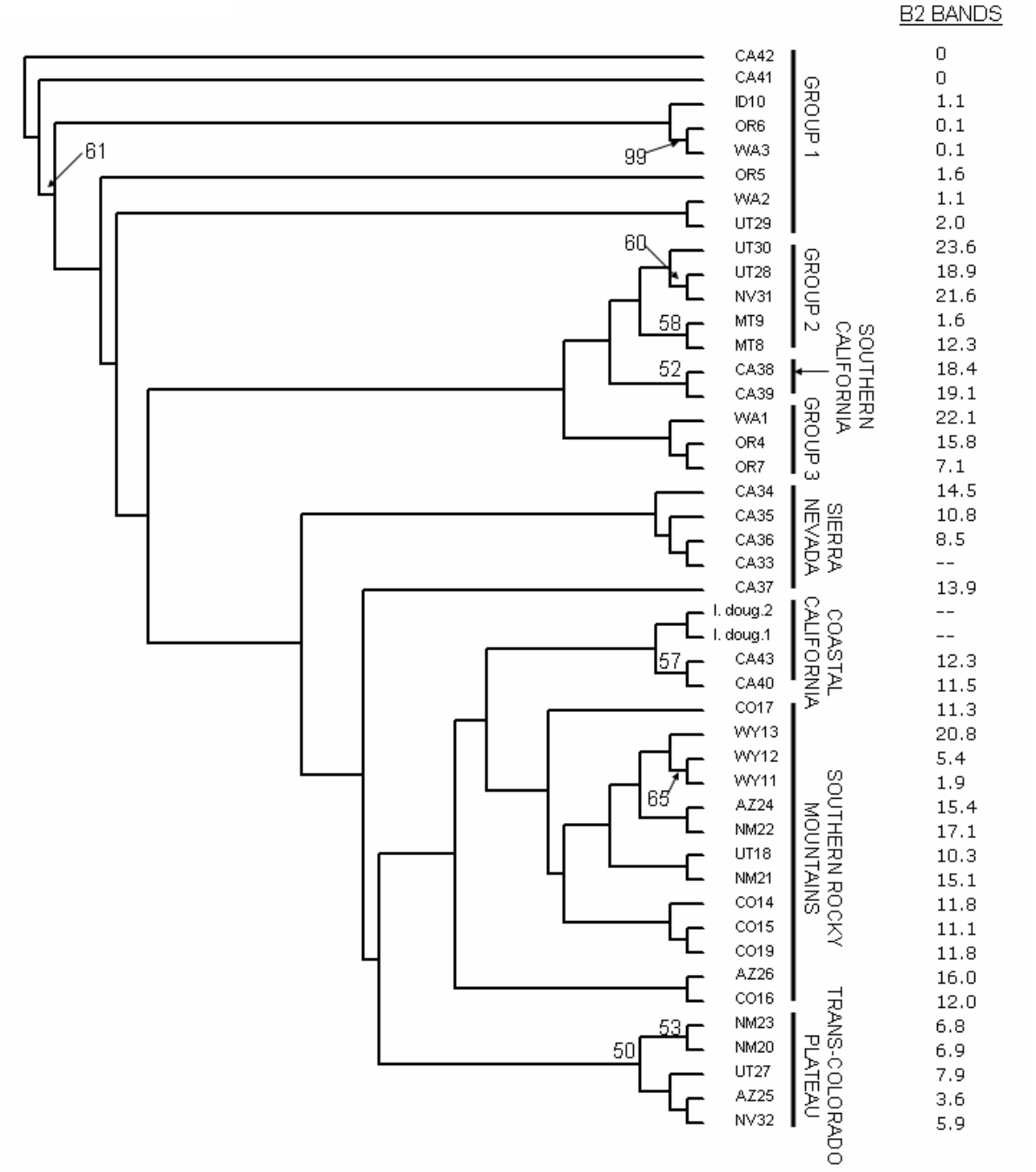

B2 BANDS
CA42 0
CA41 0
ID10 1.1
OR6 0.1
WA3 0.1
OR5 1.6
WA2 1.1
UT29 2.0
UT30 23.6
UT28 18.9
NV31 21.6
MT9 1.6
MT8 12.3
CA38 18.4
CA39 19.1
WA1 22.1
OR4 15.8
OR7 7.1
CA34 14.5
CA35 10.8
CA36 8.5
CA33 --
CA37 13.9
l. doug.2 --
l. doug.1 --
CA43 12.3
CA40 11.5
CO17 11.3
WY13 20.8
WY12 5.4
WY11 1.9
AZ24 15.4
NM22 17.1
UT18 10.3
NM21 15.1
CO14 11.8
CO15 11.1
CO19 11.8
AZ26 16.0
CO16 12.0
NM23 6.8
NM20 6.9
UT27 7.9
AZ25 3.6
NV32 5.9
GROUP 1
GROUP 2
GROUP 3
SOUTHERN CALIFORNIA
SIERRA NEVADA
COASTAL CALIFORNIA
SOUTHERN ROCKY MOUNTAINS
TRANS-COLORADO PLATEAU
61
99
60
58
52
57
65
53
50

Figure 4-3. Plot of the first two principal components axes for the population genetic-distance matrix, as determined with GenAlEx 6 (Peakall and Smouse 2005). The percentage of variation explained by each axis is presented on the axis label. Population codes are as given in Table 1.

