



This is to certify that the
dissertation entitled

SPECIATION IN THE WESTERN NORTH AMERICAN
WILDFLOWER GENUS *MIMULUS*

presented by

JAMES MICHAEL SOBEL

has been accepted towards fulfillment
of the requirements for the

PhD	degree in	Plant Biology
		Ecology, Evolutionary Biology
		and Behavior


Major Professor's Signature

9/1/10
Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

SPECIATION IN THE WESTERN NORTH AMERICAN WILDFLOWER GENUS
MIMULUS

By

James Michael Sobel

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Plant Biology
Ecology, Evolutionary Biology and Behavior

2010

ABSTRACT

SPECIATION IN THE WESTERN NORTH AMERICAN WILDFLOWER GENUS *MIMULUS*

By

James Michael Sobel

A central task of evolutionary biologists is to identify the forms of reproductive isolation that are most important in speciation. The presented study approaches this issue by examining the strength of multiple forms of reproductive isolation across a group of recently diverged species pairs spanning the western North American wildflower genus, *Mimulus*. Chapter 1, Unification of Methods for Estimating the Strength of Individual Reproductive Barriers, reviews methods that have been used by previous authors for calculating the strength of individual reproductive barriers and presents a simple linear explanation for unifying the disparate approaches. Chapter 2, Ecogeographic Isolation Among Recently Diverged Species Pairs in the Genus *Mimulus*, presents a method for separating the intrinsic biological effects of differences in geographic range for estimating the strength of ecogeographic isolation using an ecological niche modeling approach. This work examines the strength of ecogeographic isolation for 12 recently diverged species pairs in the genus, and shows that this barrier is commonly very strong, with an average strength of 0.64 to 0.69 depending on which threshold of habitat suitability is employed. Because isolating barriers act sequentially throughout the life history of organisms, this first-to-act barrier is therefore responsible for the majority of isolation experienced by recently diverged species. Chapter 3, entitled The Evolution of Reproductive Isolation Across the Genus *Mimulus*, provides estimates of the strength of other forms of reproductive isolation that can act upon the gene flow remaining after the

effect of ecogeographic isolation. Additional reproductive barriers examined include temporal isolation, isolation due to relative seed set, intrinsic postzygotic isolation due to relative hybrid viability, and relative hybrid fertility. Individually, intrinsic postzygotic isolation due to inviability is highly variable but relatively strong. Temporal isolation, seed set isolation showed moderate strength, and relative hybrid fertility was relatively weak. When viewed in light of the linear sequential stages at which reproductive barriers act in nature, ecogeographic isolation dominates the relative contribution of barriers to total isolation, while other forms including postzygotic isolation due to hybrid inviability prevents only a limited amount of gene flow. Several species pairs show evidence of incomplete total reproductive isolation, and it is unclear whether this represents cases of incomplete speciation or the effect of unmeasured barriers such as pollinator isolation or extrinsic postzygotic isolation. In Chapter 4, Contrasting Patterns of Introgression in Two Pairs of *Mimulus* Species, an introgression analysis of neutral molecular sequences is presented to corroborate the measured values of isolation for two species pairs representing relatively low and high reproductive isolation estimates. The species pair with relatively high reproductive isolation measured in the lab (*M. constrictus* and *M. whitneyi*) shows strong evidence of molecular sequence divergence, while the species pair with lower estimates of isolation (*M. bicolor* and *M. filicaulis*) shows evidence consistent with rampant gene exchange. Taken together, these analyses suggest that measurements of isolation estimated in the lab accurately represent effects on gene flow in nature, and ecogeographic isolation is of primary importance in the origin of *Mimulus* species.

ACKNOWLEDGMENTS

This dissertation could not have been completed without the support of many people and programs. I would like to thank the members of my guidance committee, Doug Schemske, Jeff Conner, Kay Gross, Rich Lenski, and Alan Prather, for providing advice through all stages of this project, reading drafts of chapters, and providing helpful comments. I would like to especially thank Jeff Conner and Doug Schemske. Jeff provided encouragement and mentorship both before and during graduate school, and has been a great influence on me. Doug has provided immeasurable amounts of encouragement and guidance through all stages of my dissertation. He has taught me a tremendous amount, and I truly could not have asked for a better mentor. Other professors at MSU have also been very supportive, especially Gary Mittelbach, Pete Murphy, and Barb Sears.

Interactions with many colleagues have helped shape my thoughts on this project. Previous graduate students in the Schemske lab have improved the quality of this work tremendously through helpful discussions and criticisms including, Kathleen Kay, Amy Angert, Grace Chen, Jeff Evans, Lorna Watt, and Emily Dittmar. In particular, my interaction with Grace on many aspects of the work presented in Chapter 1 has been a very fruitful and enjoyable collaboration. Fellow graduate students in other labs have also helped shape ideas and provided encouragement including, Orlando Alvarez-Fuentes, Frances Knapczyk, Chris Gaukler, Mike Grillo, Emily Grman, Chris Hamm, Kane Keller, Lauren Kinsman, Jason Kilgore, Melissa Kjelvik, Raffica LaRosa, Jason Martina, Todd Robinson, Angela Roles, Anne Royer, and Heather Sahli. I am also grateful for the

insight I gained through discussions with other researchers utilizing *Mimulus* as a model organism including Toby Bradshaw, Andrea Case, Arielle Cooley, Lila Fishman, Nick Griffin, Dena Grossenbacher, Chris Ivey, David Lowry, Noland Martin, Jason Sexton, Matt Streisfeld, Andrea Sweigart, John Willis, and Carrie Wu.

I am indebted to Schemske lab technician, Mark Hammond for all of his assistance. His advice, support, and comradery have been greatly valued throughout this project. Much of the work presented would not have been possible without the assistance of many undergrads, including Christina Cooper, Emily Hall, Anna Kilbourn, Christine Leduc, Liana Nichols, Lindsay Petroff, Joshua Rilko, Maria Roberts, and Julia Smith. It is very difficult to single out especially valuable members of this list because so many of them have contributed incredible amounts of time and energy to the project. Interacting with them has been one of the true joys of my graduate school experience. I would also like to again thank my advisor, Doug Schemske, for providing the financial means to hire this incredible team of people.

Several people provided valuable assistance in establishing the field component of this work. In particular, Steve Schoenig contributed much time and information that helped get me started in finding species of *Mimulus* in the mountains of California, and Joe Callizo was also incredibly generous with his time and knowledge of the plants of Napa and Lake counties. I spent time in the field with several other excellent botanists, including Paul Beardsley, Paul Aigner, and Cathy Koehler, and their advice and information helped considerably. In addition, I would like to thank the many National Forest and National Park Service botanists who assisted with permits and collection information.

Funding for this work is gratefully acknowledged. The National Science Foundation provided support through both a 1-½ year GK-12 fellowship and a 2-year NCEAS graduate fellowship. In addition, NSF provided research support in the form of a FIBR grant awarded to John Willis (Doug Schemske co-PI; DBI-0328636) and a Doctoral Dissertation Improvement Grant (Doug Schemske & J. Sobel, co-PI's; DEB-0808447). In addition, several units within Michigan State University provided funds either for travel or stipends, including the Department of Plant Biology, Ecology, Evolutionary Biology & Behavior Program, Kellogg Biological Station, College of Natural Science, and the Graduate School. In particular, the Plant Biology Department's Paul Taylor Travel Fund made much of the travel to field sites and meetings possible.

Many former teachers and professors were inspirational in guiding me towards a life in science, and I would like to take this opportunity to formally thank them for their efforts. Dale Rosene and Gloria Wheeler were instrumental in introducing me to science in 8th and 9th grades, and their passion for their subjects first sparked my interest in science in general and biology in particular. During my undergraduate work, several professors and scientists also served to inspire continued work in the subject, including Steve Bertman, Jeff Conner, Woody Ehrle, David Karowe, and Steve Roberds.

Finally, I am deeply grateful to the love and support from my family during this project. My parents, Mike and Cindy Sobel, and grandparents, Andrew and Kathleen Sobel & Jim and Judy Huggett, have always encouraged and supported my education, for which I am eternally thankful. Most of all, I want to thank my wife, Darcy Sobel, and son, Luke Sobel, for their love, encouragement, and patience throughout this project.

TABLE OF CONTENTS

LIST OF TABLES.....	ix
---------------------	----

LIST OF FIGURES.....	xi
----------------------	----

CHAPTER 1

UNIFICATION OF METHODS FOR ESTIMATING THE STRENGTH OF INDIVIDUAL REPRODUCTIVE BARRIERS.....

Introduction.....	1
Currently Used Methods.....	3
A Simple Linear Solution.....	5
Accommodation of Unequal Null Expectations.....	10
Reproductive Barriers that Affect Co-Occurrence.....	15
Summary.....	17
Tables.....	21
Figures.....	23
Literature Cited.....	24
	29

CHAPTER 2

ECOGEOGRAPHIC ISOLATION AMONG RECENTLY DIVERGED SPECIES PAIRS IN THE GENUS *MIMULUS*.....

Introduction.....	32
Materials & Methods.....	34
Results.....	38
Discussion.....	48
Summary.....	52
Tables.....	57
Figures.....	59
Literature Cited.....	64
	131

CHAPTER 3

THE EVOLUTION OF REPRODUCTIVE ISOLATION ACROSS THE GENUS *MIMULUS*.....

Introduction.....	137
Materials & Methods.....	140
Results.....	145
Discussion.....	155
Summary.....	161
Tables.....	171
Figures.....	172
Literature Cited.....	176
	192

CHAPTER 4
CONTRASTING PATTERNS OF INTROGRESSION IN TWO PAIRS OF *MIMULUS*
SPECIES.....197

 Introduction.....200

 Materials & Methods.....203

 Results.....209

 Discussion.....212

 Summary.....215

 Tables.....216

 Figures.....225

 Literature Cited.....227

LIST OF TABLES

Table 1.1. Alternative methods for calculating reproductive isolation indexes.....	23
Table 2.1. Summary of <i>Mimulus</i> species pairs used and methods employed for ecological niche modeling.....	59
Table 2.2. Environmental layers used in ecological niche modeling.....	60
Table 2.3. Correlation coefficients among environmental layers used in ecological niche modeling.....	61
Table 2.4. Results of one-way MANOVA analysis on environmental variables between <i>Mimulus</i> species pairs.....	62
Table 2.5. Ecogeographic reproductive isolation between species pairs in <i>Mimulus</i>	63
Table 3.1. <i>Mimulus</i> species pairs used in comparative study of reproductive isolation.....	172
Table 3.2. Estimates of the strength of individual reproductive barriers in pairs of species in <i>Mimulus</i>	173
Table 3.3. Relative parent and F1 viability in <i>Mimulus</i> species pairs.....	174
Table 3.4. Comparison of individual barrier strength and relative contributions to total reproductive isolation in recently diverged species of <i>Mimulus</i>	175
Table 4.1A. Molecular markers sequenced for analysis of introgression within <i>Mimulus bicolor</i> / <i>M. filicaulis</i> species pair.....	216
Table 4.1B. Molecular markers sequenced for analysis of introgression within <i>Mimulus constrictus</i> / <i>M. whitneyi</i> species pair.....	217
Table 4.2. Primer sequences used to amplify <i>MgSTS</i> markers in this study.....	218
Table 4.3A. Results of polymorphism analysis using SITES software for <i>Mimulus bicolor</i> / <i>M. filicaulis</i> species pair.....	219
Table 4.3B. Results of polymorphism analysis using SITES software for <i>Mimulus constrictus</i> / <i>M. whitneyi</i> species pair.....	220
Table 4.4A. Pairwise F_{ST} values from molecular polymorphism analysis performed in SITES for <i>Mimulus bicolor</i> / <i>M. filicaulis</i> species pair.....	221

Table 4.4B. Pairwise F_{ST} values from molecular polymorphism analysis performed in
SITES for *Mimulus constrictus* / *M. whitneyi* species pair.....222

Table 4.5A. Estimates of Tajima’s D and the mutation rate parameter θ ($4N\mu$) per base
pair for *Mimulus bicolor* / *M. filicaulis* species pair.....223

Table 4.5B. Estimates of Tajima’s D and the mutation rate parameter θ ($4N\mu$) per base
pair for *Mimulus constrictus* / *M. whitneyi* species pair.....224

LIST OF FIGURES

Figure 1.1. Relationship between the probability of heterospecific gene flow and the measure of reproductive isolation obtained through three commonly used methods.....	24
Figure 1.2. Diagram showing similarities between pre- and postzygotic isolation index measures of equal magnitude.....	25
Figure 1.3. The effect of changes in expected frequency of hybrid formation on the value of reproductive isolation using equation RI_6	26
Figure 1.4. Hypothetical data showing the relationship between three methods for calculating the strength of temporal isolation.....	28
Figure 2.1. Phylogenetic relationships among <i>Mimulus</i> species in this study.....	64
Figure 2.2A. Ecological niche model of <i>Mimulus androsaceus</i>	65
Figure 2.2B. Ecological niche model of <i>Mimulus shevockii</i>	66
Figure 2.2C. Ecological niche model of <i>Mimulus angustatus</i>	67
Figure 2.2D. Ecological niche model of <i>Mimulus pulchellus</i>	68
Figure 2.2E. Ecological niche model of <i>Mimulus bicolor</i>	69
Figure 2.2F. Ecological niche model of <i>Mimulus filicaulis</i>	70
Figure 2.2G. Ecological niche model of <i>Mimulus bigelovii</i>	71
Figure 2.2H. Ecological niche model of <i>Mimulus bolanderi</i>	72
Figure 2.2I. Ecological niche model of <i>Mimulus brevipes</i>	73
Figure 2.2J. Ecological niche model of <i>Mimulus johnstonii</i>	74
Figure 2.2K. Ecological niche model of <i>Mimulus cardinalis</i>	75
Figure 2.2L. Ecological niche model of <i>Mimulus lewisii</i>	76
Figure 2.2M. Ecological niche model of <i>Mimulus constrictus</i>	77
Figure 2.2N. Ecological niche model of <i>Mimulus whitneyi</i>	78
Figure 2.2O. Ecological niche model of <i>Mimulus cusickii</i>	79

Figure 2.2P. Ecological niche model of <i>Mimulus nanus</i>	80
Figure 2.2Q. Ecological niche model of <i>Mimulus douglasii</i>	81
Figure 2.2R. Ecological niche model of <i>Mimulus kelloggii</i>	82
Figure 2.2S. Ecological niche model of <i>Mimulus floribundus</i>	83
Figure 2.2T. Ecological niche model of <i>Mimulus norrisii</i>	84
Figure 2.2U. Ecological niche model of <i>Mimulus gracilipes</i>	85
Figure 2.2V. Ecological niche model of <i>Mimulus palmeri</i>	86
Figure 2.2W. Ecological niche model of <i>Mimulus parryi</i>	87
Figure 2.2X. Ecological niche model of <i>Mimulus rupicola</i>	88
Figure 2.3A. Variation in environmental variables between the species pair <i>Mimulus androsaceus</i> and <i>M. shevockii</i>	90
Figure 2.3B. Variation in environmental variables between the species pair <i>Mimulus angustatus</i> and <i>M. pulchellus</i>	92
Figure 2.3C. Variation in environmental variables between the species pair <i>Mimulus bicolor</i> and <i>M. filicaulis</i>	94
Figure 2.3D. Variation in environmental variables between the species pair <i>Mimulus bigelovii</i> and <i>M. bolanderi</i>	96
Figure 2.3E. Variation in environmental variables between the species pair <i>Mimulus brevipes</i> and <i>M. johnstonii</i>	98
Figure 2.3F. Variation in environmental variables between the species pair <i>Mimulus cardinalis</i> and <i>M. lewisii</i>	100
Figure 2.3G. Variation in environmental variables between the species pair <i>Mimulus constrictus</i> and <i>M. whitneyi</i>	102
Figure 2.3H. Variation in environmental variables between the species pair <i>Mimulus cusickii</i> and <i>M. nanus</i>	104
Figure 2.3I. Variation in environmental variables between the species pair <i>Mimulus douglasii</i> and <i>M. kelloggii</i>	106

Figure 2.3J. Variation in environmental variables between the species pair <i>Mimulus floribundus</i> and <i>M. norrisii</i>	108
Figure 2.3K. Variation in environmental variables between the species pair <i>Mimulus gracilipes</i> and <i>M. palmeri</i>	110
Figure 2.3L. Variation in environmental variables between the species pair <i>Mimulus parryi</i> and <i>M. rupicola</i>	112
Figure 2.4A Identity tests using Schoener's D statistic.....	114
Figure 2.4B Identity tests using I statistic.....	116
Figure 2.5A. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus androsaceus</i> and <i>M. shevockii</i>	117
Figure 2.5B. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus angustatus</i> and <i>M. pulchellus</i>	118
Figure 2.5C. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus bicolor</i> and <i>M. filicaulis</i>	119
Figure 2.5D. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus bigelovii</i> and <i>M. bolanderi</i>	120
Figure 2.5E. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus brevipes</i> and <i>M. johnstonii</i>	121
Figure 2.5F. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus cardinalis</i> and <i>M. lewisii</i>	122
Figure 2.5G. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus constrictus</i> and <i>M. whitneyi</i>	123
Figure 2.5H. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus cusickii</i> and <i>M. nanus</i>	124
Figure 2.5I. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus douglasii</i> and <i>M. kelloggii</i>	125
Figure 2.5J. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus floribundus</i> and <i>M. norrisii</i>	126
Figure 2.5K. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus gracilipes</i> and <i>M. palmeri</i>	127

Figure 2.5L. Overlay plot showing the amount of ecogeographic overlap in the species pair <i>Mimulus parryi</i> and <i>M. rupicola</i>	128
Figure 2.6. Estimates of ecogeographic isolation in <i>Mimulus</i> obtained using threshold 2 regressed on sequence divergence between the species pairs.....	129
Figure 2.7. Relationship between niche similarity and ecogeographic isolation.....	130
Figure 3.1. Recently diverged pairs of species in the genus <i>Mimulus</i> used in this study.....	176
Figure 3.2A-D. Differences in flowering time based on herbarium collection for the species pairs <i>Mimulus angustatus</i> / <i>pulchellus</i> , <i>M. bicolor</i> / <i>filicaulis</i> , <i>M. brevipes</i> / <i>johnstonii</i> , <i>M. constrictus</i> / <i>whitneyi</i>	177
Figure 3.2E-H. Differences in flowering time based on herbarium collection for the species pairs <i>Mimulus cusickii</i> / <i>nanus</i> , <i>M. douglasii</i> / <i>kelloggii</i> , <i>M. floribundus</i> / <i>norrisii</i> , <i>M. gracilipes</i> / <i>palmeri</i>	178
Figure 3.3A-I. Comparisons of flowering time differences between <i>Mimulus</i> species under common garden conditions in the greenhouse.....	179
Figure 3.4A-C. Postmating reproductive isolation in the species pairs <i>Mimulus angustatus</i> / <i>pulchellus</i> , <i>M. bicolor</i> / <i>filicaulis</i> , and <i>M. brevipes</i> / <i>johnstonii</i>	180
Figure 3.4D-F. Postmating reproductive isolation in the species pairs <i>Mimulus constrictus</i> / <i>whitneyi</i> , <i>M. cusickii</i> / <i>nanus</i> , and <i>M. douglasii</i> / <i>kelloggii</i>	181
Figure 3.4G-I. Postmating reproductive isolation in the species pairs <i>Mimulus floribundus</i> / <i>norrisii</i> , <i>M. gracilipes</i> / <i>palmeri</i> , and <i>M. jungermannioides</i> / <i>washingtonensis</i>	182
Figure 3.5A-B. Relative pollen fertility measures in species pairs <i>Mimulus angustatus</i> / <i>pulchellus</i> and <i>M. bicolor</i> / <i>filicaulis</i>	183
Figure 3.5C-E. Relative pollen fertility measures in species pairs <i>Mimulus douglasii</i> / <i>kelloggii</i> and individual species <i>M. constrictus</i> and <i>M. floribundus</i>	184
Figure 3.5F-G. Relative pollen fertility measures in species pairs <i>Mimulus gracilipes</i> / <i>palmeri</i> and <i>M. jungermannioides</i> / <i>washingtonensis</i>	185
Figure 3.6. Individual estimates of reproductive isolation across the genus <i>Mimulus</i>	186
Figure 3.7A. Individual reproductive barrier strengths in species pairs of <i>Mimulus</i>	188

Figure 3.7B. Relative contribution of reproductive barriers to total reproductive isolation in species pairs of <i>Mimulus</i>	190
Figure 3.8. Relationship between total reproductive isolation of species pairs in <i>Mimulus</i> and genetic distance.....	191
Figure 4.1A. Distribution of populations sampled for molecular analysis in <i>Mimulus</i> <i>bicolor</i> and <i>M. filicaulis</i>	225
Figure 4.1B. Distribution of populations sampled for molecular analysis in <i>Mimulus</i> <i>constrictus</i> and <i>M. whitneyi</i>	226

**CHAPTER 1: UNIFICATION OF METHODS FOR ESTIMATING THE STRENGTH
OF INDIVIDUAL REPRODUCTIVE BARRIERS**

CHAPTER 1: UNIFICATION OF METHODS FOR ESTIMATING THE STRENGTH OF INDIVIDUAL REPRODUCTIVE BARRIERS

Under the biological species concept, understanding the evolution of reproductive isolation is tantamount to describing the origin of species. A major goal of speciation research is to identify which of the many forms of reproductive isolation contribute most to total isolation at the time of speciation. In order to achieve this goal, the strength of multiple forms of isolation must be compared in an equivalent manner. Despite this necessity, there exists a wide diversity of methods that have been employed to estimate isolation, falling into several mathematically non-equivalent categories. This has resulted in comparisons between forms of isolation and between taxa in which the measure of isolation is not directly comparable. In this study, a simple linear formulation for the isolation index with the most biological support is given. This method for estimating isolation provides three distinct advantages: 1) it is directly related to gene flow between population, 2) it is symmetrical about the origin, such that measures of disassortative mating and heterosis are comparable to measures of isolation in the positive range, and 3) it is equivalent between broad categories of reproductive isolation. This linear formulation can be adopted in cases where expected amounts of con- and heterospecific gene flow differ under a null model and can be adjusted for use in all forms of isolation.