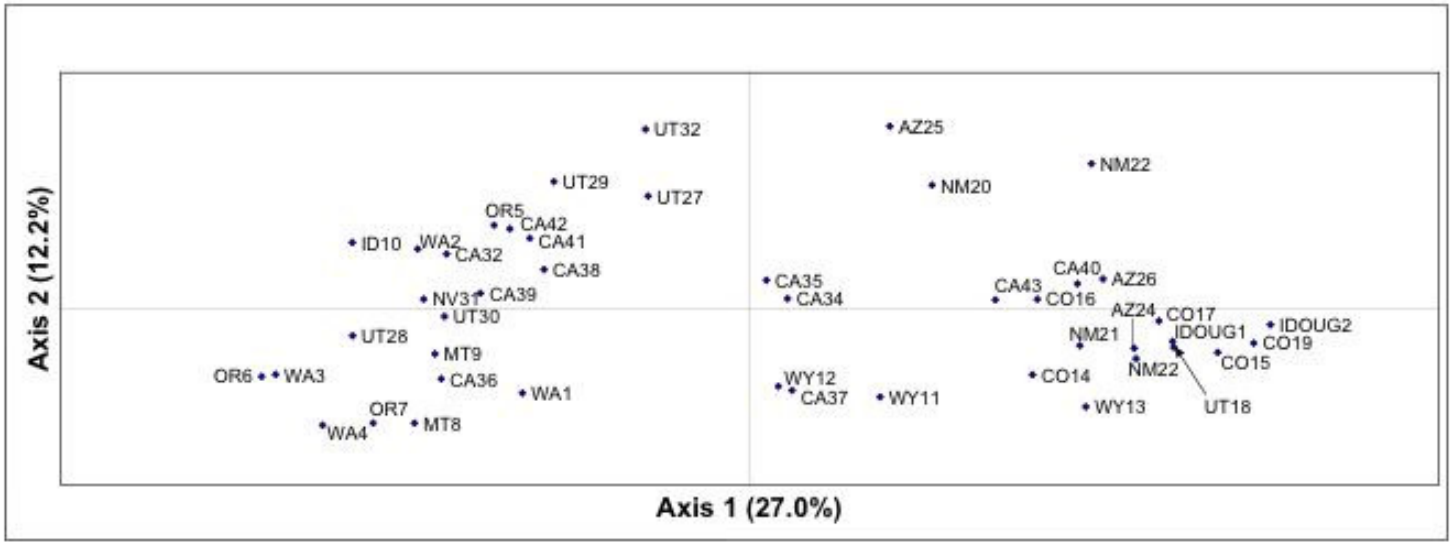
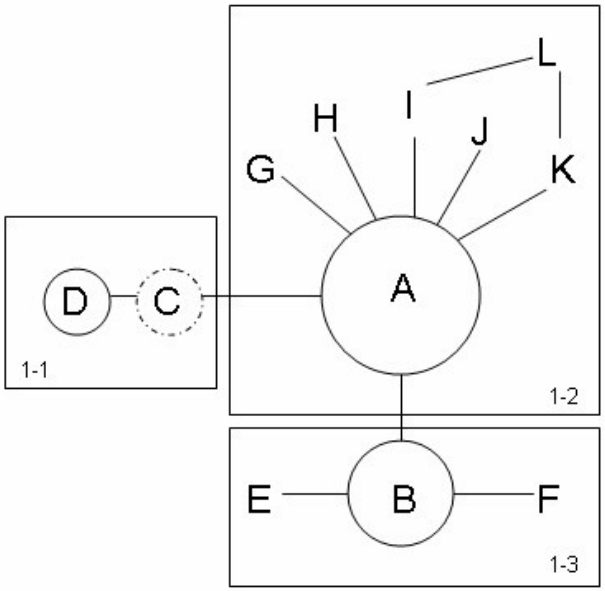
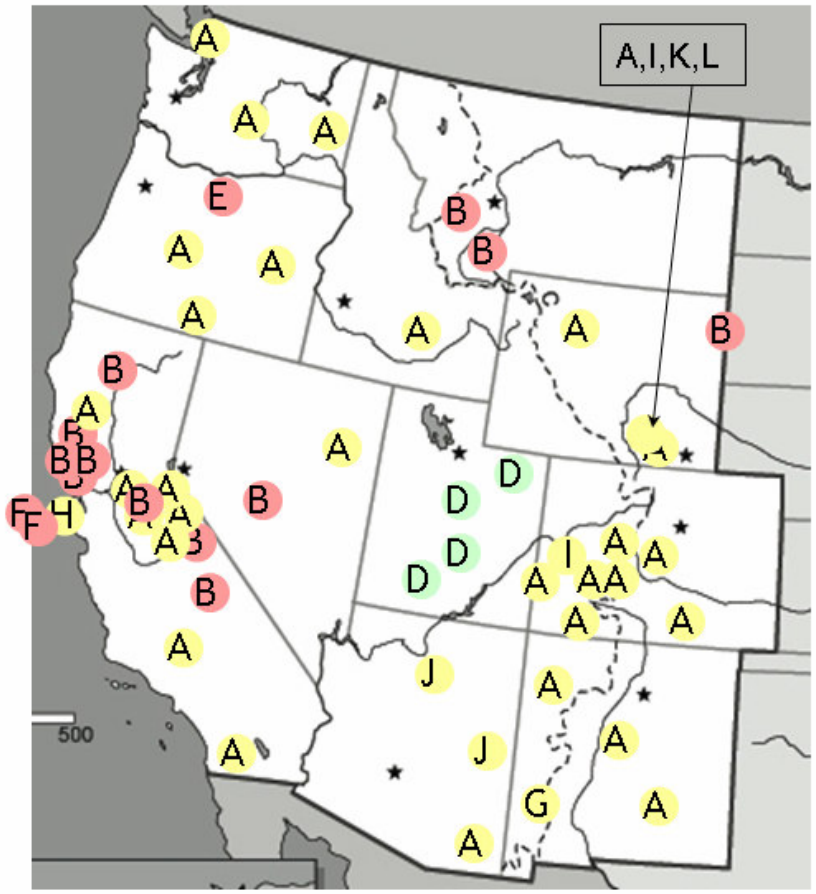

Axis 2 (12.2%)
Axis 1 (27.0%)
UT32
UT29
UT27
OR5
CA42
ID10
WA2
CA41
CA32
CA38
NV31
CA39
UT30
UT28
MT9
OR6
WA3
CA36
WA1
OR7
WA4
MT8
AZ25
NM22
NM20
CA35
CA40
AZ26
CA43
CO16
CA34
AZ24
CO17
IDOUG2
NM21
IDOUG1
CO19
CO15
NM22
WY12
CO14
CA37
WY11
WY13
UT18

Figure 4-4. A) The minimum spanning network of chloroplast haplotypes. B) Map of chloroplast haplotype distributions by sampling location.