Introduction

Adherents of the biological species concept (Mayr 1942) have developed a research program into the origin of species based upon understanding the evolution of barriers to reproduction (Coyne and Orr 2004). Within this framework, the process of speciation involves the accumulation of barriers to gene exchange leading to complete reproductive isolation (Coyne and Orr 2004). It has been long recognized (Dobzhansky 1937; Mayr 1942; Poulton 1908) that reproductive barriers can take many forms from ecological isolation to intrinsic hybrid inviability. These barriers are often separated into those that operate before the formation of a hybrid (prezygotic) and those that act once a hybrid has formed (postzygotic), and there is longstanding debate about which of these broad categories of isolating barriers are most important in speciation (Schemske 2000; Coyne and Orr 2004; Rice and Hostert 1993).

Early work on the nature of reproductive barriers typically consisted of gathering data that showed that barriers existed (e.g. Dobzhansky 1938) without necessarily estimating a metric that related the barrier strength to an amount of gene flow reduced by it. However, a central task of speciation studies is to identify which forms of reproductive isolation are the most important in the process (Coyne and Orr 2004); therefore, some method for comparison is clearly necessary. In their highly influential study in *Drosophila*, Coyne and Orr (1989; 1997) used data from the literature on 171 interspecific hybridization attempts, and developed a metric that related the relative frequency of con- and heterospecific matings to reproductive isolation. This work provided evidence that the evolution of prezygotic isolation outpaces postzygotic due to differences in sympatric taxa and that Haldane's rule is a common feature of postzygotic

isolation evolution. A surge of interest in calculating the strength of reproductive isolation resulted (e.g. Bolnick and Near 2005; Coyne and Orr 1989; 1997; Dopman et al. 2010; Matsubayashi and Katakura 2009; Mendelson 2003; Moyle et al. 2004; Presgraves 2002; Price and Bouvier 2002; Ramsey et al. 2003; Sasa et al. 1998), and most authors adopted the mathematical formulation provided by Coyne and Orr (1989).

However, Martin and Willis (2007) recently challenged whether the isolation coefficients estimated by these methods are truly representative of the declines in gene flow experienced by species. Among other concerns, they raise issues about whether pre- and postzygotic isolation are being calculated equivalently and propose methods for accounting for different null expectations of gene flow due to unequal population or gamete abundances. While these issues have been appreciated by some (e.g. Matsubayashi and Katakura 2009), others have failed to adopt the proposed methods (e.g. Dopman et al. 2010), creating substantial confusion over the correct methods for analyzing isolation.

Given the importance of these measurements, the lack of consensus on the methods used to calculate indexes of isolation is problematic. As a result, isolation indexes are used inconsistently, even within a single study. Clearly, identifying the most appropriate indexes for each form of isolation would be of great benefit, both when assessing the relative strength of individual barriers in single species pair studies and when assessing the strength of isolation among groups in a comparative context. Our goal in this paper is to review the most common methods used by previous authors, present a simplified view of the alternative with the most mathematical and biological justification, and provide examples to illustrate how to utilize this method for both ideal and realistic

datasets. Ultimately, we hope this unification of methods will allow researchers the opportunity to compare measures of reproductive isolation among disparate forms of isolation and/or taxa.

Currently Used Methods

The purpose of calculating the strength of a reproductive isolating barrier is to estimate how much gene flow is (or is potentially) reduced by a barrier (Coyne and Orr 2004). In most cases, previous workers have used equations to describe reproductive isolation that range from 0 when there is no isolation to 1 when there is complete isolation (Table 1.1). However, as long as a metric satisfies this one requirement, the relationship between the strength of a barrier and the amount of isolation attributed to it has been essentially unexplored. Because pre- and postzygotic isolation have traditionally been approached in non-equivalent ways (Martin and Willis 2007), we will begin by reviewing these separately.

Prezygotic Isolation

Metrics of prezygotic isolation attempt to predict how a barrier will affect the probability of heterospecific zygote formation. Sexual isolation due to mating preferences has been widely studied as a potential agent of isolation (Coyne et al. 2005; Coyne and Orr 1997; Ehrman 1965; Matsubayashi and Katakura 2009; Mendelson 2003; Tilley et al. 1990). One equation used to describe the relationship between mating preferences and isolation is presented in Coyne and Orr (1989) as:

$$RI = 1 - \frac{\text{frequency of heterospecific matings}}{\text{frequency of conspecific matings}},$$

which we will henceforward simplify as:

$$RI_I = 1 - \frac{H}{C}, \quad (1)$$

where H represents the frequency of heterospecific matings and C represents the frequency of conspecific matings. This equation (or minor deviations from it) is the most commonly used approach to estimating the contribution of mating preferences to reproductive isolation (Table 1.1). The metric indeed results in an isolation index that equals 1 when sexual preferences insure no heterospecific mating and zero when there are no preferences. However, the relationship is not linear between 0 and 1, and ranges to $-\infty$ in cases of disassortative mating (Figure 1.1A). As an example, in a choice mating trial, if species X females mate heterospecifically 25% of the time and conspecifically 75% of the time, there is clearly a preference for within species matings, and equation 1 would result in an isolation index of 0.67. Alternatively, if females of species X actually prefer males of species O resulting in the reversal of the above frequencies (25% conspecific mating and 75% heterospecific mating), then equation 1 would give an isolation index of -2. Given that the departure from random mating preferences are equivalent, the effect on gene flow would be identical, but in opposite directions. Obviously, this isolation index does not capture this symmetry, and the resulting indexes are difficult to ascribe to the probability of gene flow. This creates situations in which instances of disassortative mating have to be removed from consideration (or are replaced with a zero) because they are not directly comparable to measures within the normal

bounds. While disassortative mating is relatively rare in interspecific studies (Coyne & Orr 1989 note at least 5 examples in *Drosophila*), a method that can accommodate its effects on gene flow is warranted for situations where it may be more common, such as studies of pre-speciation population divergence.

Another equation used for estimating the strength of prezygotic isolation appears in Ramsey et al. (2003) for calculating the strength of reproductive isolation due to pollinator preferences on monkeyflowers *Mimulus lewisii* and *M. cardinalis*. They present an equation using the relative frequency of cross-species foraging bouts to all visits to estimate isolation:

$$RI = 1 - \frac{\text{number of cross - species foraging bouts}}{\text{total number of foraging bouts}},$$

which we will henceforward simplify into:

$$RI_2 = 1 - \frac{H}{C + H}. \quad (2)$$

RI_2 can range from 0 (all heterospecific gene flow) to 0.5 (random mating) to 1 (no heterospecific gene flow) (Figure 1.1B). The biological interpretation of RI_2 may be very different from the same RI_1 value, presenting some problems. For instance, when $RI_1 = 0.5$, heterospecific mating is half of conspecific mating, while an $RI_2 = 0.5$ results from hetero- and conspecific mating being equivalent. Clearly, a situation that doesn't deviate from a null expectation under random mating is most appropriately considered to have no isolation.

Yet another form of the equation used to calculate prezygotic isolation is presented in Mendelson's (2003) study of reproductive isolation across fish in the genus *Etheostoma*. Sexual isolation was calculated as:

$$RI = \frac{\# \text{ conspecific spawning events} - \# \text{ heterospecific spawning events}}{\text{total} \# \text{ spawning events}}.$$

This equation was referenced from Stalker's (1942) study of sexual isolation in the *Drosophila virilis* complex where the equation was presented as:

$$RI = \frac{\% \text{ conspecific females inseminated} - \% \text{ alien females inseminated}}{\% \text{ conspecific females inseminated} + \% \text{ alien females inseminated}}.$$

We will simplify this expression into:

$$RI_3 = \frac{C - H}{C + H}. \quad (3)$$

This index ranges from -1 (all heterospecific mating) to 0 (random mating) to 1 (no heterospecific mating), which is mathematically easy to understand; yet the biological meaning of the ratio between the numerator and the denominator is not made clear. The symmetrical nature of this form of the equation presents significant benefits, and we will return to its derivation in the following section.

Postzygotic Isolation

Isolation indexes of postzygotic isolation estimate the amount of gene flow reduced by the inviability or sterility of hybrids (either through intrinsic or extrinsic means). As pointed out by Martin & Willis (2007), one of the most serious difficulties in comparing pre- and postzygotic isolation within studies is that the two isolation indexes are commonly non-equivalent. For example, Coyne & Orr (1989) counted the number of instances of complete sterility or complete inviability in a reciprocal cross and divided by

4 to make a metric that could only be 0.25, 0.5, 0.75, or 1. Others have attempted to calculate postzygotic isolation in a more quantitative way, with equations that are algebraically equivalent or similar to one of the three presented thus far. For example, many studies (Table 1.1) calculate postzygotic isolation as:

$$RI = 1 - \frac{\text{average fitness of offspring from heterospecific cross}}{\text{average fitness of offspring from conspecific cross}},$$

which is mathematically equivalent to RI_I .

Using this form of the equation presents difficulties when considering cases in which hybrids outperform the parents. Heterosis is relatively common among recently diverged species (e.g. Taylor et al. 2009), but the value that RI_I gives in these instances is not proportional to the amount of gene flow that may be facilitated by the phenomenon. For example, if a plant of species *X* makes twice as many seeds when mated interspecifically to species *O*, the isolation calculated by RI_I would be -1. However, if the seed sets are reversed (intraspecific matings produce twice as many seeds), RI_I would yield an isolation coefficient of 0.5. This is analogous to the problem discussed above for cases of disassortative mating in prezygotic measures, and can present a considerable challenge when making comparisons of the isolation strength or combining multiple forms of isolation into composite metrics (as in Lowry et al. 2008).

Additional studies use equations that can be equivalent to RI_3 under certain circumstances. For example, Palmer and Feldman (2009) present this equation:

$$RI = 1 - \frac{2 \times \text{average fitness of hybrid offspring}}{\text{sum of average fitnesses from two allopatric populations}}.$$

While this form of the equation is the same as the simple linear solution discussed below, the authors did not provide a justification for the equation used, and it is not made clear what is gained by using this form of the equation.

The variety of isolation indexes available makes it difficult to compare the strength of isolation among disparate taxa. Sufficient data on reproductive isolation exist to allow researchers to approach higher order questions about the processes that increase the rate of reproductive isolation evolution (e.g. Funk et al. 2006; Lowry et al. 2008). However, if different forms of reproductive isolation indexes are used for different barriers or taxa, it is not possible to combine the data into global analyses. Within individual studies that measure more than one reproductive barrier, the goal is often to test hypotheses concerning which form of isolation evolves at a faster rate (e.g. Coyne and Orr 1989; 1997). The relative rate of isolation evolution among barriers is an actively debated subject in evolutionary biology; however, it is impossible to legitimately compare different forms of isolation if they are calculated by different measures. Clearly, a unification of methods would benefit those trying to make such comparisons.

A Simple Linear Solution

As described above, most would agree that metrics of reproductive isolation should be scaled to 0 when no isolation exists, and 1 when isolation is complete, but the manner in which this is achieved can greatly affect the resulting interpretation of the value obtained. The most straightforward way to achieve this goal is to use a simple linear equation that describes the relationship between the probability of gene flow past a specific barrier and the reproductive isolation index with the following conditions: 1) if

the probability of interspecific gene flow is 0, reproductive isolation is 1, II) if the probability of gene flow is 0.5 (e.g. random mating), reproductive isolation is 0, III) if the probability of gene flow is 1 (e.g. complete disassortative mating), reproductive isolation is -1. These conditions are satisfied by a line with y-intercept of 1 and slope of -2 (Figure 1.1C), producing the linear equation:

$$y = -2x + 1 \quad (4)$$

where y is the reproductive isolation index and x is the probability of heterospecific gene flow. The probability of heterospecific gene flow can be estimated in several ways, but the simplest approach is to calculate the relative proportion of heterospecific to total gene flow with the following term: $\frac{H}{C + H}$. Substituting $\frac{H}{C + H}$ for x gives our simplest proposed method for calculating reproductive isolation as:

$$RI_5 = 1 - 2 * \left(\frac{H}{C + H} \right). \quad (5)$$

As alluded to above, this equation is algebraically equivalent to equation 3 under certain circumstances, and has been used in different forms in several previous treatments of isolation (e.g. Mendelson 2003; Palmer and Feldman 2009, Table 1.1). However, the previous uses of this metric did not describe the purpose of the reformulation, and failed to demonstrate its advantages over previous versions.

Consider a case of prezygotic isolation in which a mating study showed that females of one species mated conspecifically 90% of the time and heterospecifically 10%. Equation RI_1 gives a reproductive isolation index for those females of 0.89 because conspecific matings represent 8/9 that of heterospecific. Equation RI_2 gives a

reproductive isolation index of 0.9 because conspecific matings represent 90% of overall matings. However, assuming both species are equally abundant in a mating trial, the null expectation is that random mating would produce a 50:50 ratio of heterospecific to conspecific mating. Using our proposed RI_5 equation, the isolation index is instead 0.8, which represents the average of the proportional increase in conspecific mating and decrease in heterospecific mating compared to this random expectation (a conspecific mating frequency of 0.9 is an 80% increase over the random expectation of 0.5, and a heterospecific frequency of 0.1 is an 80% reduction of 0.5).

There are several major advantages to using this simple linear formulation for estimating the strength of reproductive isolation. The first is that the index calculated can be directly related to gene flow as it represents the relative under-representation of heterospecific reproduction and over-representation of conspecific mating relative to expectations under random mating; i.e. how isolated a species is based on a specific barrier.

In addition to being directly applicable to gene flow, another major advantage is that RI_5 gives a meaningful value when heterospecific gene flow is favored. Previous studies have eliminated these data or set their values equal to zero because the potentially large negative values are not directly comparable to measures of isolation from the typical part of the range (Coyne and Orr 1989; 1997). This is not necessary when using RI_5 . This can be illustrated by imagining a case of disassortative mating in which females prefer males of the opposite species. In an extreme example, if the mating preferences used above were reversed such that females of species X prefer males of species O and

mate heterospecifically 90% of the time and conspecifically only 10%, RI_1 would yield a value of -8. The proposed method using RI_5 would instead give a value of -0.8. Unlike the value obtained by RI_1 , this index has biological significance. The negative value indicates that the preference trait facilitates gene flow rather than confers isolation, and the magnitude denotes that there is 80% more heterospecific mating (i.e. gene flow) than expected by chance. This property allows the inclusion of these data when comparing to data from the positive range, and facilitates creating composite metrics that estimate the combined strength of multiple forms of isolation (e.g. Ramsey et al. 2003).

The third advantage of using the proposed RI_5 is that the same equation can be applied to both pre- and postzygotic isolation barriers and produce comparable values. While the explanation for applying this equation to prezygotic is reasonably intuitive, it may seem less so for postzygotic isolation. For example, if species X has relative fitness of 1 and hybrids that result from mating females of species X to males of species O have half the fitness, RI_5 would result in an isolation coefficient for females of species X of 0.33. The biological reasoning for this result becomes clear when imagining a simplified scenario where postzygotic isolation is the only barrier present and manifests as a difference in hybrid survival rates. Under random mating, females of species X will mate with males of species O with a frequency of 0.5. With a hypothetical starting population of 20 X females, 10 will mate with other X males while 10 will mate with species O males. If each of those matings result in one offspring, the next generation will have 10 species X offspring and 10 X/O hybrids. If half the X/O hybrids die as a result of postzygotic isolation, the total population consists of 15 individuals, 5 of which are

hybrids. A null expectation for 15 offspring of species *X* females is that 7.5 will be species *X* and 7.5 will be *X/O* hybrids; therefore, the 5 viable hybrids that are observed represents a reduction in gene flow of 1/3 relative to the random expectation (and 10 conspecific offspring represents 1/3 more than the null expectation of 7.5).

To illustrate that measures of pre- and postzygotic isolation using the proposed method of RI_5 have equivalent effects on gene flow, Figure 1.2 provides a pair of hypothetical scenarios in which a population begins with 10 females of species *O*. Figure 1.2A shows the effect of a mating preference that results in conspecific mating with a frequency of 0.9 and heterospecific frequency of 0.1. Using RI_5 , the reproductive isolation value obtained from this hypothetical scenario is 0.8. If each female produces 10 offspring, after reproduction there should be 90 species *O* offspring and 10 *O/X* hybrids based on these preferences. We impose postzygotic isolation of 0.9 on both conspecific and heterospecific offspring (the exact value is irrelevant as long as the two classes of offspring are equal). This results in a population with 81 species *O* offspring and 9 *O/X* hybrids. We can again see the biological implication of a reproductive isolation value of 0.8 in the fact that there are 80% fewer heterospecific offspring than expected by chance.

Figure 1.2B shows that the same amount of postzygotic isolation will result in an equivalent effect on gene flow. In this scenario the same number of females of species *O* begin with no mating isolation, such that the offspring produced in the next generation are 50 species *O* and 50 *O/X* hybrids. To impose postzygotic isolation of 0.8, we set the survival rate of conspecific offspring to 0.9 and the survival of heterospecific offspring to 0.1. After passing through this selective filter, the population consists of 45 species *O* offspring and 5 *O/X* hybrids. The isolation coefficient of 0.8 again makes biological sense

because in a population of 50 offspring of species *O* females, the expectation under a null model is 25 of each offspring class. The coefficient of 0.8 represents both the proportional increase in conspecific offspring and the proportional decrease in heterospecific offspring. Again this direct estimate of the reduction of heterospecific offspring is directly related to the decline in gene flow experienced compared to the null expectation.

Accommodation of Unequal Null Expectations

In the above examples, the simplifying assumption of equivalent null expectations of con- and heterospecific gene flow allowed the utilization of R/I_5 . However, as Martin & Willis (2007) argue, it is sometimes necessary to scale indexes of reproductive isolation to different null expectations, such as when there are consistent differences in relative abundances or gamete numbers between species. Using a similar convention to that proposed by Martin & Willis (2007), the observed frequencies of con- and heterospecific gene flow can be made proportional to their individual expected frequencies: *observed C/expected C* and *observed H/expected H*. The average over-representation of conspecific and under-representation of heterospecific can be calculated by simply dividing the sum by two:

$$\frac{(\textit{observed C/expected C}) + (\textit{observed H/expected H})}{2}.$$

To calculate isolation, the relative proportion of the term (*observed H/expected H*) to this average is subtracted from 1:

$$1 - \frac{(observed\ H/expected\ H)}{(observed\ C/expected\ C) + (observed\ H/expected\ H)} \cdot \frac{1}{2}$$

This equation can be rearranged to give:

$$RI_6 = 1 - \frac{2(observed\ H/expected\ H)}{(observed\ C/expected\ C) + (observed\ H/expected\ H)} \quad (6)$$

This form is easily related to our simplified RI_5 version, and indeed, if *expected H* and *expected C* are equivalent, the equation is identical to RI_5 . The probability of

heterospecific gene flow from RI_5 of $\frac{H}{C + H}$ is elaborated to accommodate the unequal

expected values as $\frac{observed\ H/expected\ H}{observed\ C/expected\ C + observed\ H/expected\ H}$. The RI_6 form of

the equation results in a non-linear relationship between the reproductive isolation index and observed frequency of heterospecific gene flow for all values of expected H and C except 0.5, but is still bounded by 1 and -1 at the extremes of complete isolation and disassortative mating/heterosis (Figure 1.3). The line passes through the x-axis at the expected value of heterospecific gene flow (i.e. reproductive isolation equals zero when the *observed H* equals the *expected H*).

As an example, if species *X* and species *O* differ in relative abundance in areas of sympatry such that on average, the frequency of species *X* is 0.6 and the frequency of species *O* is 0.4, then there is no longer an expectation of a 50:50 ratio in hetero:conspecific mating under a random expectation. Instead, even if there are no mating preferences, species *X* will mate conspecifically 60% of the time by chance. Therefore, if species *X* is observed mating heterospecifically 20% of the time and

conspecifically 80%, there is reproductive isolation due to mating preference ($RI_6 = 0.45$), but it is milder than would be calculated without taking the altered null expectation into consideration ($RI_5 = 0.6$).

Reproductive barriers that affect co-occurrence

The isolating barriers discussed so far have been based on traits that can both limit (reproductive isolation) or facilitate gene flow (e.g. disassortative mating, heterosis). However, forms of isolation that affect the potential for co-occurrence in either time or space are generally based on traits that can not act to facilitate heterospecific gene flow, and are therefore bound by 0 (no isolation) and 1 (complete isolation) rather than -1 and 1. These include ecogeographic isolation, temporal isolation, and some forms of pollinator isolation data, such as lists of shared and non-shared pollinators. These barriers have some unique features that affect the calculation of isolation. For example, the distribution of flowering times for species X may be completely separate from species O , resulting in complete reproductive isolation. This is equivalent to a mating isolation situation in which species X mates conspecifically 100% of the time. Flowering time could also overlap completely, which would result in zero reproductive isolation. This is equivalent to a situation in which no mating preferences exist, and therefore random mating occurs. The difference between mating isolation and temporal isolation is that mating preferences can occur that would actually facilitate gene flow (i.e. disassortative mating), while there is no equivalent ability for flowering time traits to increase gene flow beyond random mating.

While the simple form of equation RI_4 still applies, calculating the probability of heterospecific gene flow for barriers that affect co-occurrence is less straightforward than above. In these instances there are both shared regions of a distribution and non-shared regions. Within the shared region the expectation of heterospecific gene flow is 0.5 (or can be altered to a different expectation when necessary), while in the non-shared region, the expected heterospecific gene flow is 0. Therefore, the probability of heterospecific gene flow is equal to the relative proportion of half the shared overlap area to the total area occupied, and can be substituted for x in equation 4 to yield:

$$RI = 1 - 2\left(\frac{1/2 \text{ Shared}}{\text{Unshared} + \text{Shared}}\right),$$

which simplifies to:

$$RI_7 = 1 - \frac{\text{Shared}}{\text{Unshared} + \text{Shared}}. \quad (7)$$

It is necessary to use this form of the equation under any circumstances in which there are shared and non-shared portions of a distribution in which interactions can occur. Using equation RI_7 assumes that the shared region of overlap between species results in a 50:50 ratio of hetero:conspecific mating. This may be nearly true in many cases as relative abundances may be close to 50% each when averaged across the entire shared region of the distribution.