A



B



CHAPTER 5

CONCLUSIONS

Collectively, the studies described in the previous three chapters provide new insights into the forces that structure gene flow in natural populations of *Iris*. We have pursued both fine-scale (Chapter Two) and range-wide (Chapters Three and Four) characterizations of genetic structure that are each in their own way relevant to the process of speciation. Our study of Louisiana Iris hybrids (Chapter Two) demonstrates the importance of postzygotic selection in structuring hybrid zones, but in a manner different from what has been commonly assumed by much of the literature on natural hybridization until recently (reviewed in Arnold 1997). Previous studies of Louisiana Iris hybrids (Emms and Arnold 1997, Burke et al. 1998a, Johnston et al. 2003) had shown that some hybrid genotypes have high fitness and that selection is at least in part environment-dependent, contrary to the assumptions of the tension-zone model of hybridization (Barton and Hewitt 1985). Our study (Chapter Two) revealed that postzygotic selection can maintain partially isolated hybrid lineages in close proximity, providing the conditions for lineages of high fitness to expand demographically. Furthermore, additional greenhouse studies (Burke et al. 1998b and Martin et al. 2005) and field studies (Martin et al. 2006) have demonstrated the importance of postzygotic, epistatic interactions in determining hybrid fitness, including the identification of QTL with large fitness effects. Thus, there appears to be ample support for the existence of genetic factors that could quickly generate strong selection against recombination once a fit hybrid lineage has been created. Indeed, postzygotic selection has previously been shown to canalize hybridization outcomes in *Helianthus*, another model system

of hybridization and speciation, independently of phenotypic selection (Rieseberg et al. 1996). It is intriguing to speculate that the mechanisms of inbreeding depression and postzygotic selection on recombinant genotypes observed in Louisiana Irises have a similar basis and are related to the high abundance of IRRE retroelements in their genomes. This speculation is sparked by the high mutation rate inferred for the IMRE1 element in *I. missouriensis* (Chapter Three), which might reflect in part a high rate of nonhomologous recombination. Nonhomologous recombination may result in relatively large genomic deletions (Devos et al. 2002) which could result in an accelerated production of null alleles for 'host' genes. Based on theoretical expectations regarding the phenotypic effect of null alleles (Orr 1991), negative selection would most likely be greatest in homozygotes and thus could be an important component of inbreeding depression. Moreover, transposon-mediated deletions that interact epistatically could contribute to the inferred postzygotic isolation of hybrid lineages. Unfortunately, isolation of the actual genetic basis of selection is a Herculean task even for the most advanced model systems such as Arabidopsis, so it is unlikely that this speculative hypothesis will be testable in the near future.

BIBLIOGRAPHY

Adams, W., A. Griffin, and G. Morgan. 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. *Am. Nat.* 140: 762-780.

Anderson, E. 1949. Introgressive Hybridization. New York, John Wiley and Sons.

Anderson, E., and G. L. Stebbins. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8: 378-388.

Arbogast, B. and G. J. Kenagy 2001. Comparative phylogeography as an integrative approach to historical biogeography. *J. Biogeog.* 28: 819-825.

Arnold, M. L. 1993. *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. *Am. J. Bot.* 80: 577-583.

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

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, Oxford University Press.

Arnold, M. L., D. D. Shaw, and N. Contreras. 1987. Ribosomal RNA encoding DNA introgression across a narrow hybrid zone between two subspecies of grasshopper. *Proc. Natl. Acad. Sci. USA* 84: 3946-3950.

Arnold, M. L., J. J. Robinson, C. M. Buckner, and B. D. Bennett. 1992. Pollen dispersal and interspecific gene flow in Louisiana irises. *Heredity* 68: 399-404.

Arnold, M. L., J. L. Hamrick, and B. D. Bennett. 1993. Interspecific pollen competition and reproductive isolation in *Iris*. *J. Heredity* 84: 13-16.

Arnold, M. L., A. C. Bouck, and R. S. Cornman. 2004. Verne Grant and Louisiana Irises: Is there anything new under the sun? *New Phyt.* 161: 143-149.

Avise, J. C. 2000. Phylogeography. Cambridge MA, Harvard University Press.

Barton, N. H. 2001. The role of hybridization in evolution. *Mol. Ecol.* 10: 551-568.

Barton, N. H. and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16: 113-148.

Beardsley, P. M., S. E. Schoenig, J. B. Whittall, and R. G. Olmstead. 2004. Patterns of evolution in western North American *Mimulus* (Phrymaceae). *Am. J. Bot.* 91: 474-489.

Bennett, B. D. and J. B. Grace. 1990. Shade tolerance and its effect on the segregation of two species of Louisiana iris and their hybrids. *Am. J. Bot*. 77:100-107.