However, estimates of reproductive isolation can be refined when more specific data about relative abundance of each species can be collected (Martin and Willis 2007). Using temporal isolation of flowering plants as an example, when each species can be assumed to be of similar abundance overall, the probability of heterospecific gene flow on day i can be calculated by estimating the proportion of species A flowering on day i to

its total abundance throughout the growing season (A_i/A_{total}) and multiplying by the relative abundance of the heterospecific species on day i ($B_i/(A_i+B_i)$). Summing these products across all days of the growing season of species A and substituting into the probability of heterospecific from equation RI_4 , the isolation index can be calculated by:

$$RI_8 = 1 - 2 \sum_i \left(\frac{A_i}{A_{total}} \times \frac{B_i}{A_i + B_i} \right). \quad (8)$$

This equation can be further complicated by incorporating different global expected frequencies of each species based on consistent differences in relative abundance. In these cases, the terms involving both heterospecific and conspecific mating must be scaled by the non-equivalent expected values, resulting in:

$$RI_9 = 1 - 2 \frac{\frac{\text{observed } H}{\text{expected } H}}{\frac{\text{observed } C}{\text{expected } C} + \frac{\text{observed } H}{\text{expected } H}}; \text{ which can be expressed as:}$$

$$RI_9 = 1 - 2 \frac{\frac{\sum_i \left(\frac{A_i}{A_{total}} \times \frac{B_i}{A_i + B_i} \right)}{B_{total} / (A_{total} + B_{total})}}{\frac{\sum_i \left(\frac{A_i}{A_{total}} \times \frac{A_i}{A_i + B_i} \right)}{A_{total} / (A_{total} + B_{total})} + \frac{\sum_i \left(\frac{A_i}{A_{total}} \times \frac{B_i}{A_i + B_i} \right)}{B_{total} / (A_{total} + B_{total})}} \quad (9)$$

This equation is very similar to the one employed by Martin & Willis (2007); however, their base equation was equivalent to the form shown in RI_1 , which we argue should be replaced with our base of RI_4 .

Figure 1.4 gives a simple example of data that can be calculated using these methods. Figure 1.4A represents the data that are most available to researchers as the total number of days in which there was co-occurrence and non co-occurrence of sexually

mature individuals. In the example, species *X* mates on days 1 through 20, while species *O* mates on days 15 through 34. This results in an overlap of 6 days in which both species co-occur. Using these simplest data available and form of the equation (RI_7), an isolation index of 0.7 would be reached for species *X*. If the abundance of individuals on a given date are available (e.g. number of reproductive adults on each day), a more precise estimate can be achieved using RI_8 or RI_9 , depending on the equivalency of the overall abundance of the two species. When the overall abundances of the two species are the same, temporal isolation for species *X* would be calculated as $RI_8 = 0.81$ (Figure 1.4B). However, if species *O* is twice as abundant as species *X*, then the global expectation of heterospecific mating under the null model is no longer 50%. Instead, in reference to species *X*, we expect that overall it will mate heterospecifically 2/3 of the time if mating is random; i.e., *expected H* = 0.67 and *expected C* = 0.33. Using RI_9 , we most accurately represent isolation as 0.87 (Figure 1.4C).

Geographic isolation is another form of isolation that affects co-occurrence of species, and should be estimated using equations RI_7 , RI_8 or RI_9 depending on the data available. Geographic isolation can be incorporated into the biological species concept (Mayr 1942) by estimating the portion of geographic isolation that is due to intrinsic biological differences between organisms (Sobel et al. 2010). This form of isolation is very similar mathematically to the temporal case presented above, with only some additional complexity added by an extra dimension. As above, for simplification it is assumed that there is a shared region of overlap in which species can interbreed freely and a region of non-overlap in which no interbreeding occurs (though in reality this may

occur infrequently due to long distance dispersal). As an example, consider two species in which geographic distributions are based on intrinsic genetic differences, such that ecogeographic isolation exists (Sobel et al. 2010). For simplification, species *X* starts with 1000 individuals spread evenly across its geographic distribution, and each individual produces one offspring. If the ecogeographic extent of species *X* overlaps species *O* by 20% (i.e. 1/5 of the ecogeographic range of species *X* is shared with *O* and 4/5 is unshared), then after reproduction, 800 offspring will be species *X* from the offspring produced in the unshared portion of the range. In the shared portion random mating would produce 100 species *X* offspring and 100 *X/O* hybrids. In total 100 offspring are *X/O* hybrids, representing a 80% reduction in heterospecific offspring relative to a null expectation, and 900 offspring are species *X*, representing an 80% increase in conspecific offspring. Therefore, isolation is equal to 0.8, which is the value obtained when using RI_7 . Of course, the assumption that all individuals will be evenly distributed across the range of a species will often be violated. Therefore, to employ equations RI_8 or RI_9 , relative abundances across multiple contact zones can be incorporated into the calculation.

Summary

Many would agree that one of the ultimate goals of speciation research is to uncover which forms of reproductive isolation are most important to the process. Ideally, we want to know what the relative contributions of each barrier to total reproductive isolation are at the exact moment of speciation. Different researchers may approach this problem in various ways, by either trying to understand the evolution of specific forms of

reproductive isolation through time or by focusing on differences among the strength of barriers experienced by recently diverged species. Regardless, estimating the strength of reproductive isolation is an important part of any approach to understanding the origin of species, but no single method for making the calculations necessary has emerged.

We here provide a simple linear solution with many advantages for estimating the strength of individual barriers. While the proposed method does not differ significantly over much of the useable range of the metric, it is both simple to understand (in its base form) and easily related to differences in the probability of heterospecific gene flow. It can also be used in cases that depart from the general 50:50 null expectation, allowing the most accurate measures of isolation to be estimated. The proposed method also enables the incorporation of values below zero that previously had been discarded because of the asymmetry of previous equations. It therefore allows researchers the opportunity to make comparisons among forms of isolation and taxa by providing metrics that have equivalent effects on gene flow. RI_4 provides a framework in which to place estimates of the probability of heterospecific gene flow, but calculating this probability can sometimes be challenging. The estimate for each form of isolation can be refined further by adjusting the equation to non-equivalent null expectations when appropriate data exist. (RI_6 , RI_9). Differences from equivalency can occur by a number of factors including differences in relative abundance, gamete abundance, mating frequency, etc. By selecting the appropriate form of these equations, future workers can be assured of producing estimates of isolation that are as accurate as possible while allowing direct comparisons with other estimates.

Table 1.1. Alternative methods for calculating reproductive isolation indexes. *H* refers to an aspect of heterospecifics (e.g. heterospecific mating, offspring production, etc.) while *C* refers to conspecifics.

Equation	Type of isolating barriers	Algebraic equivalent	Reference
$RI = 1 - \frac{\text{frequency of heterospecific matings}}{\text{frequency of homospecific matings}}$	Prezygotic (mating preference)	$RI_1 = 1 - \frac{H}{C}$	(Coyne and Orr 1989; 1997)
$RI = \frac{\text{number of homotypic matings} - \text{number of heterotypic matings}}{\text{number of homotypic matings} + \text{number of heterotypic matings}}$	Prezygotic (mating preference)	$RI_5 = 1 - \frac{2H}{C + H}$	(Cayol et al. 1999)
$RI = 1 - \frac{\text{number of cross - species foraging bouts}}{\text{total number of foraging bouts}}$	Prezygotic (pollinator fidelity)	$RI_2 = 1 - \frac{H}{C + H}$	Ramsey et al. (2003)
$RI = 0, 0.25, 0.5, 0.75, \text{ or } 1$	Postzygotic (hybrid viability and sterility)	NA	(Coyne and Orr 1989; 1997)
$RI = 1 - \frac{\text{observed/expected heterospecific matings}}{\text{observed/expected homospecific matings}}$	Prezygotic (flowering time, etc.)	$= RI_1$	Martin and Willis (2007)
$RI = \frac{(1 - \text{proportion fertile hybrid males/proportion fertile pure males})}{2}$	Postzygotic (hybrid sterility)	$= \frac{1}{2} RI_1$	(Bono and Markow 2009)
$RI = 1 - \frac{\text{proportion of hybrid seed}}{(1 - \text{proportion of hybrid seed})}$	Prezygotic (hybridization rate)	$= RI_1$	(Brock 2009)
$RI = 1 - \frac{2 \times \text{average fitness of the hybrid offspring}}{\text{average fitnesses within the two allopatric populations}}$	Postzygotic (incompatibility)	$= RI_5$	(Palmer and Feldman 2009)
$RI = 1 - \frac{\text{proportion of seeds sired by the minor donor}}{\text{proportion of seeds sired by the major donor}}$	Postpollination (seed production)	$= RI_1$	(Ruane 2009)
$RI = 1 - \frac{\text{Prob. mating with heterospecific male}}{\text{Prob. mating under random choice } (= 0.5)}$	Prezygotic (mating preference)	$= RI_5$	(Takami et al. 2007)
$RI = \frac{\% \text{ of successful allopatric pairings} - \% \text{ of successful sympatric pairings}}{\% \text{ of successful sympatric pairings}}$	Prezygotic (mating preference)	$= -RI_1$	(Hosken et al. 2009)

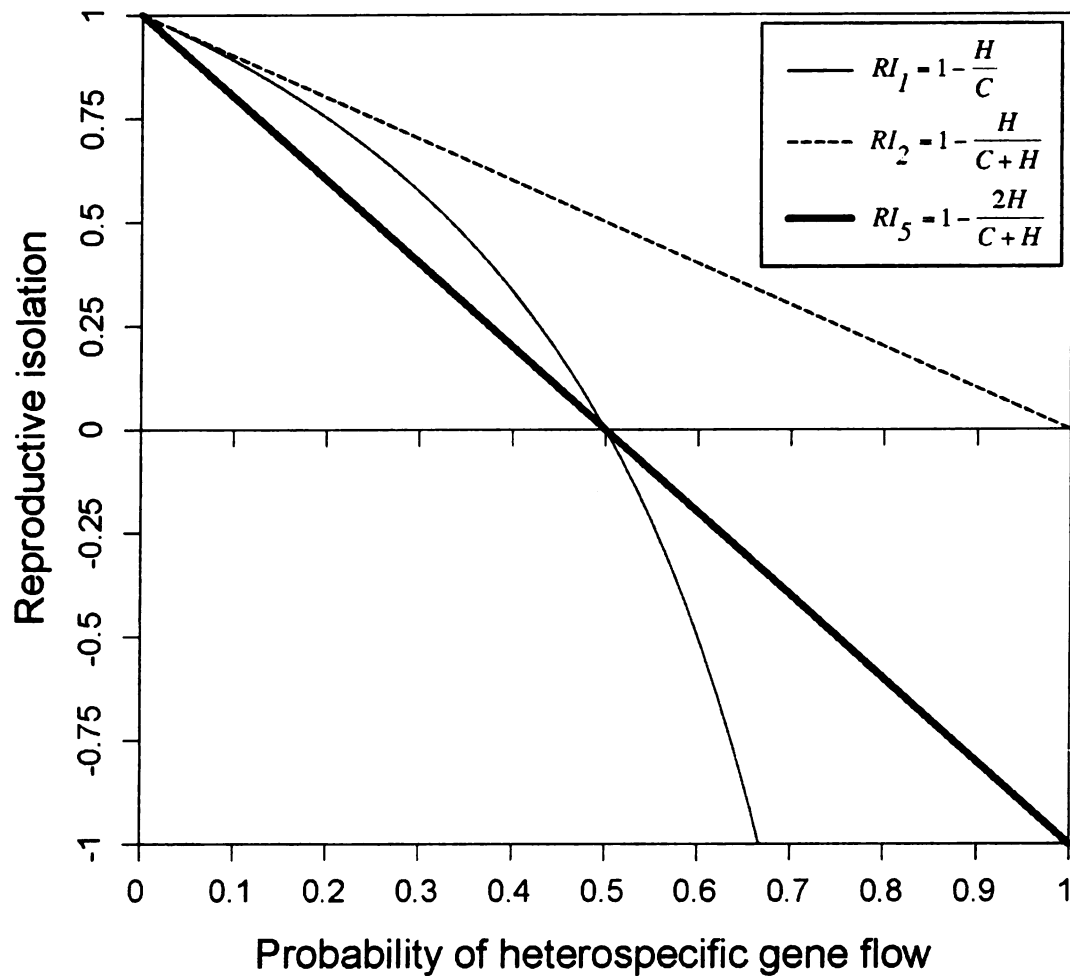


Figure 1.1. Relationship between the probability of heterospecific gene flow and the measure of reproductive isolation obtained through three commonly used methods. RI_1 ranges from 1 at complete isolation, through 0 at random mating (probability of gene flow = 0.5), and to $-\infty$ when gene flow is facilitated. RI_2 ranges from 1 at complete isolation to 0.5 at random mating and 0 when gene flow is facilitate. RI_5 (and RI_3) range from 1 at complete isolation, 0 at random mating, and -1 when gene flow is facilitated.

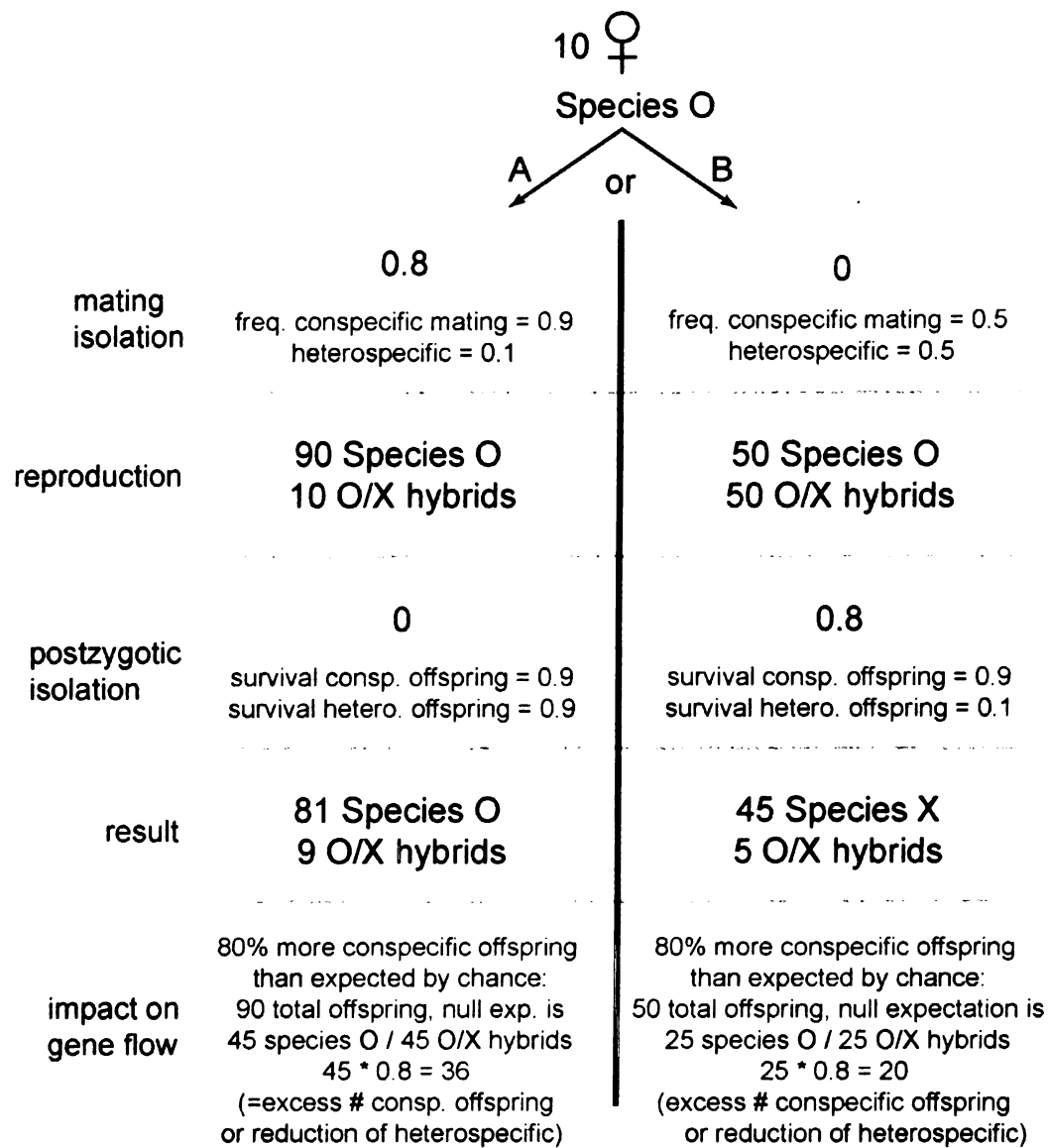


Figure 1.2. Diagram showing similarities between pre- and postzygotic isolation index measures of equal magnitude ($RI = 0.8$). In both A & B, the population begins with ten females, and each produces 10 offspring. A) Species O females have strong mating preferences, mating with conspecific males with a frequency of 0.9 and heterospecific males of species X with a frequency of 0.1. Of the 100 offspring produced, 90 are pure species O and 10 are O/X hybrids. The isolation coefficient of 0.8 represents the proportional decrease in heterospecific offspring below a null expectation. B) Species O females have no mating preferences, but do exhibit strong postzygotic isolation such that the survival rate of conspecific offspring is 0.9 while the heterospecific offspring have a lower survival rate of 0.1. This results in an equivalent effect on gene flow with 80% fewer heterospecific offspring (i.e. gene flow) than would have been produced under the null model.

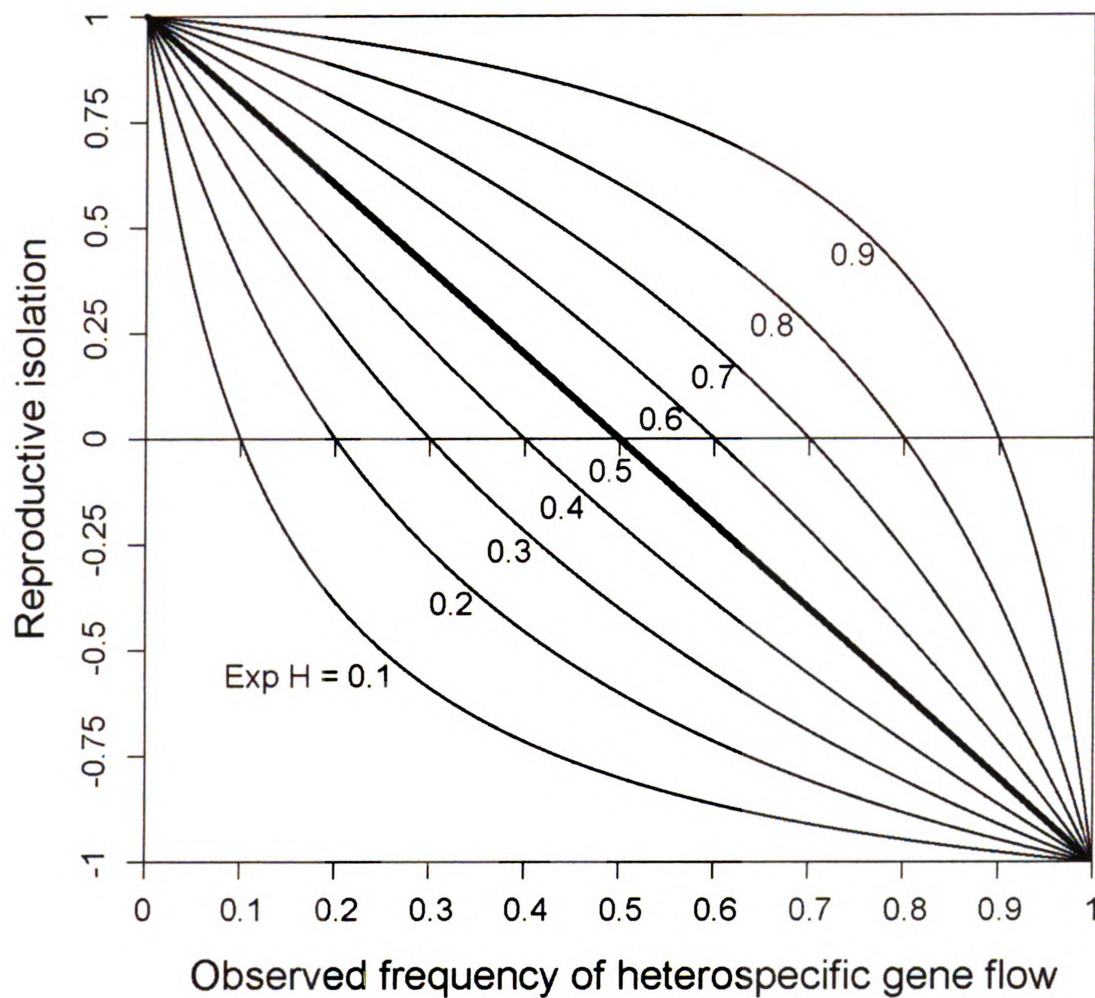
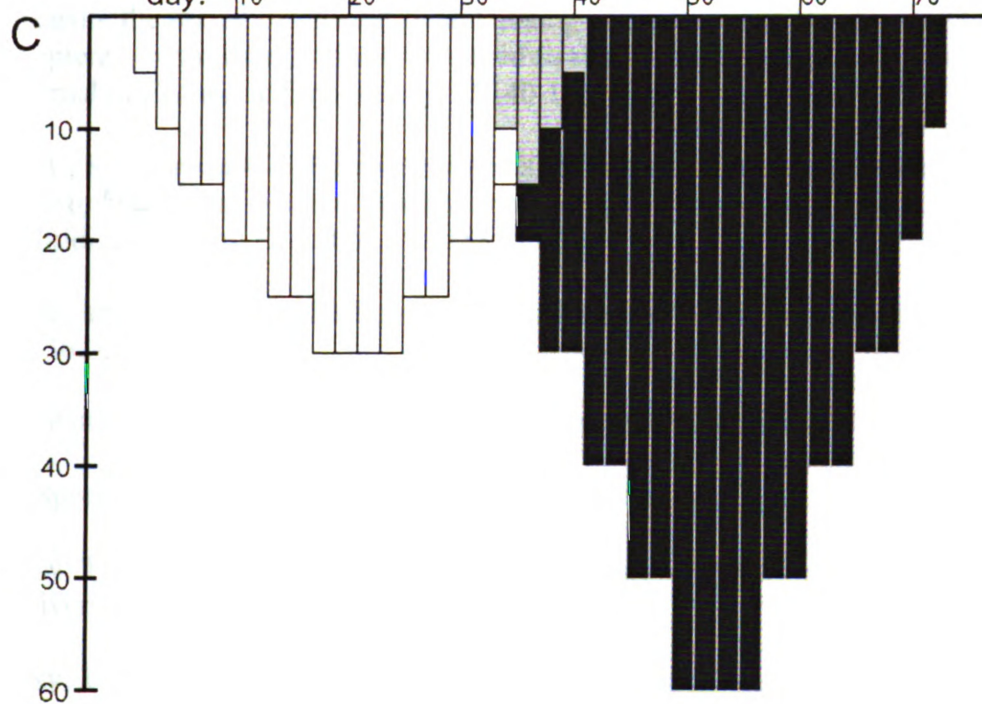
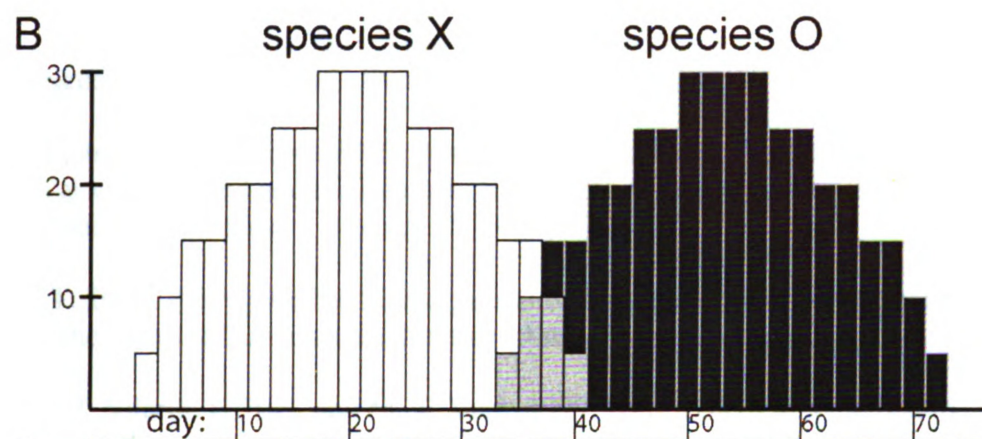
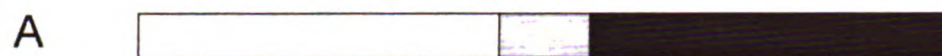


Figure 1.3. The effect of changes in expected frequency of hybrid formation on the value of reproductive isolation using equation RI_6 . Each line represents a different expectation of hybrid formation (due to differences in relative abundance for example) ranging from 0.1 to 0.9 (labeled below each line). The point at which the expected H line intersects the x-axis is the point at which observed heterospecific gene flow equals expected heterospecific gene flow; therefore reproductive isolation equals zero. The thick line represents the point of equal expectation of con- and heterospecific gene flow, which is equivalent to RI_5 from Figure 1.1.

Figure 1.4. Hypothetical data showing the relationship between three methods for calculating the strength of temporal isolation. Days of reproduction is on the x-axis and number of reproductive individuals is on the y-axis (mirrored). A) Only data on days of reproduction is available. Species *X* has sexually mature individuals on days 1 to 20 while species *O* is mature on days 15 to 34. Given the data available, the most simple equation (RI_7) is used, yielding an isolation coefficient of 0.7 for species *X*. B) If data on relative abundance throughout the mating season are available, and both species have roughly equivalent relative numbers, the estimate of isolation can be refined by using equation RI_8 , giving an isolation index of 0.81 for species *X*. C) When species *X* and *O* exhibit consistent differences in relative abundance, the null expectation under random mating can be adjusted using equation RI_9 , resulting in an isolation coefficient of 0.87.



Literature Cited

- Ackermann, M., M. Achatz, and M. Weigend. 2008. Hybridization and crossability in *Caiophora* (Loasaceae subfam. Loasoideae): Are interfertile species and inbred populations results of a recent radiation? *American Journal of Botany* 95:1109-1121.
- Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* 59:1754-1767.
- Bono, J. M., and T. A. Markow. 2009. Post-zygotic isolation in cactophilic *Drosophila*: larval viability and adult life-history traits of *D. mojavensis*/*D. arizonae* hybrids. *Journal of Evolutionary Biology* 22:1387-1395.
- Brock, M. T. 2009. Prezygotic barriers to gene flow between *Taraxacum ceratophorum* and the invasive *Taraxacum officinale* (Asteraceae). *Oecologia* 161:241-251.
- Cayol, J. P., J. Vilardi, E. Rial, and M. T. Vera. 1999. New indices and method to measure the sexual compatibility and mating performance of *Ceratitis capitata* (Diptera : Tephritidae) laboratory-reared strains under field cage conditions. *Journal of Economic Entomology* 92:140-145.
- Coyne, J. A., S. Elwyn, and E. Rolan-Alvarez. 2005. Impact of experimental design on *Drosophila* sexual isolation studies: Direct effects and comparison to field hybridization data. *Evolution* 59:2588-2601.
- Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43:362-381.
- . 1997. "Patterns of speciation in *Drosophila*" revisited. *Evolution* 51:295-303.
- . 2004. *Speciation*. Sunderland, MA, Sinauer Associates.
- Dobzhansky, T. 1937, *Genetics and the Origin of Species*. New York, Columbia University Press.
- Dobzhansky, T., and P. C. Koller. 1938. An experimental study of sexual isolation in *Drosophila*. *Biologisches Zentralblatt* 58:589-607.
- Dopman, E. B., P. S. Robbins, and A. Seaman. 2010. Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution* 64:881-902.
- Ehrman, L. 1965. Direct observation of sexual isolation between allopatric and between sympatric strains of the different *Drosophila paulistorum* Races. *Evolution* 19:459-464.

- Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America* 103:3209-3213.
- Hosken, D. J., O. Y. Martin, S. Wigby, T. Chapman, and D. J. Hodgson. 2009. Sexual conflict and reproductive isolation in flies. *Biology Letters* 5:697-699.
- Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363:3009-3021.
- Martin, N. H., and J. H. Willis. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61:68-82.
- Matsubayashi, K. W., and H. Katakura. 2009. Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles. *Evolution* 63:2563-2580.
- Mayr, E. 1942. *Systematics and the origin of species*. New York, Columbia University Press.
- Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma*). *Evolution* 57:317-327.
- Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58:1195-1208.
- Palmer, M. E., and M. W. Feldman. 2009. Dynamics of hybrid incompatibility in gene networks in a constant environment. *Evolution* 63:418-431.
- Poulton, E. G. 1908. What is a species? Pages 46-94 *Essays on Evolution 1889-1907*. London, Oxford at the Clarendon Press.
- Presgraves, D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. *Evolution* 56:1168-1183.
- Price, T. D., and M. M. Bouvier. 2002. The evolution of F-1 postzygotic incompatibilities in birds. *Evolution* 56:2083-2089.

- Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57:1520-1534.
- Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we learned in 40 years? *Evolution* 47:1637-1653.
- Ruane, L. G. 2009. Post-pollination processes and non-random mating among compatible mates. *Evolutionary Ecology Research* 11:1031-1051.
- Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. *Evolution* 52:1811-1820.
- Schemske, D. W. 2000. Understanding the origin of species. *Evolution* 54:1069-1073.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295-315.
- Stalker, H. D. 1942. Sexual isolation studies in the species complex *Drosophila virilis*. *Genetics* 27:238-257.
- Takami, Y., N. Nagata, M. Sasabe, and T. Sota. 2007. Asymmetry in reproductive isolation and its effect on directional mitochondrial introgression in the parapatric ground beetles *Carabus yamato* and *C. albrechti*. *Population Ecology* 49:337-346.
- Taylor, S. J., M. Arnold, and N. H. Martin. 2009. The genetic architecture of reproductive isolation in Louisiana irises: Hybrid fitness in nature. *Evolution* 63:2581-2594.
- Tilley, S. G., P. A. Verrell, and S. J. Arnold. 1990. Correspondence between sexual isolation and allozyme differentiation: A test in the salamander *Desmognathus ochrophaeus*. *Proceedings of the National Academy of Sciences of the United States of America* 87:2715-2719.

CHAPTER 2: ECOGEOGRAPHIC ISOLATION AMONG RECENTLY DIVERGED
SPECIES PAIRS IN THE GENUS *MIMULUS*

CHAPTER 2: ECOGEOGRAPHIC ISOLATION AMONG RECENTLY DIVERGED SPECIES PAIRS IN THE GENUS *MIMULUS*

Abstract

Despite a long history of examining the geographic context of speciation, evolutionary biologists rarely consider differences in geographic range as a legitimate form of reproductive isolation. This stems mainly from the fact that geographic ranges are both the product of historic and ecological processes. The presented work provides a method for estimating the ecological aspects of geographic isolation, bringing this potentially important form of isolation under the umbrella of the biological species concept.

Ecological niche modeling was used to evaluate the potential ranges of twelve pairs of species in the genus *Mimulus*. There was substantial variation in which environmental variables used to construct niche models were most important, with aspects of geology and rainfall providing the most explanatory power. Examination of differences in niche models generated for species pairs showed that recently diverged species of *Mimulus* differ significantly in the niches they occupy. These differences in niche translate into strong ecogeographic isolation, with an average strength of at least 0.64. Given the position of this barrier in the order of sequentially acting reproductive barriers, this value leads to the conclusion that ecogeographic isolation will be the highest contributor to total isolation among all other forms of reproductive isolation.

Introduction

Identifying the mechanisms of speciation is an area of considerable importance in evolutionary biology, and one of the most debated topics in speciation biology is the relative impact of different forms of reproductive isolation (Coyne and Orr 2004). This debate typically simplifies into an argument of whether barriers that act prior to fertilization (prezygotic) or after fertilization (postzygotic) are more important (Coyne and Orr 2004; Poulton 1908; Rice and Hostert 1993; Schemske 2000). Despite the fact that a complete inventory of potential reproductive barriers has been recognized since the modern synthesis (Dobzhansky 1937; Mayr 1942), several forms of reproductive isolation have been essentially ignored in this debate. In particular, the role of geographic separation in limiting gene flow has been largely ignored, with very few estimates of its potential impact on gene flow (but see Kay 2006; Kirkpatrick and Ravigne 2002; Ramsey et al. 2003).

Although its effects on gene flow are rarely measured as an actual prezygotic barrier, the geography of speciation is far from an unexplored topic. Jordan first recognized the role of geography in generating species formulating the rule, *"Given any species, in any region, the nearest related species is not to be found in the same region nor in a remote region, but in a neighboring district separated from the first by a barrier of some sort or at least by a belt of country, the breadth of which gives the effect of a barrier."* (Jordan 1908; Jordan and Kellogg 1907, p. 120)

Indeed, much of the speciation research in the past several decades has been squarely focused on whether Jordan's rule is correct, asking if allopatry is a requirement for speciation (Bolnick 2004; Bolnick and Fitzpatrick 2007; Coyne and Orr 2004; Feder et al.

2005; Mayr 1947; Nosil 2008; Turelli et al. 2001; Via 2001). Some authors have measured the degree of overlap in geographic space occupied by recently diverged species (Barraclough and Vogler 2000; Fitzpatrick and Turelli 2006), but almost exclusively within the context of determining whether allopatric or sympatric speciation is most common. Despite this substantial attention, most workers have treated allopatric separation itself as a completely historical process with very little regard for how the biology of organisms might be involved in creating or maintaining allopatry (Wiens 2004). This has resulted in geographic isolation being considered a potent part of speciation, but mostly as a force that allows the evolution of additional forms of isolation rather than as a legitimate form of isolation itself.

Historically, several authors pointed out the need to consider geographic isolation as a barrier *per se*, though often with some trepidation. For example, Stebbins notes that Jordan's rule is often correct, but adds that closely related species "... are usually separated from their relatives by ecological barriers as well as by geographic ones (Stebbins 1950, p. 197)." Dobzhansky also noted that, "... the occupation of separate areas by two species may be due not only to the fact that they have developed there, but also to the presence of physiological characteristics that make each species attached to the environment...(Dobzhansky 1937, p. 231)." These views suggest a revision to Jordan's rule that might read, closely related species are most often near each other but separated by a geological barrier, and also exhibit ecological divergence that may enhance the fidelity of each to its home range.

As argued in Sobel et al. (2010), an examination of the definition of reproductive barriers under the Biological Species Concept (BSC) makes it possible to reconcile issues

involved in treating geographic isolation as a legitimate prezygotic barrier. The BSC identifies reproductive barriers as biological differences between populations that actually or potentially limit gene flow (Mayr 1942; Mayr 1963; Mayr 1984). Therefore, the aspects of geographic isolation that are based on historical factors and not biological differences are not legitimate barriers, but geographic separation that arises (or is maintained) by intrinsic biological differences between populations can and should be measured as a component of prezygotic isolation, which is referred to as ecogeographic isolation (Sobel et al. 2010).

Several advances have made it possible to explore the relationship between history and biology in deconstructing their relative role in geographic isolation. The first is the availability of species-level molecular phylogenetic hypotheses for many groups of organisms (Barraclough and Vogler 2000; Graham et al. 2004b), providing an objective identification of species pairs that have recently diverged. Another is GPS (geographic positioning system) technology for collecting detailed spatial information on the distribution of species, and the computing power required to manage large databases of geographic information (GIS technology). Combining these advances with the availability of museum and herbaria records of species collections (Graham et al. 2004a), the ability to catalog and analyze the spatial patterns of speciation have grown immensely in recent years.

Ecological niche modeling has emerged as a powerful method for examining the climatic tolerances and geographic ranges of species (Ortega-Huerta and Peterson 2008), identifying suitable habitat for species in under-sampled regions (Jarnevich et al. 2006), and predicting the outcome of species invasions (Peterson 2003). In addition, it has been

recently adopted by evolutionary biologists (Kozak et al. 2008), and has been applied to questions of whether ecological niche conservatism (Kozak and Wiens 2006) or divergence (Nakazato et al. 2008; Nakazato et al. 2010; Warren et al. 2008) are general components of speciation.

This method has the potential to also aid in deconstructing the historical and biological components that determine geographic separation between species (Sobel et al. 2010). Because ecological niche models show both where a species does live and where it can live, it provides an estimate of both the actual and potential range of a species. By overlaying individual niche models of a pair of closely related species, it is therefore possible to simultaneously assess the actual and potential gene flow limited by the geographic range occupied by species. Indeed, the overlap of two species' spatially projected ecological niche model provides an estimate of how much physical space in a region that two species are capable of co-occurring over. This provides an estimate of the degree to which adaptation to different habitats results in actually and/or potentially geographically distinct ranges, and allows a measure of geographic isolation relevant under the BSC called ecogeographic isolation. In the study presented, ecological niche modeling was applied to a diverse group of western North American wildflowers in the genus *Mimulus*. Ecological niche models were developed for 12 pairs of species that represent recent diverged speciation events across the genus, and overlap in niche models was used to estimate the relative contribution of this potentially important barrier.

Materials & Methods

Study system

The genus *Mimulus* consists of approximately 120 species of wildflowers, approximately 75% of which are found in western North America (Grant 1924). The genus exhibits a broad diversity in interesting ecological and evolutionary traits including: growth form (Lowry et al. 2008), mating systems (Fishman et al. 2002; Willis 1993), ploidy levels (Beardsley et al. 2004; Vickery 1995), pollination syndromes (Bradshaw and Schemske 2003; Bradshaw et al. 1995; Schemske and Bradshaw 1999; Streisfeld and Kohn 2005), edaphic specialization (Macnair 1983), and climatic tolerances (Angert and Schemske 2005). Previous phylogenetic work has elaborated relationships at broad taxonomic levels (Beardsley and Olmstead 2002), and a species-level phylogeny is available with nearly complete coverage of the North American species (Beardsley et al. 2004). A rich history of ecological and genetic work (Hiesey et al. 1971; Kiang 1972; Vickery 1959; Vickery 1964), combined with more recent advances in developing genomic resources (Wu et al. 2008) has elevated the genus to a model system in ecology and evolutionary biology, especially in studies of speciation (Martin and Willis 2007; Ramsey et al. 2003).

Selection of species pairs

The phylogeny presented in Beardsley et al. (2004) was used to aid species selection. For inclusion, species pairs exhibited recent speciation events and had high bootstrap support from the molecular phylogenetic hypothesis presented. In order to verify that the species pairs selected for inclusion were more closely related to each other

than to any other species in the study, sequence data (*trnL/F*, *ITS*, *ETS*) presented in Beardsley et al. (2004) were reanalyzed. A maximum likelihood phylogenetic analysis was performed in PHYLIP (Felsenstein 2005), and the majority rule consensus tree supports the relationships drawn from the genus-wide phylogeny (Figure 2.1). In most cases, these species pairs represent each other's closest extant relative; however, in some cases alignment with species used in a broader study of reproductive barriers (Sobel, Chapter 3) necessitated choosing species pairs where another species may be more closely related to one of the included species. However, each pair of species used is more closely related to each other than any other species in the study (Figure 2.1), ensuring phylogenetic independence of reproductive isolation measured between each pair (Felsenstein 1985). In total, 12 pairs of closely related species were selected for inclusion (Table 2.1), spanning the phylogenetic breadth of the genus with members from traditional morphological taxonomic sections *Eunanus*, *Erythranthe*, *Oenoe*, and *Paradanthus* (Grant 1924).

Collection of georeferenced locality data

Georeferenced locality data were collected for each species using the Jepson Online Interchange for California Floristics (UC-JEPS2004). This database consists of specimen records of California plants compiled from seventeen different herbaria across the state, providing an unparalleled collection of *Mimulus* records. Recent accessions in the database are often georeferenced, but a large number of specimen records from older collections lack this detail. Using a combination of cartographic resources including Google Earth (<http://earth.google.com>) and National Geographic TOPO! software

(National Geographic TOPO! Maps, San Francisco, CA; <http://maps.nationalgeographic.com/topo>), latitude and longitude coordinates were added to samples without this data, resulting in over 2000 georeferenced accessions at an average of 84 specimens per species (Table 2.1).

Selection and manipulation of environmental layers

Eight climatic variables were selected from the publicly available dataset WORLDCLIM (<http://worldclim.org>) that represented a mix of both temperature and precipitation variables (Table 2.2). These climatic layers consist of a grid of 1 km^2 resolution pixels that span the globe. Variables were selected to minimize correlations among climatic layers within the study area; however, correlations remain high among some of the variables (all are <0.9 , Table 2.3). In addition, a geological map of California was obtained from the U.S. Geologic Survey (<http://usgs.gov>). This layer consists of a qualitative assessment of basic geologic parent material broken into 51 categories (e.g. sandstone, granodiorite, serpentine, etc.). All environmental variables were processed and visualized using ArcGIS 9.2 (ESRI, Redlands, California, USA). The original polygon geological shapefile was converted to a raster layer with the same spatial extent and resolution as the WORLDCLIM climatic variables using the ‘Polygon to Raster’ conversion tool within ArcGIS. Cells that contained multiple values were assigned using the maximum area criterion. Each WORLDCLIM layer was clipped to limit the spatial extent of analysis to the state of California in order to align with the geologic map. Limiting analysis to California is also justifiable given that the vast majority of accessions in the Jepson Interchange are from collections within the state, and

extrapolating to outside regions could prove problematic. Prior to analysis, any species occurrence data collected under different coordinate systems (e.g. NAD27, UTM, etc.) were converted to a common datum (WGS84) to match the environmental layers.

Ecological niche modeling

Ecological niche models (ENMs) were constructed for each species using Maxent software (version 3.3.1; <http://cs.princeton.edu/~shapire/maxent>) (Phillips et al. 2006). Maxent utilizes a machine-learning algorithm which employs the general principal that when estimating an unknown probability distribution, the best approximation will satisfy all known constraints on the distribution without imposing unfounded ones (Jaynes 1957). When applied to modeling the distribution of species, the approach therefore finds the probability distribution of maximum uniformity (entropy) while satisfying known constraints imposed by the observed distribution of species and environmental variables across the study area. The approach has been consistently shown to offer improved ability to accurately predict the ranges of species over previously employed methods such as GARP or GLMs (Elith et al. 2006). It is especially useful in situations where information on species occurrences is readily available but data on absences is not (Elith et al. 2006; Phillips et al. 2006), and has been utilized in a number of recent speciation studies using museum and herbarium records (e.g. Kozak and Wiens 2006; Nakazato et al. 2010).

Common species

For species with greater than 30 records, Maxent ENMs were generated using default settings of 500 maximum iterations and convergence threshold of 0.00001. The

random test percentage was set to 50%, so that a random half of the data was used to train the model and the other half was used to test the model. The raw Maxent output is a grid of pixels with the same spatial extent as the inputted environmental variables, where each pixel in the grid is assigned a non-negative probability which sums to 1 over all pixels. Given the large spatial extent of analysis (state of California), the values in each pixel are extremely small. Therefore, a cumulative representation was selected that results in each pixel carrying a value equal to the sum of that pixel and all other pixels of equal or lesser probability. The sum is multiplied by 100 to give a percentage, which can be thought of as a relative suitability score for each pixel. Jackknifes were performed in Maxent to measure the relative importance of each environmental variable to the predicted model. In addition, the threshold independent test indicator, AUC (area under the curve) of the ROC (receiver operator characteristic), was recorded as an indicator of Maxent performance in predicting niches.

Another test of Maxent niche modeling performance was omission rates of test samples based on delineated thresholds of the cumulative probability value (representing the cutoff for suitable and unsuitable habitat). Establishing threshold levels appropriate for producing binary suitable/unsuitable maps for species is an area of active research (Liu et al. 2005). Different threshold criteria have been shown to be optimal under different conditions, so to assess the performance of Maxent in producing *Mimulus* ENMs, two thresholds were established and the relative success rate of each was compared. The first threshold represents a maximized sum of the sensitivity and specificity values (Liu et al. 2005), which has been shown to perform relatively well compared to other methods and has been employed in recent studies of speciation (e.g.

Nakazato et al. 2010). This threshold attempts to balance the incidence of true positives (test species occurrence points assigned to suitable habitat) with the incidence of correctly identified negatives. Maxent simulates negative calls based on a random sampling of the background of the area of study, so correct negative calls are not actually based on known absences. The second threshold used is a cutoff of the 10th percentile of cumulative probability values that occur in pixels occupied by species occurrence records. While arbitrary, this cutoff has been shown to be an effective predictor of distributions (Pearson et al. 2007) and has also been used in recent studies of speciation (e.g. Kozak and Wiens 2006). In both cases, threshold performance was assessed by examining the rate of omission of test samples from the trained model.