Bouck, A. C., R. Peeler, M. L. Arnold, and S. R. Wessler. 2005. Genetic mapping of species boundaries in Louisiana Irises using IRRE retrotransposon display markers. *Genetics* 171: 1289-1303.

Brunsfeld, S. J., J. Sullivan, D. E. Soltis, and P. S. Soltis. 2001. Comparative phylogeography of northwestern North America: a synthesis. *In*: Integrating Ecological and Evolutionary Processes in a Spatial Context: 14th Special Symposium of the British Ecological Society. J. Silvertown and J. Antonovics, eds. Oxford, Blackwell Science: 319-340.

Buerkle, A. C., R. J. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. *Heredity* 84: 441-451.

Burke, J. M. and M. L. Arnold. 1999. Isolation and characterization of microsatellites in *Iris*. *Mol. Ecol.* 8: 1075-1092.

Burke, J. M., S. E. Carney, and M. L. Arnold. 1998a. Hybrid fitness in the Louisiana Irises: analysis of parental and F1 performance. *Evolution* 52: 37-43.

Burke, J. M., T. J. Voss, and M. L. Arnold. 1998b. Genetic interactions and natural selection in Louisiana Iris hybrids. *Evolution* 52: 1304-1310.

Burke, J. M., M. R. Bulger, R. A. Wesselingh, and M. L. Arnold. 2000. Frequency and spatial patterning of clonal reproduction in Louisiana iris hybrid populations. *Evolution* 54: 137-144.

Campbell, D. R., N. M. Waser, and P. G. Wolf. 1998. Pollen transfer by natural hybrids and parental species in an *Ipomopsis* hybrid zone. *Evolution* 52: 1602-1611.

Charlesworth, B. and D. Charlesworth 1983. The population dynamics of transposable elements. *Genet. Res.* 42: 1-27.

Clegg, M. T. 1980. Measuring plant mating systems. *Bioscience* 30: 814-818.

COHMAP Members. 1988. Climatic changes of the last 18,000 years: observations and model simulations. *Science* 241: 1043-1052.

Conroy, C. J. and J. A. Cook. 2000. Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Mol. Ecol.* 9: 165-175.

Cornman, R. S., J. M. Burke, R. A. Wesselingh, and M. L. Arnold. 2004. Contrasting genetic structure of adults and progeny in a Louisiana Iris hybrid population. *Evolution* 58: 2669-2681.

Cruzan, M. B., M. L. Arnold, S. E. Carney. and K. R. Wollenberg. 1993. cpDNA inheritance in interspecific crosses and evolutionary inference in Louisiana Irises. *Am. J. Bot.* 80: 344-350.

Cruzan, M. B. and M. L. Arnold. 1993. Ecological and genetic associations in an *Iris* hybrid zone. *Evolution* 47: 1432-1445.

Cruzan, M. B. and M. L. Arnold. 1994. Assortative mating and natural selection in an *Iris* hybrid zone. *Evolution* 48: 1946-1958.

Cruzan, M. B., J. L. Hamrick, M. L. Arnold, and B. D. Bennett. 1994. Mating system variation in hybridizing irises: effects of phenology and floral densities on family outcrossing rates. *Heredity* 72: 96-105.

Davidson, B. L. 1957. *Iris missouriensis* and *Iris longipetala*. *Bull. Amer. Iris Soc.* 147: 13-22.

DeChaine, E. G. and A. P. Martin 2005. Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. *Am. J. Bot.* 92: 477-486.

Devlin, B., K. Roeder, and N. C. Ellstrand. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. *Theor. Appl. Genet.* 76: 369-380.

Devos, K., J. K. Brown, and J. L. Bennetzen 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* 12: 1075-1079.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26: 297-302.

Dobzhansky, T. 1937. Genetics and the Origin of Species. New York, Columbia University Press.

Dobzhansky, T. 1970. Genetics of the Evolutionary Process. New York, Columbia University Press.

Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. 19: 11-15.

Duarte, J. M., J. B. dos Santos, and L. C. Melo.1999. Comparison of similarity coefficients based on RAPD markers in the common bean. *Genet. Mol. Biol.* 22: 427-432.

Dykes, W. R. 1913. The Genus Iris. New York, Dover.