Rare species

Species with highly restricted geographic ranges are common in *Mimulus*; therefore, there are several species in this study with very few collection records (Table 2.1). One of the benefits of using the Maxent platform is that it has been shown to be effective at very low sample sizes (Papes and Gaubert 2007; Pearson et al. 2007; Wisz et al. 2008), with reasonable performance with samples as few as 5 (Pearson et al. 2007). However, the 50% train/test method described above is not appropriate for these situations due to the limited number of samples. Instead following methods described by Pearson et al. (2007) a 'jackknife minus 1' cross validation technique was used for any species with fewer than 30 accessions. In brief, this method involves dividing the species occurrence data into N (# of occurrence points) subsamples, removing a single occurrence for each group. The $N-1$ portion of each group was used as training data and

the single occurrence removed was used to test the model (repeated N times so that every point is removed from one subsample). The success rate of this method was assessed by setting a threshold and recording the number of times the model successfully predicted suitable habitat for the test occurrence point. Two thresholds were selected for this analysis. The first is a conservative measure based on the minimum cumulative probability value present in the training data. The second is a subjective cutoff of the 10th percentile of the values present in the training data. Pearson et al (2007) provides a method and software for using the area of suitable habitat predicted under each threshold in a chi-square test for significant departure from randomness, and this method was used for statistical validation of model performance. As in the common species analysis presented above, AUC of the ROC (averaged across all subsample runs) was recorded as an indicator of Maxent performance in predicting niches. Environmental variables were jackknifed in Maxent to estimate the relative contribution of environmental variables to each ecological niche model.

Comparing niche models between recently diverged species pairs

Comparisons were made between species pairs by both contrasting the ecological niche model output of species pairs and by examining the environmental data associated with pixels containing species occurrence records. Ecological niche models were compared using the software ENMTools (<http://enmtools.blogspot.com>) (Warren et al. 2008). For each species pair, the niche overlap tool was used to first calculate a value of ecological niche similarity based upon both Schoener's D (Schoener 1968) and the I statistic described in Warren et al. (2008). The niche identity tool was used to randomly

assign species identity to the list of known localities for a species pair, and a Maxent ecological niche model was generated for each replicate. This was performed 100 times and a Schoener's D and I statistic were calculated for each iteration, generating a null distribution of niche overlap. Schoener's D and I statistics were compared to this null distribution to assess whether ecological niche models in species pairs of *Mimulus* differed from a chance expectation. ENMTools is not currently capable of handling categorical environmental variables, so this analysis was performed on only the WORLDCLIM set of climatic layers (Table 2.2).

Additional tests to examine the differences between species pairs were performed by extracting the environmental variable values associated with pixels containing species records. For each species, the 'Extraction-Sample' tool within the Spatial Analyst package of ArcGIS was used to compile these data. One-way MANOVA (Wilks λ) were conducted in R (R Development Core Team 2010) on the continuous environmental data extracted for each species pair to test for differences among the observed variables. Post-hoc analyses were performed on data from each pair with an α corrected for multiple comparisons using a Bonferroni adjustment ($\alpha = 0.00556$). Because many environmental variables were not normally distributed within species samples, the non-parametric Mann-Whitney U was employed to test for differences between the continuous WORLDCLIM variables. The categorical geological variable was tested using a contingency table analysis. Mann-Whitney U and contingency analyses were performed in JMP 8.0 (SAS Institute, Cary, North Carolina, USA).

Calculation of reproductive isolation

In order to calculate overlap in ecological niche models suitable for estimating reproductive isolation, threshold values need to be established so that overlap in binary suitable/unsuitable habitat can be assessed. Niche models generated in Maxent were loaded into ArcGIS, and the 'Raster Calculator' within the Spatial Analyst package was used to create binary maps of habitat. This was achieved by constructing an expression equivalent to [ecological niche model for species X] > threshold value for species X. This was done for both threshold values described above for testing the performance of niche modeling in Maxent. Threshold 1 is the maximized sum of sensitivity and specificity for 'common' species and is the minimum cumulative value present in training data for 'rare' species. Threshold 2 is the 10th percentile of cumulative probability values present within the training data for both common and rare species. The output of the raster calculator is a new raster layer with a value of 1 assigned to every pixel of suitable habitat and a 0 assigned to pixels of unsuitable habitat. This was done separately for both species 1 and 2 of a given pair. Species 2 of the pair was added to itself using the raster calculator to obtain a new raster layer identical to the original except each pixel of suitable habitat was assigned a value of 2 and each unsuitable pixel retained a value of 0. To calculate the area of overlap in suitable habitat between species pairs, the doubled threshold raster layer of species 2 was added to the single threshold layer of species 1. The result is a raster layer with values of 0- unsuitable habitat, 1- suitable habitat for species 1, 2- suitable habitat for species 2, and 3- suitable habitat for both species 1 and 2. This was performed for all species pairs in the study, and the total number of pixels (area in km²) of species 1, total number of pixels of species 2, and number of shared pixels was recorded. Reproductive

isolation was calculated using the methods described in Chapter 1 using equation RI_7 .

Reproductive isolation was calculated asymmetrically for each species as:

$$RI_{species\ x} = 1 - \frac{Shared\ Area\ x + y}{(Shared\ Area\ x + y) + Unshared\ Area\ x}.$$

Reproductive isolation for both species was recorded, and an average between the two species was calculated. Isolation was estimated using both methods of threshold value definition giving a range of potential isolation depending on the conservativeness of the threshold employed.

Genetic distance and degree of overlap in ecological niche models

Estimates of sequence divergence were extracted from the phylogenetic analysis presented in Figure 2.1. Under the assumption that substitutions occur at a clock-like rate, the level of sequence divergence can be used as a proxy for the amount of time that has passed since divergence in each species pair. In order to examine if time since divergence affects the amount of overlap species pairs experience, a correlation was performed using JMP 8.0 (SAS Institute, Cary, North Carolina, USA).

Correlation between niche differentiation and ecogeographic isolation

Because the geographic consequence of differences in ecological niche can vary tremendously depending on the geographic arrangement of suitable habitat (Sobel et al. 2010), the relationship between niche differentiation and ecogeographic isolation is expected to vary. To test for this, a correlation was performed in JMP 8.0 (SAS Institute, Cary, North Carolina, USA) between the levels of niche differentiation calculated in the

Schoener's D identity test (Figures 2.3A) and the level of ecogeographic isolation calculated for each species obtained using the 10th percentile threshold (Table 2.5).

Results

Ecological niche models

Table 2.1 provides a summary of data from constructing the ecological niche models. Maxent performed reasonably well in producing models with an average AUC score of 0.95 (lowest was 0.85). For common species, threshold 1 (maximized sensitivity + specificity) correctly placed test occurrence points in suitable habitat an average of 89.7% (± 6.5), and binomial probabilities at this threshold departed from the null expectation for all species ($p < 10^{-11}$). At threshold 2 (10th percentile of training data), 75.6% (± 19.4) of test points were correctly placed in suitable habitat and binomial probabilities at this threshold were better than null as well ($p < 10^{-11}$). For rare species, threshold 1 (minimum training presence) predicted the test point accurately 86.9% (± 9.3), and Pearson et al's (2007) method of comparing this prediction rate to null resulted in $p < 0.0001$ in all cases. Threshold 2 performed slightly better on average with a success rate of 82.7% (± 7.2), and $p < 0.0001$ in all cases as well. Figure 2.2A-X shows the continuous cumulative probability maps re-projected onto the spatial extent of the study area. The shading relates to the values of the cumulative probability distribution with lighter shades and white representing the highest-ranking suitable habitat. Figures are presented in same order as Table 2.1 such that species pairs are adjacent.

Ecological niche models were predicted by Maxent with high confidence (Table 2.1), indicating that the environmental variables included in this study have significant

impacts on the geographic ranges of species in the genus *Mimulus*. The environmental variables that were most important in building ecological niche models varied considerably across species (see circles in Figure 2.2A-L). In general, precipitation layers (BIO12, 15, 16, 17) had higher relative contributions than temperature with an average of 12.5% (± 13.9 , range 0, 64.7) averaged across all four layers, versus 5.6% (± 9.6 , range 0, 67.2) for the temperature layers (BIO1, 4, 5, 6). Of the continuous climatic variables, BIO12 (mean annual precipitation) had the highest average contribution with 15.4% (± 15.89 , range 0.01-51.2). The categorical geology variable had the highest relative contribution with an average of 27.7% (± 13.8 , range 10, 57.2). It was the most important variable in 8 of the 24 species and was the second most important variable in 10 additional species.

Comparison of niche models of species pairs

MANOVA analysis revealed that environmental variables differed between species pairs (Table 2.4), indicating that substantial environmental differences between species. Figure 2.2A-L illustrates these differences and provides the results of post-hoc Mann-Whitney U (for continuous WORLDCLIM variables) and contingency table analysis (for categorical geology variable). While the analysis in the preceding section provides an indication of which variables were most important to building individual niche models, these analyses allow the identification of which layers differ significantly between species pairs. As with the relative contributions of layers presented above, the variables that are different between species exhibited considerable variation; however, all

species exhibited some niche differentiation with at least two environmental variables significantly different in all cases.

Identity tests performed within ENMTools (Warren et al. 2008) corroborated the pattern of strong niche divergence. In the analysis of Schoener's D, the actual measured value of D for each species pair was significantly different from the null distribution generated from randomizing species identities ($p < 0.05$ in all cases) (Figures 2.3A). In the analysis of the *I* niche identity statistic (Warren et al. 2008), all but one species pair also showed niches were less similar than expected under the null (Figures 2.3B).

Substantial niche differentiation was detected between recently diverged species pairs in the genus, providing a link between speciation and ecological divergence that has been argued to be common in nature (Schluter 2001; Sobel et al. 2010). Niche divergence was detected by the MANOVA analysis in all species pairs (Table 2.4), and identity tests (Warren et al. 2008) indicated that species pairs had significantly different niches with the exception of species pair *Mimulus cusickii* / *nanus* in the *I* statistic test (Figures 2.3 A&B). However, *M. cusickii* is a rare species with only 7 collection records used from the Jepson Online Interchange, and this lack of sample size most likely reduced the ability of the test to detect differences.

Reproductive isolation

Given that the above analyses showed significant divergence in ecological niche for closely related species pairs in *Mimulus*, the geographic consequence of these niche differences gives an estimate of the degree of ecogeographic reproductive isolation between these species. Figure 2.5A-L illustrates the amount of ecogeographic isolation

experienced between species pairs. These maps were generated based on the threshold 2 criterion described above because it was common to analyses of both common and rare species. Estimates of reproductive isolation obtained through measuring shared and non-shared areas for each species are given in Table 2.5, with values obtained using both threshold 1 and 2 presented. Results are given both asymmetrically (reproductive isolation experienced by species 2 with respect to species 2) and as an average composite of both species asymmetric scores. Reproductive isolation was quite strong in many cases with an average value of 0.64 to 0.69 depending on which threshold was used.

Genetic distance and degree of overlap in ecological niche models

Correlation analysis between ecogeographic isolation and sequence divergence showed no significant relationship (Figure 2.6; $p = 0.9188$). However, because these species pairs represent recent speciation events, the range of sequence divergence over which a relationship could be found was quite low.

Correlation between niche differentiation and ecogeographic isolation

Figure 2.7 shows the relationship between the degree of ecogeographic isolation and a measure of niche similarity, Schoener's D. Regression analysis shows a significant negative correlation between the two variables (slope = -0.946, $p=0.006$) indicating that ecogeographic isolation is highest in species with the least similar niches. However, r^2 is only 0.547, revealing that variation in the spatial arrangement of suitable habitat results in significant variation in how much isolation results from niche differences.

Discussion

Ecological niche modeling provides support for divergence in niches

Strong levels of niche divergence between close relatives is consistent with previous studies using similar methodologies (e.g. Graham et al. 2004b; Nakazato et al. 2010), though others have argued that conservation of niches is also common (Kozak and Wiens 2006; Peterson et al. 1999). The ways in which species pairs differed was quite idiosyncratic. While precipitation level variables were the most important in constructing niche models for individual species, these were not any more likely to be significantly different in the posthoc tests than the temperature variables (Figure 2.2A-L). Many of the species pairs that exhibited strong niche divergence occupy different altitudes where both temperature and precipitation vary considerably (e.g. *M. cardinalis/lewisii*, *M. angustatus/pulchellus*, *M. constrictus/whitneyi*). As an example, *Mimulus cardinalis* and *M. lewisii* were significantly different ($p < 0.0001$) at every environmental variable used to construct the ecological niche models (Figure 2.2F), exhibiting one of the highest levels of niche divergence in the species pairs represented in this study. This corroborates extensive work into the nature of niche differences in this species pair using demographic (Angert 2006), reciprocal transplant (Angert and Schemske 2005), and experimental evolution approaches (Angert et al. 2008). Alternatively, very few species showed evidence of weak niche divergence. In one example, *Mimulus gracilipes* and *M. palmeri* only differed significantly for two variables related to precipitation (BIO12: average annual precip. and BIO16: precip. of wettest quarter). However, as long as any aspects of niche are significantly different, a geographic consequence may be associated leading to reproductive isolation.

Correspondence between ecological niche models and intrinsic biological differences

While indicative of biological differences, results from ecological niche modeling should be interpreted with caution (Guisan and Thuiller 2005). Differences among environmental variables at collection sites only indicate that environmental conditions are different in the places in which a species currently lives. For a given species, regions identified as suitable habitat may indeed be used to predict other regions with similar environmental conditions, as in predicting invasive species potential (e.g. Peterson 2003). However, for regions identified as unsuitable through ecological niche modeling, we only know that the species does not live in that type of habitat, not necessarily that it isn't able to. This problem is especially acute for uncommon species with highly restricted ranges. Species with restricted ranges tend to have niche models with very tight parameter values, exhibiting only a very small range of potential habitat (e.g. *Mimulus filicaulis*, Figure 2.2F). This may result from very tight ecological requirements of these species, or they may have small geographic ranges for other reasons, such as purely historical factors (Sexton et al. 2009).

Even when geographic ranges do accurately represent the intrinsic tolerances of a species, it is not necessarily true that all environmental variables that appear important in niche modeling actually are. A species may actually be geographically restricted by only a subset of the variables that appear important from niche modeling. For example, *Mimulus norrisii* occurs only on limestone outcrops of the southern Sierra Nevada. This edaphic endemic has a highly restricted range based upon the occurrence of limestone outcrops in this region. The ecological niche modeling (Figure 2.2T, Figure 2.2J) shows

that this species occurs over a very narrow range of climatic conditions, but that may simply be related to the range over which limestone outcrops exist. If limestone occurred across a wider range of climatic variation, *M. norrisii* may or may not occupy a larger geographic and environmental range.

In order to substantiate the findings of these ecological niche models, reciprocal transplants or growth chamber experiments are necessary. Because many species within the genus *Mimulus* are rare and potentially threatened (Beardsley et al. 2004), reciprocal transplants may be difficult or impossible in the majority of species pairs presented. Angert & Schemske (2005) performed reciprocal transplants between habitats with *Mimulus cardinalis* and *M. lewisii*, finding that each species performed considerably better in its own range than in the reciprocal environment. This indicates that the high levels of differentiation seen in the niche modeling presented (Figure 2.2K & L, Figure 2.2F) are based on intrinsic biological differences. However, in order to decouple which environmental variables are responsible, it would be necessary to vary each independently in a growth chamber experiment.

Ecogeographic reproductive isolation

Species can occupy a distribution both because of ecological and historical and historical factors (Endler 1982; Thorpe et al. 2008), and decoupling these two factors has been a major impediment to incorporating estimates of geographic isolation into reproductive isolation measures (Sobel et al. 2010). An example of this can be seen in the ecological niche model of *Mimulus johnstonii*. *M. johnstonii* occurs over a relatively small geographic area primarily within the Transverse Range of southern California.

Ecological niche modeling reveals that conditions similar to its home range also exist in the southern Sierra Nevada and northern Coast ranges, but it does not inhabit these areas most likely due to historic reasons (Figure 2.2J). This illustrates that the ecological niche modeling process does not simply resolve where a species does live, but also shows where a species could live. In this way, comparing overlap in niche models is more akin to examining intrinsic biological differences than simply calculating the overlap in species ranges currently occupied (Sobel et al. 2010). This allows a much more accurate depiction of the isolation that is both potentially and actually acting, making these estimates relevant under the biological species concept (Mayr 1942).

Few estimates of ecogeographic isolation have been made, even though it is recognized as a potentially important form of reproductive isolation (Coyne and Orr 2004), perhaps due to a lack of methods needed to decouple historical and ecological processes. Ramsey et al. (2003) estimated ecogeographic isolation in *Mimulus cardinalis* and *M. lewisii* at 0.587. While this is considerably lower than the estimates provided here (0.906 – 0.970, Table 2.5), the previous method was much coarser, making a conservative estimate at best. As stressed by Ramsey et al. (2003) the sequential nature of reproductive barriers gives ecogeographic isolation enormous potential in preventing gene flow, as it is the first barrier to act. Within this linear sequential series methodology for calculating reproductive isolation, later acting barriers can only prevent gene flow that has not already been stopped by previously acting barriers. Because ecogeographic isolation gets the opportunity to act first, it only requires a value of 0.5 to ensure that no other barrier could possibly prevent more gene flow. This is true in 8 out of the 12 species pairs analyzed using the threshold 1 criterion, and for 11 out 12 species pairs

using threshold 2 (Table 2.5). And in total, the average ecogeographic isolation among all species pairs is above this level (0.64 to 0.69). This suggests that ecogeographic isolation must account for a sizeable proportion of the total isolation experienced by these recently diverged species, and therefore, it must be considered one of the most important barriers preventing gene flow in nature in this genus.

Improving the accuracy of ecogeographic isolation estimates

One of the biggest limitations to ecological niche modeling at present is the resolution of environmental variables available. The WORLDCLIM climatic layers employed in this study provide 1km² resolution, but many features relevant to species distributions vary at smaller spatial scales. The geology layer presents particular issues for this study in giving a more conservative measure of niche boundaries than is really experienced. For example, both *Mimulus norrisii* and *M. rupicola* are limestone endemics that occur in different regions (*M. norrisii* in the southern Sierra and *M. rupicola* in limestone cliffs around Death Valley). However, limestone often occurs as relatively small irregular outcrops surrounded by different geologic formations. For this reason, many of the occurrence points for these two species were in pixels that were classified as non-limestone areas even though all collections of these species were taken from limestone formations. If finer resolution geologic variable layers had been used, the ecological niche maps presented in Figures 2.2T and X would show far less suitable habitat, and instead only pixels with limestone present would be potential habitat for either species. As a result, estimates of ecogeographic isolation would be considerably higher than reported.

The issue of low resolution is likely to be most apparent in study regions with abrupt changes in elevation such as the Sierra Nevada. In fluctuating topography, climatic variables may vary considerably over a single square kilometer, making it possible for species occurrence records to be found in what appears to be unsuitable habitat due to pockets of microspatial climatic differences. Increased spatial resolution would also allow some assessment of microspatial habitat isolation. The overlapping threshold ecological niche models presented in Figures 2.5A-L show that many species have some degree of ecogeographic overlap. However, while some of these species pairs do co-occur over spatial scales of square kilometers, at microspatial scales, it is extremely uncommon to find two recently diverged species in close contact in the field (J. Sobel, pers. obs.). This suggests that even with the large values of ecogeographic isolation presented here, differences in ecological niche occupied likely results in even more reproductive isolation than reported. If finer resolution variables were to become available in the future, microspatial habitat isolation should be estimated with caution, as the average dispersal distance of organisms may be greater than the grain size at those scales.

Summary

Understanding the evolution of reproductive isolation is a primary goal in evolutionary biology, and many would agree that a measure of which forms of isolation are most important at the time of speciation is crucial to this. Given its position in the linear sequence of potential reproductive barriers, ecogeographic isolation has enormous potential to play an important role. However, ecogeographic isolation has been largely

ignored in past studies of reproductive barriers, so estimates of its strength are sorely needed. The ecological niche maps and analyses in the presented work shows that ecogeographic isolation is of great importance in the genus *Mimulus*, with the certainty that it will be the highest contributor to total isolation for many of the pairs under study. This finding challenges the historical ignorance of the strength of this barrier, and suggests that additional studies are needed to assess the ubiquity of this pattern in other organisms, and should direct future studies of the traits and genetics underlying isolation in the future.

Table 2.1. Summary of *Mimulus* species pairs used and methods employed for ecological niche modeling.

Species	Maxent model construction method	N	Test data AUC	Threshold 1 (Success Rate)	Threshold 2 (Success Rate)
<i>M. androsaceus</i>	50/50 training/testing split	42	0.894	17.813 (90.5)	46.560 (66.7)
<i>M. shevockii</i>	Jackknife minus 1 cross validation	16	0.993	8.539 (93.8)	46.767 (93.8)
<i>M. angustatus</i>	50/50 training/testing split	38	0.977	6.242 (100)	17.543 (78.9)
<i>M. pulchellus</i>	50/50 training/testing split	37	0.994	29.007 (94.4)	36.671 (88.9)
<i>M. bicolor</i>	50/50 training/testing split	140	0.951	11.968 (87.1)	24.183 (72.9)
<i>M. filicaulis</i>	Jackknife minus 1 cross validation	22	0.995	9.479 (95.5)	32.815 (81.8)
<i>M. bigelovii</i>	50/50 training/testing split	162	0.900	23.492 (82.7)	21.949 (82.7)
<i>M. bolanderi</i>	50/50 training/testing split	88	0.896	12.252 (79.5)	11.571 (79.5)
<i>M. brevipes</i>	50/50 training/testing split	223	0.965	14.114 (94.5)	24.646 (81.1)
<i>M. johnstonii</i>	50/50 training/testing split	62	0.958	27.434 (87.1)	34.906 (80.6)
<i>M. cardinalis</i>	50/50 training/testing split	234	0.890	30.819 (77.8)	15.277 (86.3)
<i>M. lewisii</i>	50/50 training/testing split	85	0.952	11.902 (95.2)	23.537 (76.2)
<i>M. constrictus</i>	50/50 training/testing split	74	0.985	8.916 (97.3)	31.190 (78.4)
<i>M. whitneyi</i>	50/50 training/testing split	68	0.987	9.119 (97.1)	33.131 (73.5)
<i>M. cusickii</i>	Jackknife minus 1 cross validation	7	0.979	21.419 (85.7)	21.419 (85.7)
<i>M. nanus</i>	50/50 training/testing split	72	0.916	15.361 (94.4)	20.491 (86.1)
<i>M. douglasii</i>	50/50 training/testing split	84	0.883	8.489 (88.1)	9.516 (83.3)
<i>M. kelloggii</i>	50/50 training/testing split	108	0.944	6.877 (92.6)	12.469 (81.5)
<i>M. floribundus</i>	50/50 training/testing split	329	0.849	18.067 (84.1)	21.417 (79.3)
<i>M. norrisii</i>	Jackknife minus 1 cross validation	7	0.967	40.445 (71.4)	40.445 (71.4)
<i>M. gracilipes</i>	Jackknife minus 1 cross validation	8	0.992	39.909 (87.5)	39.909 (90.0)
<i>M. palmeri</i>	50/50 training/testing split	98	0.929	26.694 (83.7)	23.045 (83.7)
<i>M. parryi</i>	Jackknife minus 1 cross validation	11	0.994	8.153 (90.9)	35.770 (81.8)
<i>M. rupicola</i>	Jackknife minus 1 cross validation	11	0.920	3.576 (90.9)	33.199 (81.8)

Table 2.2. Environmental layers used in ecological niche modeling.