Emms, S. and M. L. Arnold. 1997. The effect of habitat on parental and hybrid fitness: transplant experiments with Louisiana Irises. *Evolution* 51: 1112-1119.

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

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton NJ, Princeton University Press.

Ennos, R. 2001. Inferences about spatial processes in plant populations from the analysis of molecular markers. *In:* J. Silvertown, J. Antonovics and N. R. Webb, eds. Integrating Ecology and Evolution in a Spatial Context: 14th Special Symposium of the British Ecological Society. Oxford, Blackwell Science: 49-62.

Fay, M. F., R. S. Cowan, and I. J. Leitch. 2005. The effets of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann. Bot.* 95: 237-246.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package (Version 3.6). Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Foster, R. C. 1937. A cyto-taxonomic survey of North American species of iris. *Contrib. Gray Herbarium* 119: 3-82.

Garcia-Martinez, J. and J. A. Martinez-Izquierdo. 2003. Study on the evolution of the Grande retrotransposon in the *Zea* genus. *Mol. Biol. Evol.* 20: 821-841.

Garnier-Gere, P. and C. Dillmann. 1992. A computer program for testing pairwise linkage disequilibria in subdivided populations. *J. Hered.* 83: 239

Goldblatt, P. 1990. Phylogeny and classification of the Iridaceae. *Ann. Missouri Bot. Gard.* 77: 607-627.

Grant, B. and P. Grant. 1996. High survival of Darwin's Finch hybrids: effects of beak morphology and diets. *Ecology* 77: 500-509.

Grant, V. 1971. Plant Speciation. New York, Columbia University Press.

Guo, S. W. and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.

Haldane, J. 1949. The association of characters as a result of inbreeding and linkage. *Ann. Eugen.* 15: 15-23.

Hampe, A. and R. J. Petit. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecol. Letters* 8: 461-467.

Hamrick, J. L. and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. *Phil. Trans. Roy. Soc. Biol. Sci. Series B* 351: 1291-1296.

Hare, M. P. 2001. Prospects for nuclear gene phylogeography. *Trends Ecol. Evol.* 16: 700-706.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. *Oxf. Surv. Evol. Biol.* 7:69-128.

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

Hewitt, G. M. and K. M. Ibrahim. 2001. Inferring glacial refugia and historical migrations with molecular phylogenies. *In:* J. Silvertown, J. Antonovics and N. R. Webb, eds. Integrating Ecology and Evolution in a Spatial Context: 14th Special Symposium of the British Ecological Society. Oxford, Blackwell Science: 271-294.

Hodges, S. A., J. M. Burke, and M. L. Arnold. 1996. Natural formation of *Iris* hybrids: experimental evidence on the establishment of hybrid zones. Evolution 50: 2504-2509.

Howard, D. J., J. L. Marshall, D. D. Hampton, et al. 2002. The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. *Am. Nat.* 159: S8-S21.

Huff, D. R., R. Peakall, and P. E. Smouse. 1993. RAPD variation within and among populations of outcrossing buffalograss (*Buchloe dactyloides* (Nutt.) Engelm). *Theor. Appl. Genet.* 86: 927-934.

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

Johnston, J. A., M. L. Arnold, and L. A. Donovan. 2003. High hybrid fitness at seed and seedling life history stages in Louisiana Irises. *J. Ecol.* 91: 438-446.

Kadereit, J. W., E. M. Griebeler, and H. P. Comes. 2004. Quaternary diversification in European alpine plants: pattern and processes. *Phil. Trans. Roy. Soc. Biol. Sci. Series B* 359: 265-274.

Kentner, E. K., M. L. Arnold, and S. R. Wessler. 2003. Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. *Genetics* 164: 685-697.

Key, K. H. L. 1968. The concept of stasipatric speciation. *Syst. Zool.* 17: 14-22.

Kimball, S., P. Wilson, and J. Crowther. 2004. Local ecology and geographic ranges of plants in the Bishop Creek watershed of the eastern Sierra Nevada, California, USA. *J. Biogeog.* 31: 1637-1657.

Knowles, L. L. and W. P. Maddison. 2002. Statistical phylogeography. *Mol. Ecol.* 11: 2623-2635.