Layer	Description	Source
Bio1	Annual mean temperature (C)	WORLDCLIM (http://worldclim.org)
Bio4	Temperature seasonality (st. dev.)	WORLDCLIM (http://worldclim.org)
Bio5	Maximum temperature of warmest month (C)	WORLDCLIM (http://worldclim.org)
Bio6	Minimum temperature of coldest month (C)	WORLDCLIM (http://worldclim.org)
Bio12	Annual mean precipitation (mm)	WORLDCLIM (http://worldclim.org)
Bio15	Precipitation seasonality (coefficient of variation)	WORLDCLIM (http://worldclim.org)
Bio16	Precipitation of wettest quarter (mm)	WORLDCLIM (http://worldclim.org)
Bio17	Precipitation of driest quarter (mm)	WORLDCLIM (http://worldclim.org)
Geology	Geologic map of California	U.S. Geological Survey (http://www.usgs.gov)

Table 2.3. Correlation coefficients among environmental layers used in ecological niche modeling.

	BIO1	BIO4	BIO5	BIO6	BIO12	BIO15	BIO16	BIO17
BIO1								
BIO4		0.3277						
BIO5			0.8852	0.8686	-0.5476	0.0314	-0.4919	-0.8358
BIO6			0.6840	-0.1478	-0.5087	-0.6618	-0.5508	-0.0542
BIO12				0.5672	-0.5989	-0.2076	-0.5725	-0.6584
BIO15					-0.2496	0.3908	-0.1716	-0.8237
BIO16						0.4179	0.8938	0.5556
BIO17							0.4960	-0.3123
								0.4729

Table 2.4. Results of one-way MANOVA analysis on environmental variables between *Mimulus* species pairs.

Species pair	Wilks' λ	Degrees of Freedom	F	p
<i>M. androsaceus/shevockii</i>	0.488	1, 49	6.417	< 0.0001
<i>M. angustatus/pulchellus</i>	0.341	1, 66	15.97	< 0.0001
<i>M. bicolor/filicaulis</i>	0.792	1, 153	5.024	< 0.0001
<i>M. bigelovii/bolanderi</i>	0.106	1, 243	256.7	< 0.0001
<i>M. brevipes/johnstonii</i>	0.390	1, 276	53.90	< 0.0001
<i>M. cardinalis/lewisii</i>	0.296	1, 310	92.18	< 0.0001
<i>M. constrictus/whitneyi</i>	0.493	1, 134	17.24	< 0.0001
<i>M. cusickii/nanus</i>	0.552	1, 70	7.10	< 0.0001
<i>M. douglasii/kelloggii</i>	0.636	1, 183	13.066	< 0.0001
<i>M. floribundus/norrisii</i>	0.933	1, 327	2.9352	0.0035
<i>M. gracilipes/palmeri</i>	0.414	1, 98	17.317	< 0.0001
<i>M. parryi/rupicola</i>	0.038	1, 13	40.92	<0.0001

Table 2.5. Ecogeographic reproductive isolation between species pairs in *Mimulus*.

Species	Threshold 1				Threshold 2			
	Total area (km ²)	Shared area (km ²)	Individual RI	Average RI	Total area (km ²)	Shared area (km ²)	Individual RI	Average RI
<i>M. androsaceus</i>	139,065		0.944	0.483	45,223		0.995	0.544
<i>M. shevockii</i>	7,950	7,769	0.023		227	206	0.093	
<i>M. angustatus</i>	71,335		0.859	0.532	33,501		0.891	0.756
<i>M. pulchellus</i>	12,715	10,092	0.206		9,623	3,654	0.620	
<i>M. bicolor</i>	39,990		0.708	0.470	23,585		0.898	0.558
<i>M. filicaulis</i>	15,225	11,685	0.233		3,081	2,409	0.218	
<i>M. bigelovii</i>	100,701		0.999	0.999	105,591		0.998	0.998
<i>M. bolanderi</i>	73,596	119	0.998		76,599	163	0.998	
<i>M. brevipes</i>	51,053		0.859	0.828	30,961		0.916	0.901
<i>M. johnstonii</i>	35,420	7,184	0.927		22,805	2,599	0.886	
<i>M. cardinalis</i>	87,614		0.975	0.970	177,157		0.967	0.906
<i>M. lewisii</i>	61,563	2,147	0.904		38,512	5,928	0.846	
<i>M. constrictus</i>	34,186		0.646	0.567	10,359		0.654	0.621
<i>M. whitneyi</i>	23,645	12,100	0.488		8,717	3,589	0.588	
<i>M. cusickii</i>	27,585		0.266	0.544	27,585		0.347	0.581
<i>M. nanus</i>	112,744	20,234	0.821		97,301	18,006	0.815	
<i>M. douglasii</i>	148,866		0.551	0.389	141,866		0.631	0.418
<i>M. kelloggii</i>	86,494	66,894	0.227		65,783	52,284	0.205	
<i>M. floribundus</i>	147,050		0.989	0.658	127,805		0.988	0.672
<i>M. norrisii</i>	2,466	1,661	0.326		2,466	1,588	0.356	
<i>M. gracilipes</i>	7,775		0.501	0.721	7,775		0.305	0.618
<i>M. palmeri</i>	66,462	3,881	0.942		79,174	5,406	0.932	
<i>M. parryi</i>	17,185		0.020	0.469	3,193		0.482	0.688
<i>M. rupicola</i>	206,754	16,842	0.919		15,480	1,653	0.893	

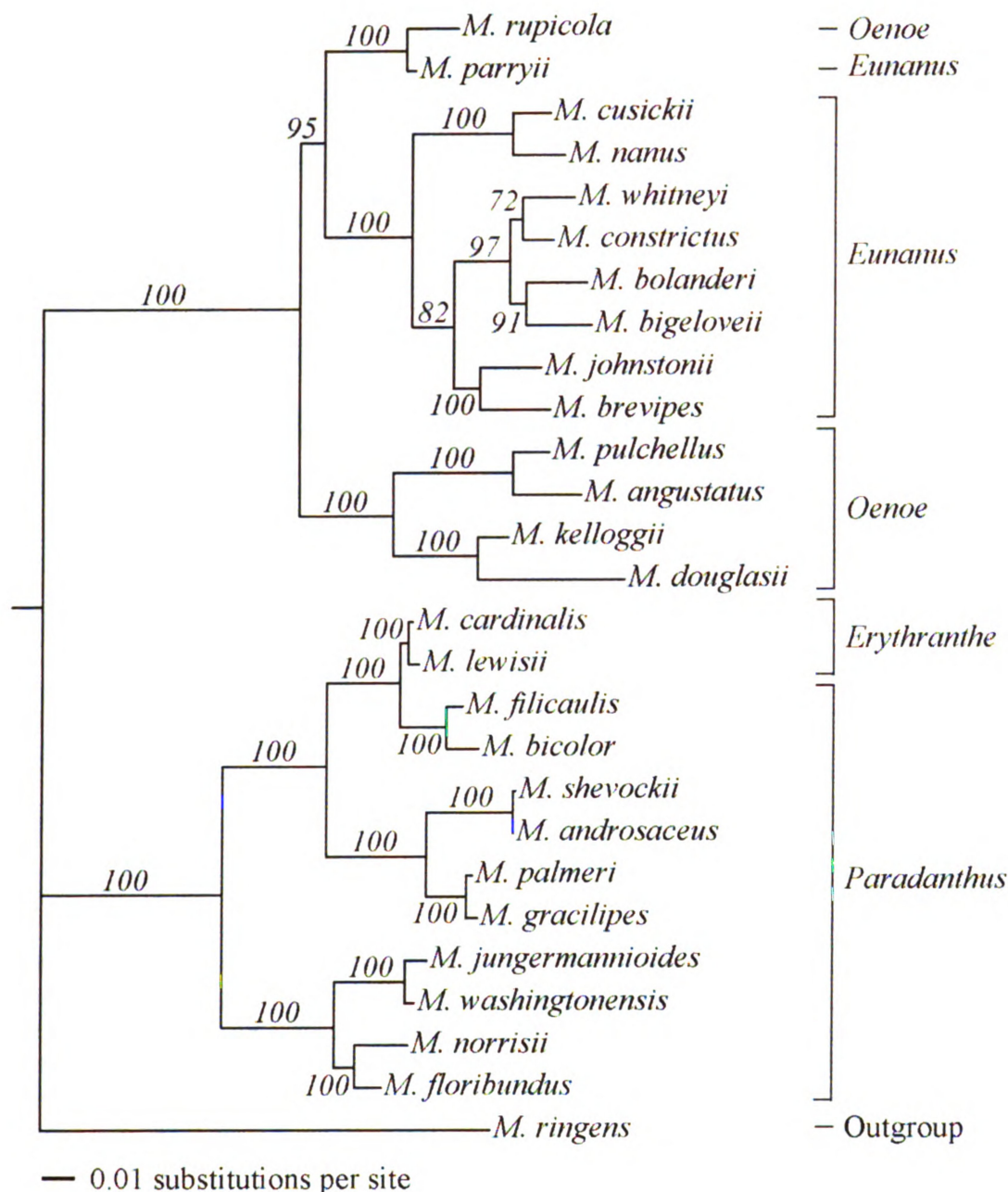


Figure 2.1. Phylogenetic relationships among *Mimulus* species in this study. Majority rule consensus tree obtained by maximum likelihood analysis in PHYLIP (Felsenstein 2005) showing relationships among *Mimulus* species (with *M. ringens* as outgroup). Sequence data (*trnL/F*, *ITS*, and *ETS*) were the same as used in Beardsley et al. (2004), trimmed to only include species in this study. Labeled brackets indicate traditional delineations from morphological taxonomy (Grant 1924). Branch lengths are proportional to average substitutions per site, and numbers on branches indicate the number of times out of 100 that the partition of species separated by that branch occurred among all trees.

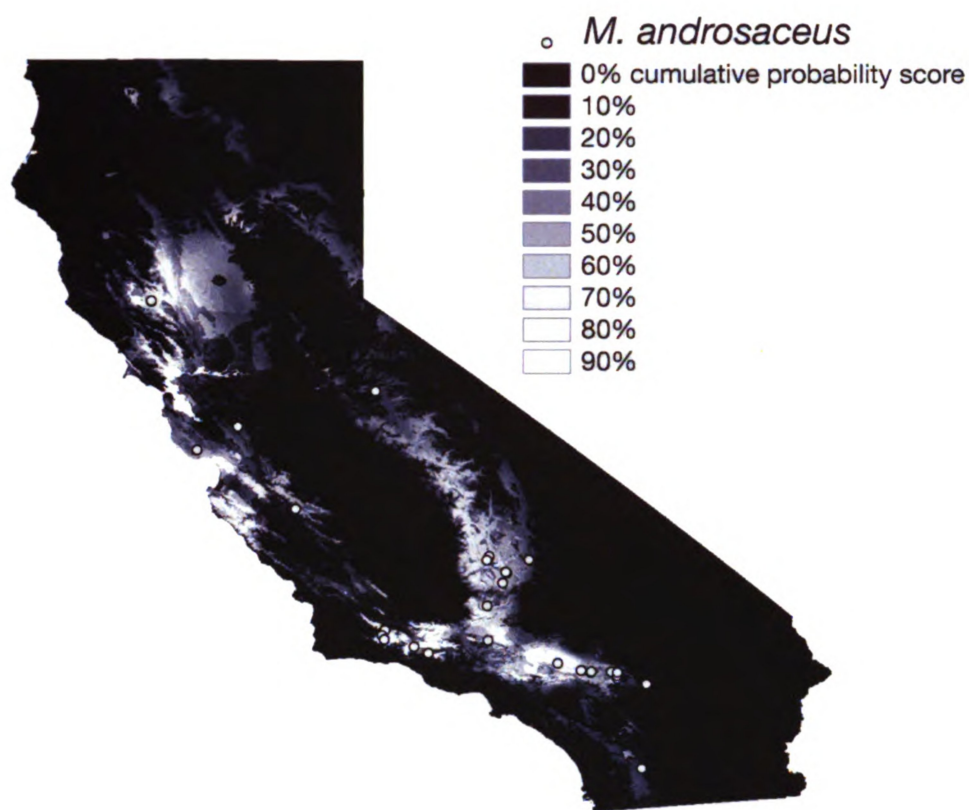


Figure 2.2A. Ecological niche model of *Mimulus androsaceus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. androsaceus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

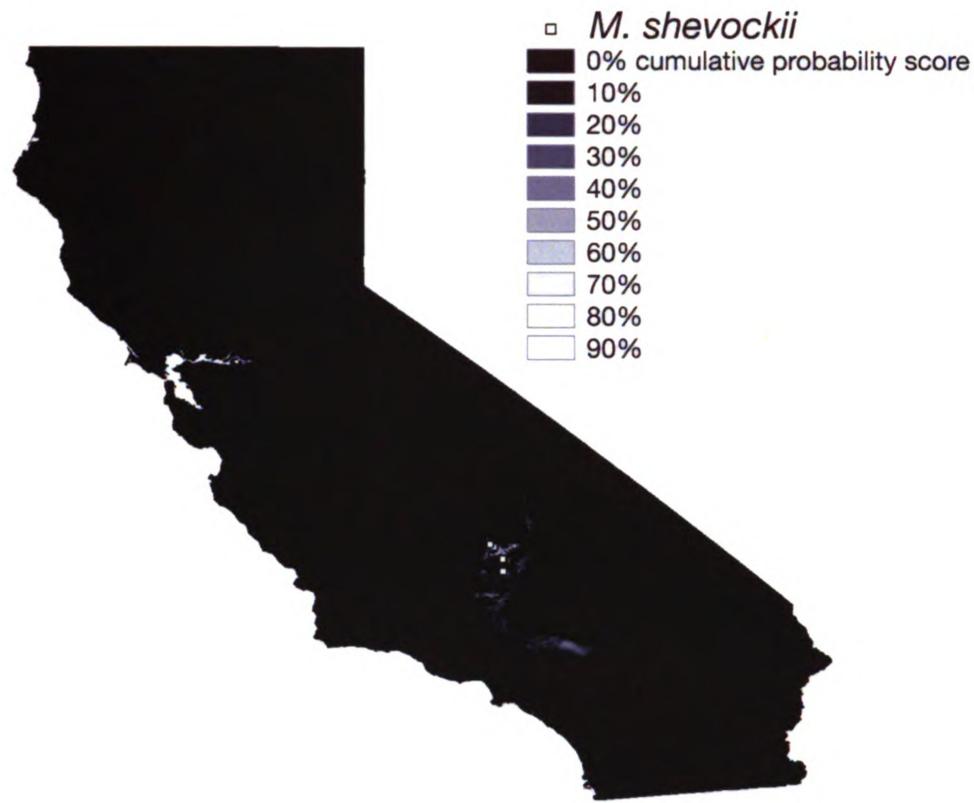


Figure 2.2B. Ecological niche model of *Mimulus shevockii*. Ecological niche model generated in Maxent using the ‘jackknife minus 1’ cross validation procedure for the species, *Mimulus shevockii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

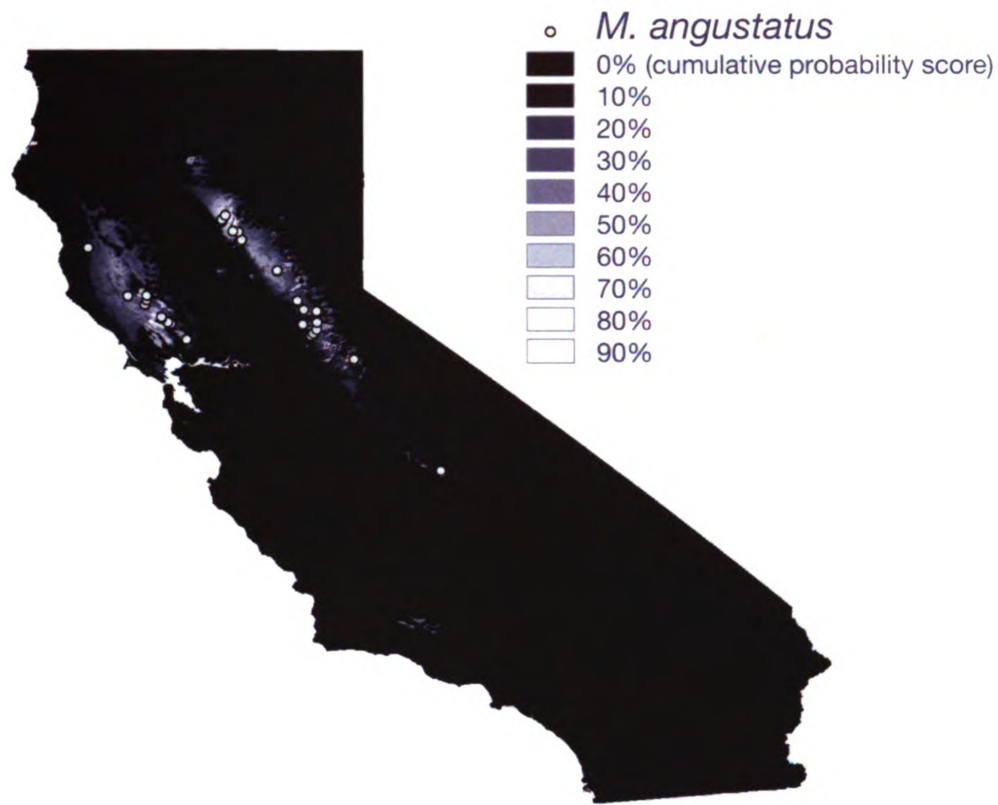


Figure 2.2C. Ecological niche model of *Mimulus angustatus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. angustatus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

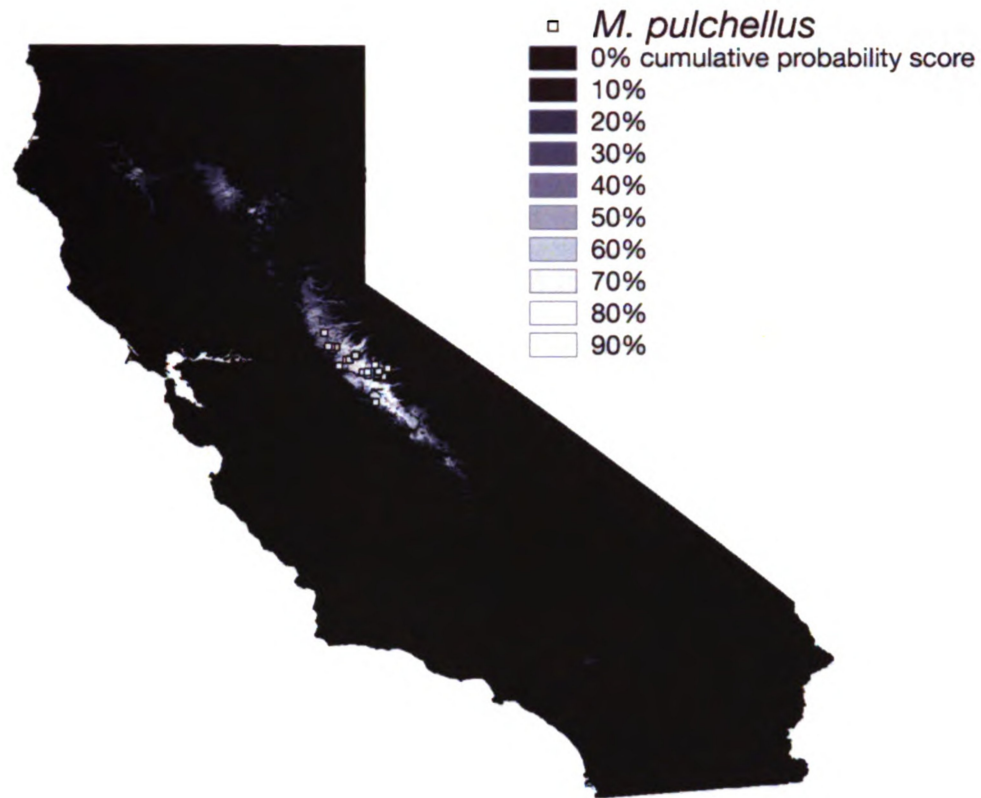


Figure 2.2D. Ecological niche model of *Mimulus pulchellus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. pulchellus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

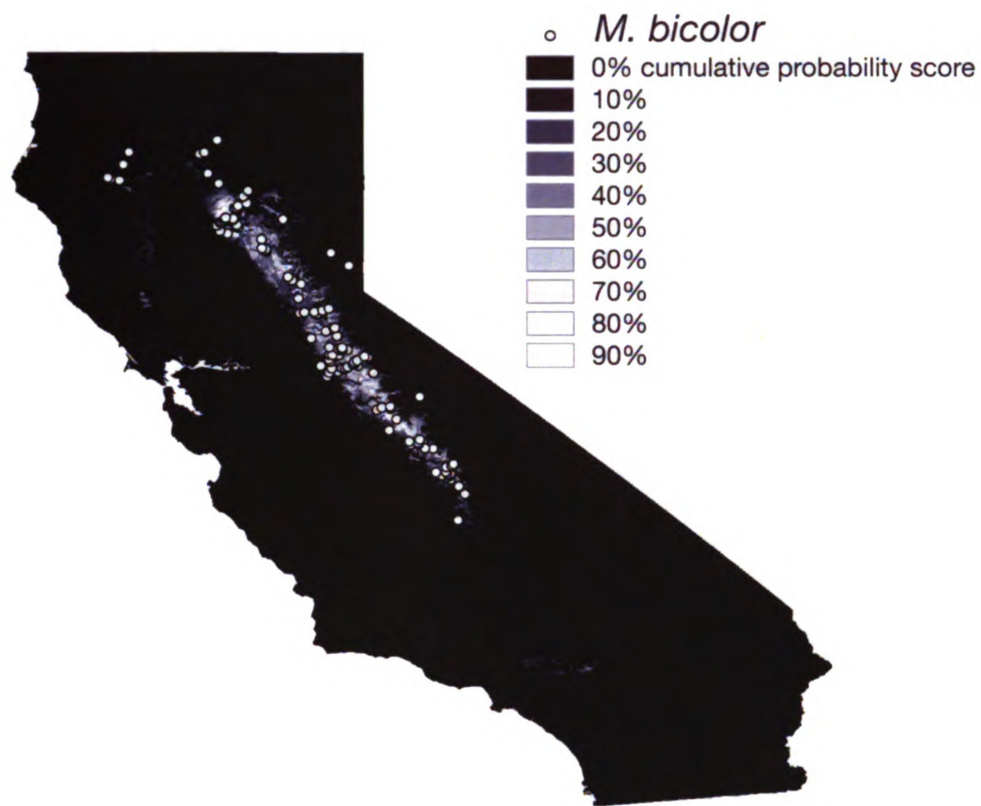


Figure 2.2E. Ecological niche model of *Mimulus bicolor*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. bicolor*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