Kumar, A. and J. L. Bennetzen. 1999. Plant retrotransposons. *Annu. Rev. Genet.* 33: 479-532.

Lenz, L. W. 1959. Hybridization and speciation in the Pacific Coast Irises. *Aliso* 4: 1-72.

Levin, D. A. 2000. The origin, expansion, and demise of plant species. New York, Oxford University Press.

Lewontin, R. C. and L. C. Birch. 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution* 20: 315-336.

Lexer, C., M. E. Welch, O. Raymond, and L. H. Rieseberg. 2003. The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evolution* 57: 1989-2000.

Lyon, D. L. 1973. Territorial and feeding activity of Broad-Tailed Hummingbirds (*Selasphorus platycercus*) in *Iris missouriensis*. *The Condor* 75: 346-349.

Ma, J. and J. L. Bennetzen. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Nat. Acad. Sci. USA* 101: 12404-12410.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7: 639-655.

Martin, N. H., A. C. Bouck, and M. L. Arnold. (2005). Loci affecting long-term hybrid survivorship in Louisiana irises: implications for reproductive isolation and introgression. *Evolution* 59: 116-124.

Martin, N. H., A. C. Bouck, and M. L. Arnold. (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* 172: 1-9.

Mathew, B. 1989. The Iris. 2nd ed. Portland, Timber Press.

McCarthy, E. M., M. A. Asmussen, and W. W. Anderson. 1995. A theoretical assessment of recombinational speciation. *Heredity* 74: 502-509.

Metcalf, H. N. 1978. The Longipetalae. In: The World of Irises. B. Warburton and M. Hamblen, eds.. Witchita, KS, American Iris Society: 295-297.

Morgan, M. T. 2001. Transposable element number in mixed mating populations. *Genetical Research* 77: 261-275.

Natural Resources Conservation Service. 2006. Plants Profile: *Iris missouriensis* Nutt. 2006. http://plants.usda.gov/index.html

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Orians, C. M., D. I. Bolnick, B. M. Roche, R. S. Fritz, and T. Floyd. 1999. Water availability alters the relative performance of *Salix sericeae, Salix eriocephala*, and their F1 hybrids. *Can. J. Bot.* 77: 514-522.

Orr, H. A. 1991. A test of Fisher's theory of dominance. *Proc. Nat. Acad. Sci. USA* 88: 11413-11415.

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

Paran, I. and R. W. Michelmore 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* 85: 985-993.

Patterson, T. B. and T. J. Givnish 2003. Geographic cohesion, chromosomal evolution, parallel adaptive radiations, and consequent floral adaptations in *Calochortus* (Calochortaceae): evidence from a cpDNA phylogeny. *New Phytol.* 161: 253-264.

Peakall, R., and P. E. Smouse. 2001. GenAlEx V5: Genetic Analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia http://www.anu.edu.au/BoZo/GenAlEx

Peakall, R. and P. E. Smouse. 2005. GenAlEx V6: Genetic Analysis in Excel. Population genetic softwarefor teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Petit, R. J., R. Bialozyt, S. Brewer, R. Cheddadi, and B. Comps. 2001. From spatial patterns of genetic diversity to postglacial migration processes in forest trees. *In:* Integrating Ecology and Evolution in a Spatial Context: 14th Special Symposium of the British Ecological Society. J. Silvertown and J. Antonovics, eds. Oxford, Blackwell Science: 295-318.

Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial, and nuclear diversity in plant populations. *Mol. Ecol.* 14: 689-701.

Posada, D., K. A. Crandall, and A. R. Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9: 487-488.

Posada, D. and A. R. Templeton. 2005. GeoDis 2.4 Documentation. 2005. http://darwin.uvigo.es

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

Queller, D. C. and K. F. Goodnight. 1989. Estimating relatedness using genetic markers. *Evolution* 43: 258-275.

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

Randolph, L. F., J. Mitra, and I. S. Nelson. 1961. Cytotaxonomic studies of Louisiana irises. *Bot. Gaz.* 123: 125-133.

Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am. J. Bot.* 78: 1218-1237.

Rieseberg, L. H. and J. M. Burke. 2001. The biological reality of species: gene flow, selection, and collective evolution. *Taxon* 50: 47-67.