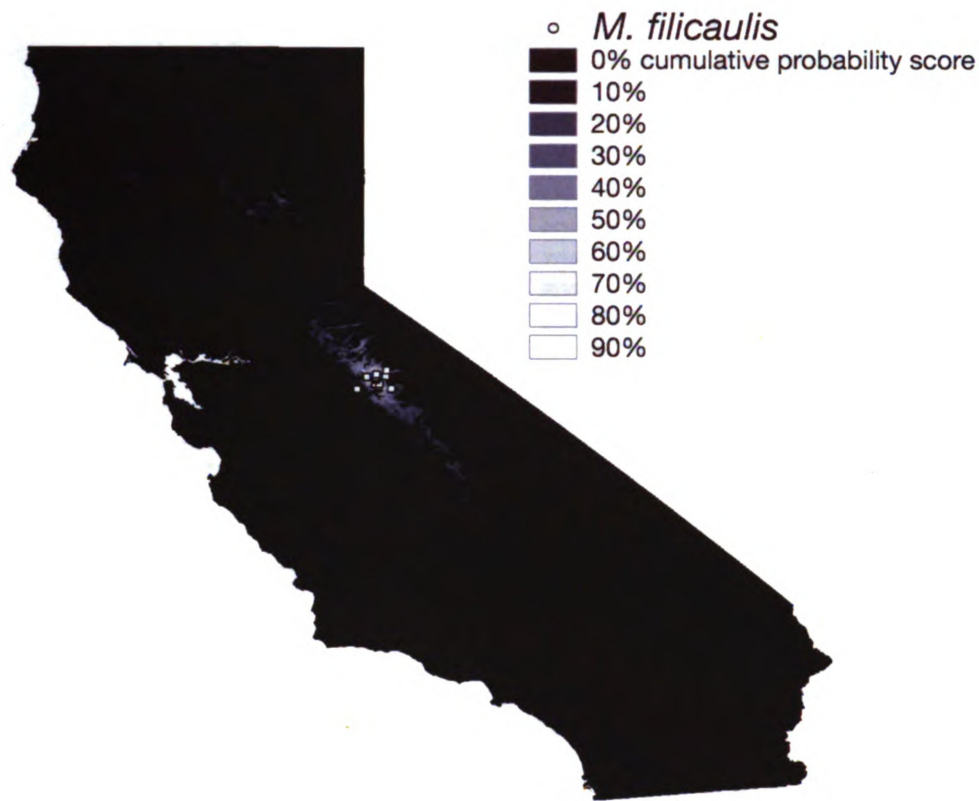


Figure 2.2F. Ecological niche model of *Mimulus filicaulis*. Ecological niche model generated in Maxent using the 'jackknife minus 1' cross validation procedure for the species, *M. filicaulis*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

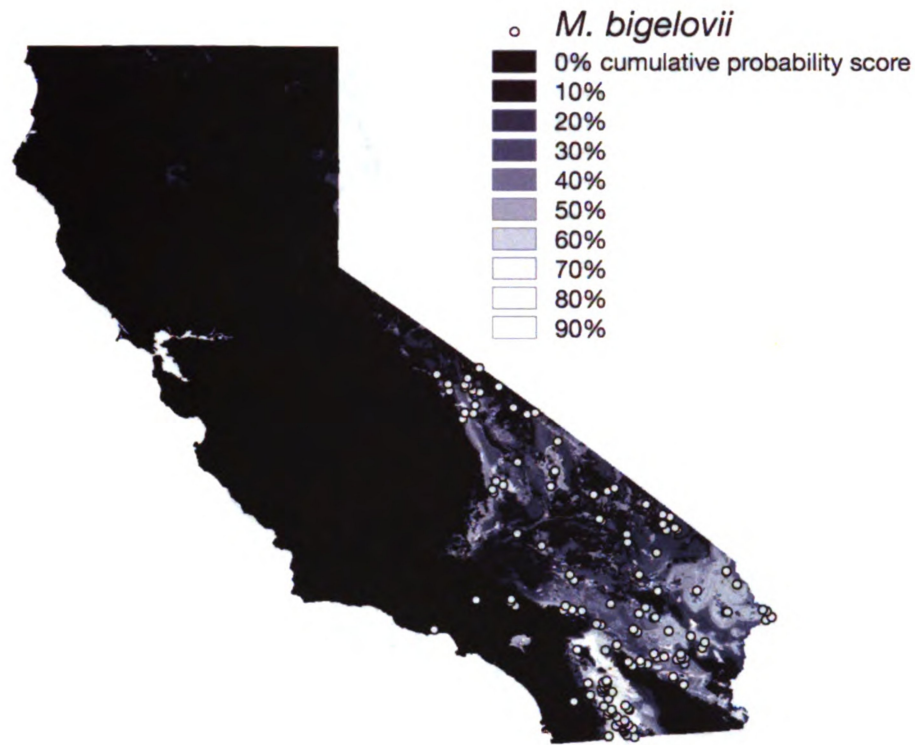


Figure 2.2G. Ecological niche model of *Mimulus bigelovii*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. bigelovii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

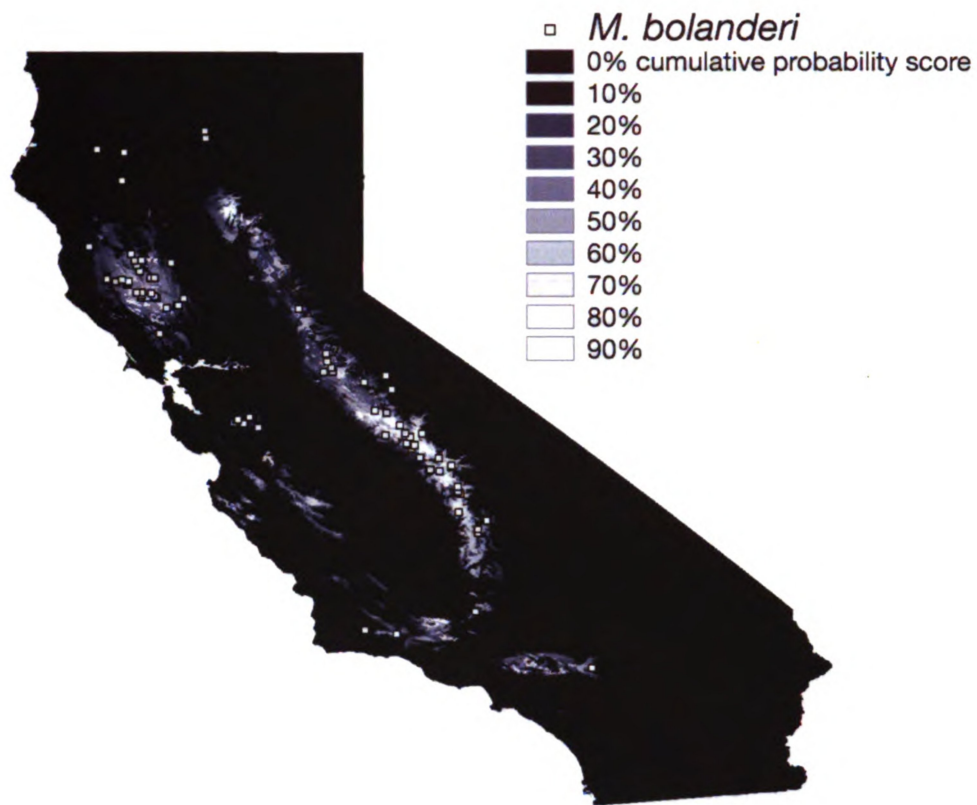


Figure 2.2H. Ecological niche model of *Mimulus bolanderi*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. bolanderi*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

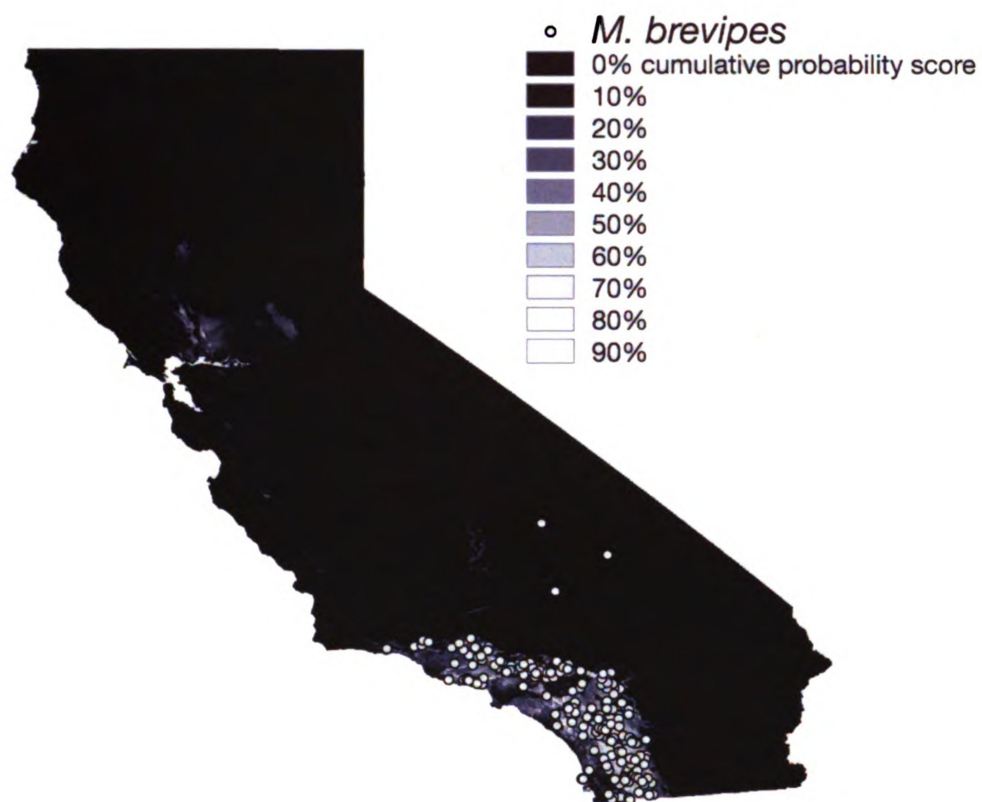


Figure 2.2I. Ecological niche model of *Mimulus brevipes*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. brevipes*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

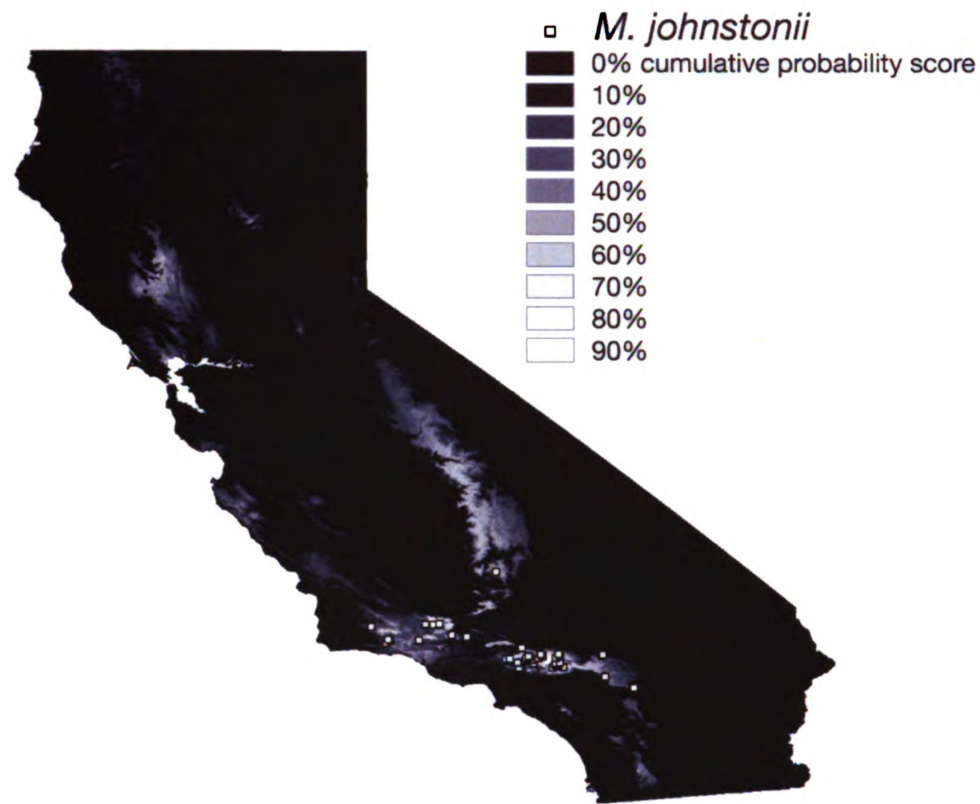


Figure 2.2J. Ecological niche model of *Mimulus johnstonii*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. johnstonii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

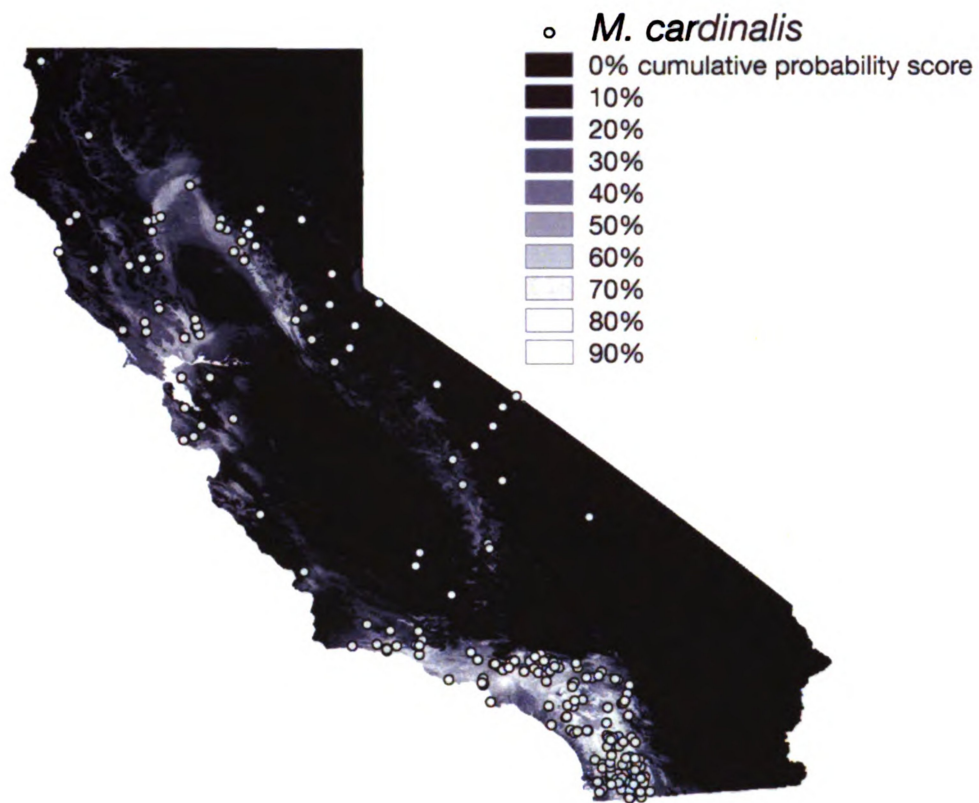


Figure 2.2K. Ecological niche model of *Mimulus cardinalis*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. cardinalis*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

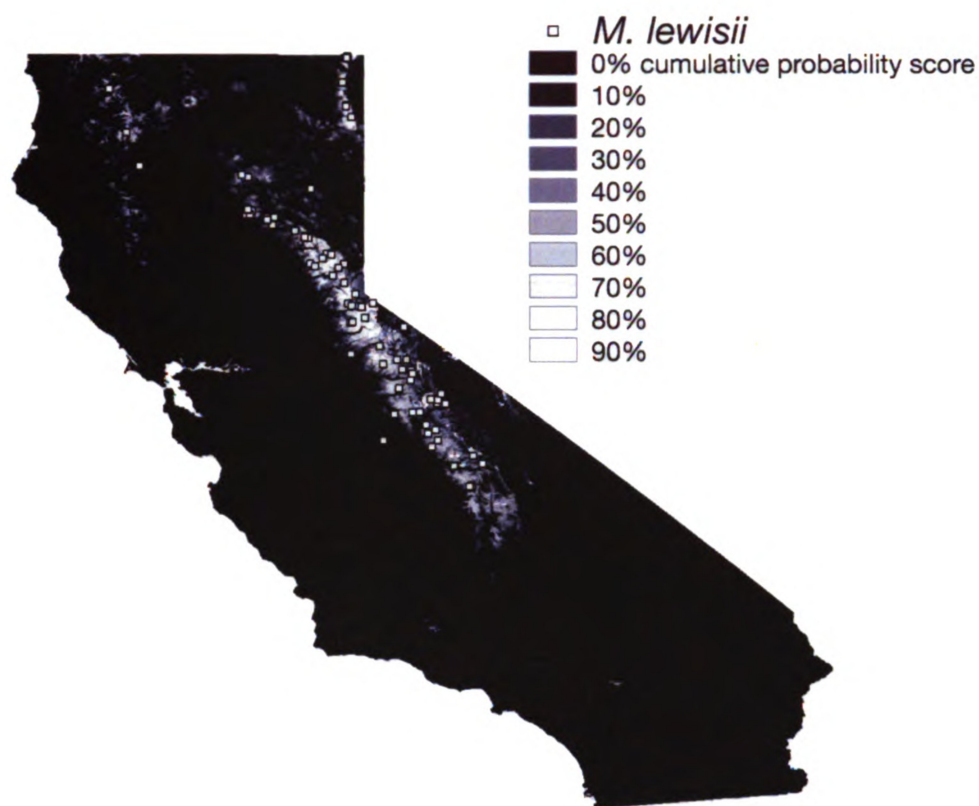


Figure 2.2L. Ecological niche model of *Mimulus lewisii*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. lewisii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

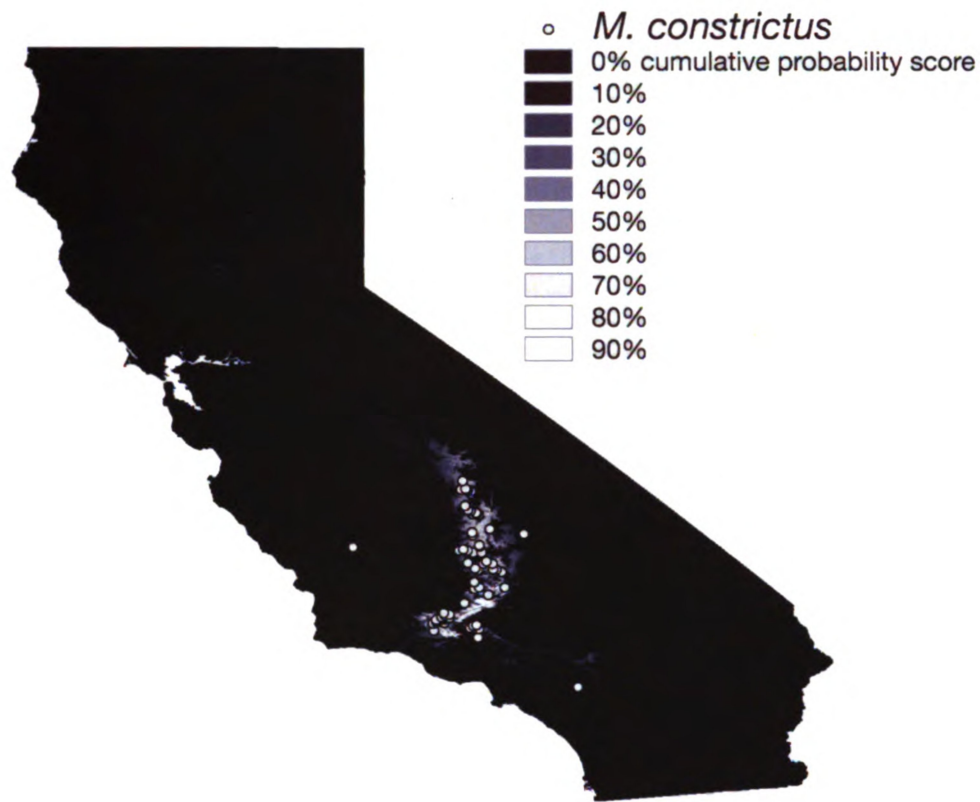


Figure 2.2M. Ecological niche model of *Mimulus constrictus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. constrictus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