Rieseberg, L. H., B. Sinervo, C. R. Linder, M. C. Ungerer, and D. M. Arias. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272: 741-745.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, et al. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211-1216.

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

Ritland, K. 1990. Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* 44: 1230-1241.

Ritland, K. 2002. Extensions of models for the estimation of mating systems using n independent loci. *Heredity* 88: 221-228.

Schoen, D. J. and S. C. Stewart. 1986. Variation in male reproductive investment and male reproductive success in white spruce. *Evolution* 40: 1109-1120.

Schweitzer, J., G. D. Martinsen, and T.G. Whitham. 2002. Cottonwood hybrids gain fitness traits of both parents: a mechanism for their long-term persistence? *Am. J. Bot.* 89: 981-990.

Simonet, M. 1934. The Genus *Iris*: Cytological and Genetic Research. Doctoral thesis. Warburton, B., trans. Westborough MA, MIS Press.

Small, J. K. and E. J. Alexander. 1931. Botanical interpretation of the iridaceous plants of the Gulf States. *Contrib. New York Bot. Gard.* 327: 325-358.

Smouse, P. E. 1998. To tree or not to tree. *Mol. Ecol.* 7: 399-412.

Smouse, P. E. and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82: 561-573.

Species Group of the British Iris Society. 1997. A Guide to Species Irises: their Identification and Cultivation. New York, Cambridge University Press.

Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Syst. Evol.* 206: 353-373.

Steele, K. P. and R. Vilgalys. 1994. Phylogenetic analysis of *Polemoniaceae* using nucleotide sequences of the plastid gene *matK*. *Syst. Bot.* 19: 126-142.

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

Szymura, J. M., and N. H. Barton. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* 45: 237-261.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17: 1105-1109.

Templeton, A. R. 1981. Mechanisms of speciation - a population genetic approach. *Annu. Rev. Ecol. and Syst.* 12: 23-48.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7: 381-397.

Thompson, R. S. and K. H. Anderson. 2000. Biomes of western North America at 18,000, 6000 and 0 14C yr BP reconstructed from pollen and packrat midden data. *J. Biogeog.* 27: 555-584.

University of California Jepson Herbarium. 2004. SMASCH Project: Specimen Management System for California Herbaria. http://www.mip.berkeley.edu/www_apps/smasch/

van den Broeck, D., T. Maes, M. Sauer, et al. 1998. Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* 13: 121-129.

Vicient, C. M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, and A. Beharav. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784.

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

Vitte, C. and O. Panaud. 2003. Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in Rice *Oryza sativa* L. *Mol. Biol. Evol.* 20: 528-540.

Vos, P., R. Hogers, et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nuc. Acids Res.* 23: 4407-4414.

Vuylsteke, M., R. Mank, et al. 1999. Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. Appl. Genet.* 99: 921-935.

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

Weir, B. S. 1979. Inferences about linkage disequilibrium. Biometrics 35: 235-254.

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

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

Wesselingh, R. A. and M. L. Arnold. 2003. A top-down hierarchy in fruit set on inflorescences in *Iris fulva* (Iridaceae). *Plant Biol.* 5: 651-660.

Williams, J. G. K., A. R. Kubelik, and D. Dunkerley. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nuc. Acids Res.* 18: 6531-6535.

Williams, M., J. Chappell, K. J. Livak, and J. A. Rafalski. 1998. Quaternary Environments. New York, Oxford University Press.

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

Wilson, C. A. 2004. Phylogeny of *Iris* based on chloroplast *matK* gene and *trnK* intron sequence data. *Mol. Phyl. Evol.* 33: 402-412.

Wolfe, A. D., Q. Y. Xiang, and S. R. Kephart. 1998. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proc. Natl. Acad. Sci. USA* 95: 5112-5115.

Xia, X. and Z. Xie. 2001. DAMBE: Data analysis in molecular biology and evolution. *J. Heredity* 92: 371-373.

Young, N. D. 1996. Concordance and discordance: a tale of two hybrid zones in the Pacific Coast irises (Iridaceae). *Am. J. Bot.* 83: 1623-1629.

Zhang, D.-X. and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol. Ecol.* 12: 563-584.

Zink, R. A. and N. T. Wheelwright. 1997. Facultative self-pollination in Island Irises. *Am. Mid. Nat.* 137: 72-78.