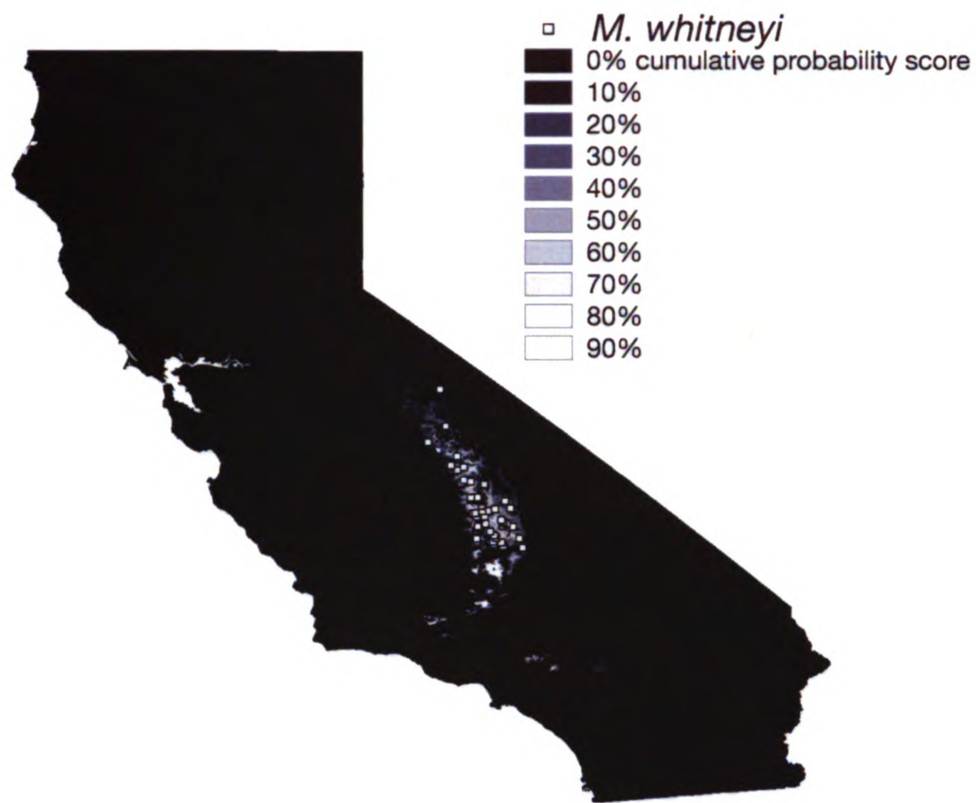


Figure 2.2N. Ecological niche model of *Mimulus whitneyi*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. whitneyi*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

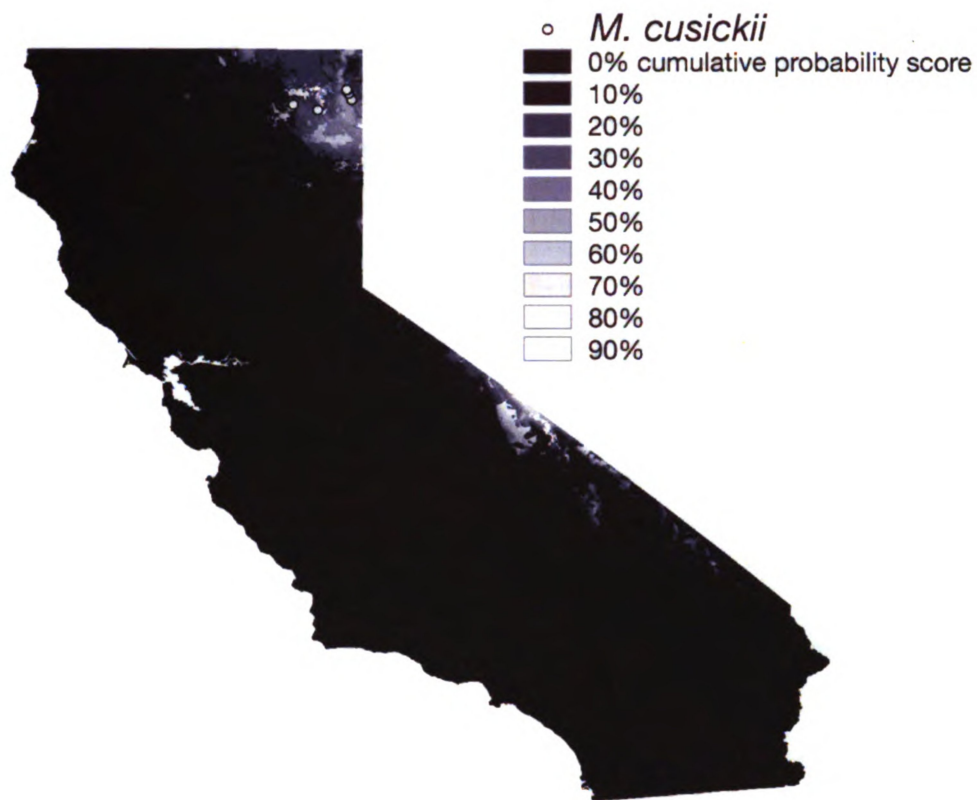


Figure 2.20. Ecological niche model of *Mimulus cusickii*. Ecological niche model generated in Maxent using the 'jackknife minus 1' cross validation procedure for the species, *M. cusickii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

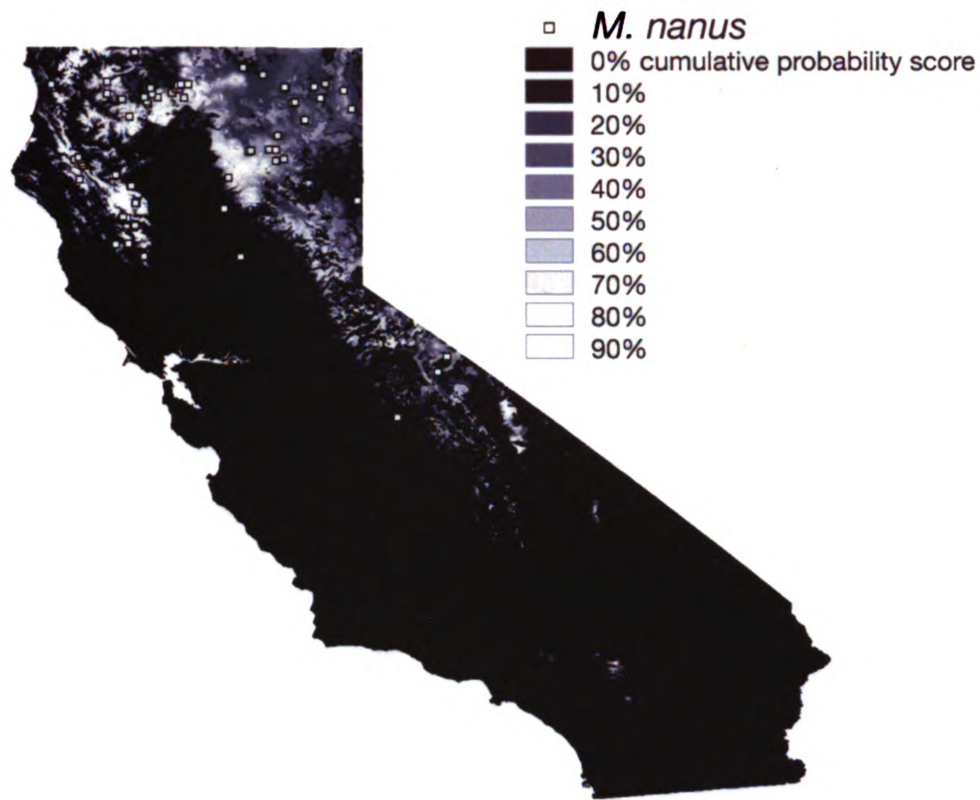


Figure 2.2P. Ecological niche model of *Mimulus nanus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. nanus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

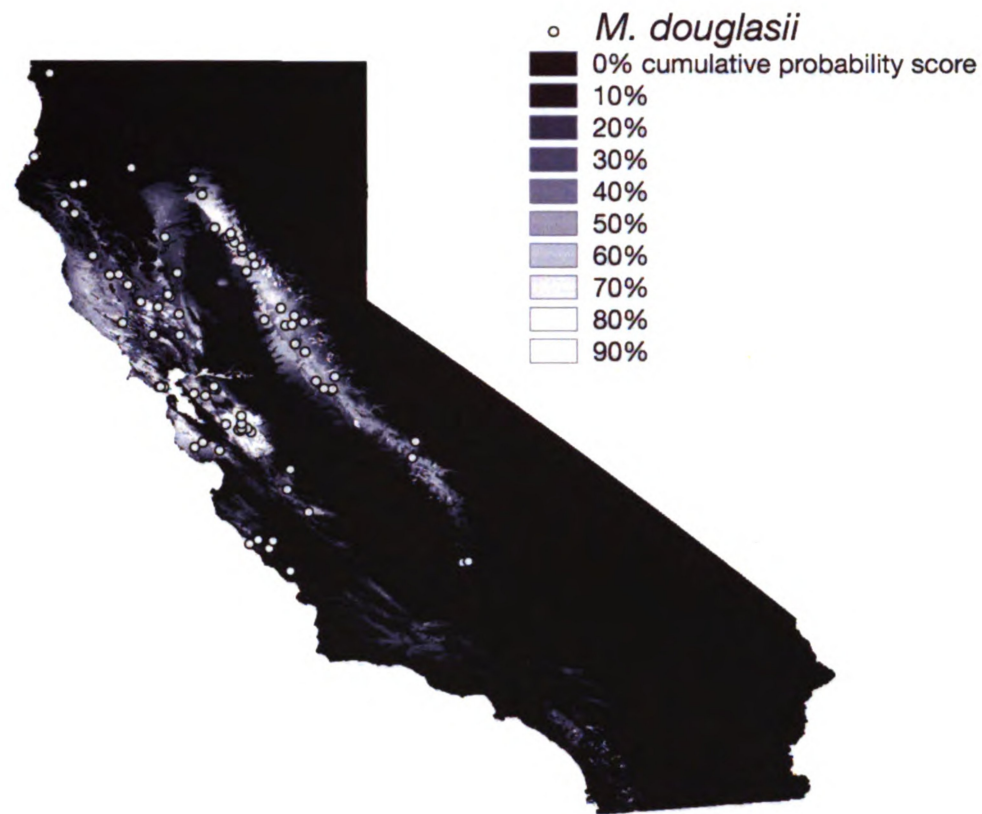


Figure 2.2Q. Ecological niche model of *Mimulus douglasii*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. douglasii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

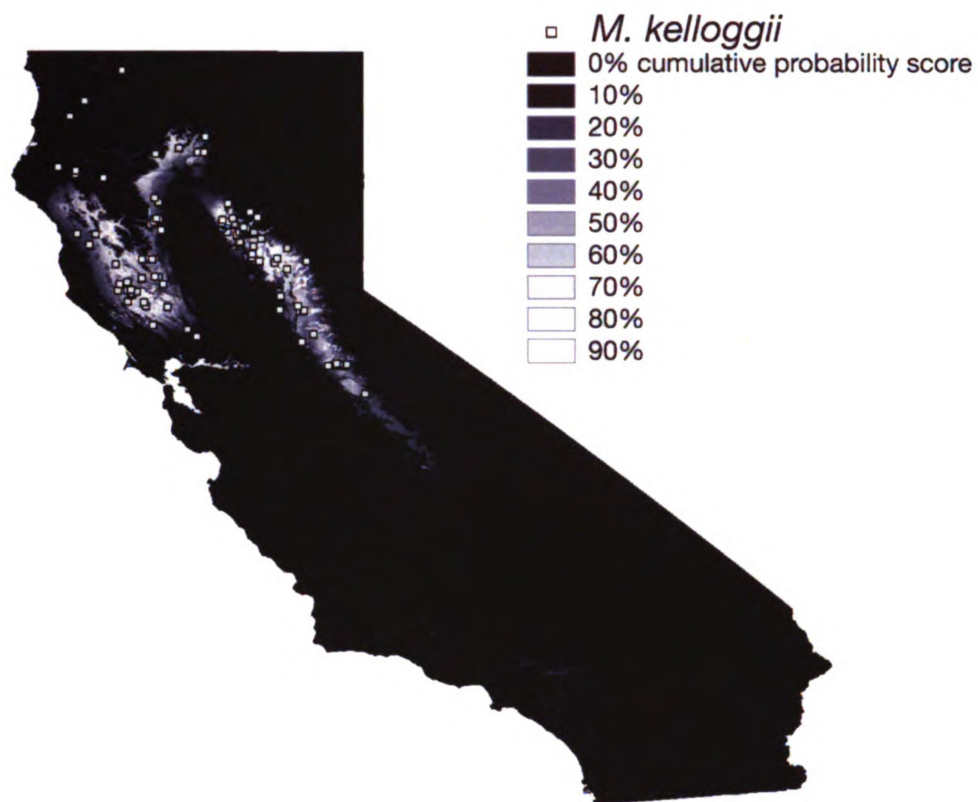


Figure 2.2R. Ecological niche model of *Mimulus kelloggii*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. kelloggii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

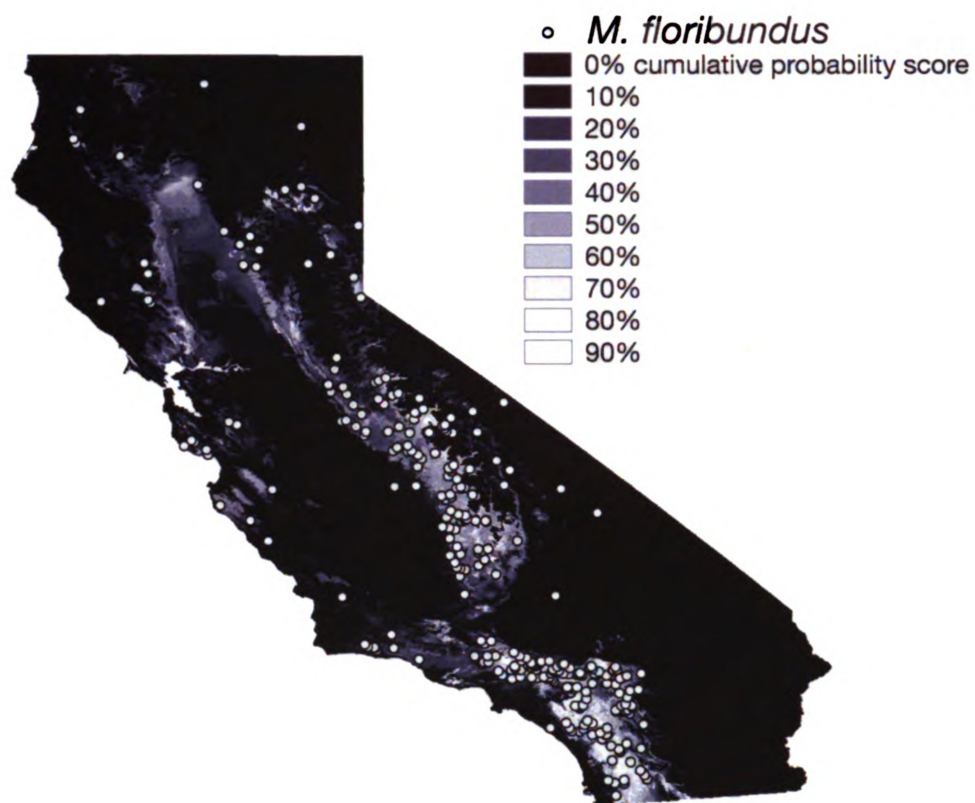


Figure 2.2S. Ecological niche model of *Mimulus floribundus*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. floribundus*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

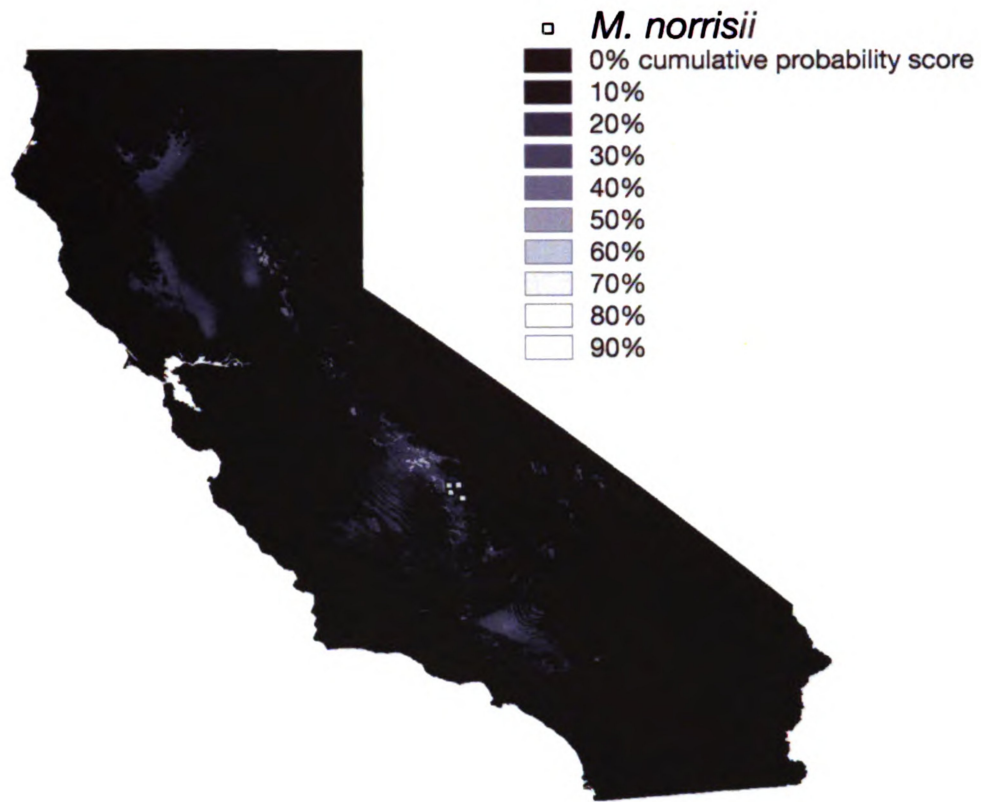


Figure 2.2T. Ecological niche model of *Mimulus norrisii*. Ecological niche model generated in Maxent using the ‘jackknife minus 1’ cross validation procedure for the species, *M. norrisii*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

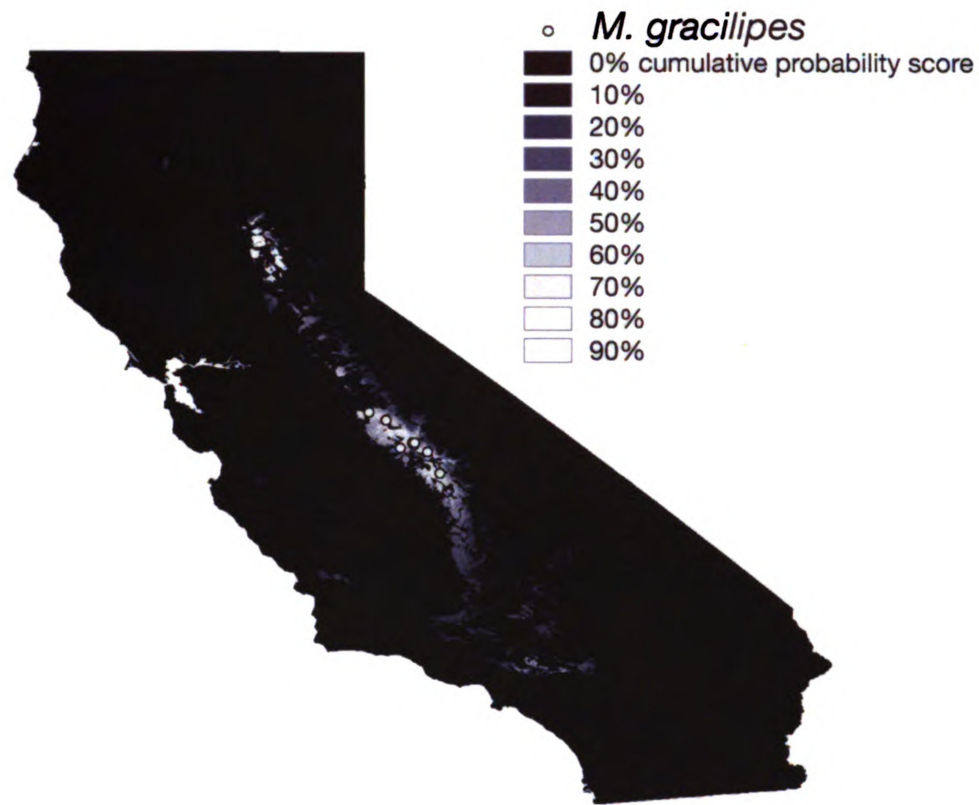


Figure 2.2U. Ecological niche model of *Mimulus gracilipes*. Ecological niche model generated in Maxent using the ‘jackknife minus 1’ cross validation procedure for the species, *M. gracilipes*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

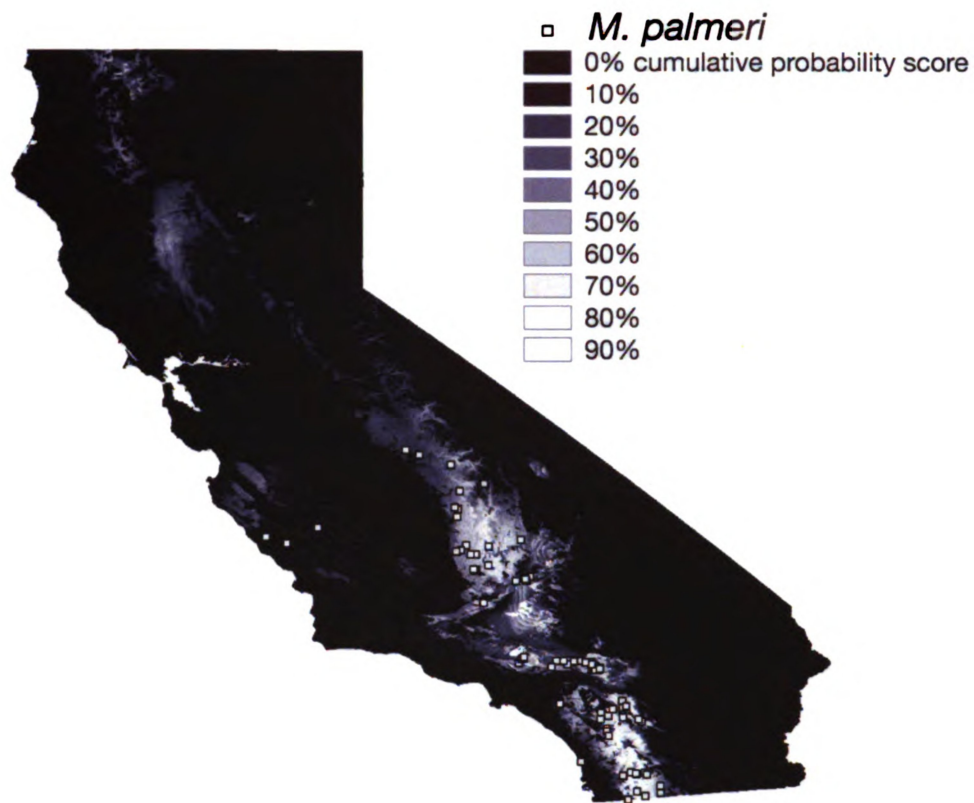


Figure 2.2V. Ecological niche model of *Mimulus palmeri*. Ecological niche model generated in Maxent using the 50% training/testing procedure for the species, *M. palmeri*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

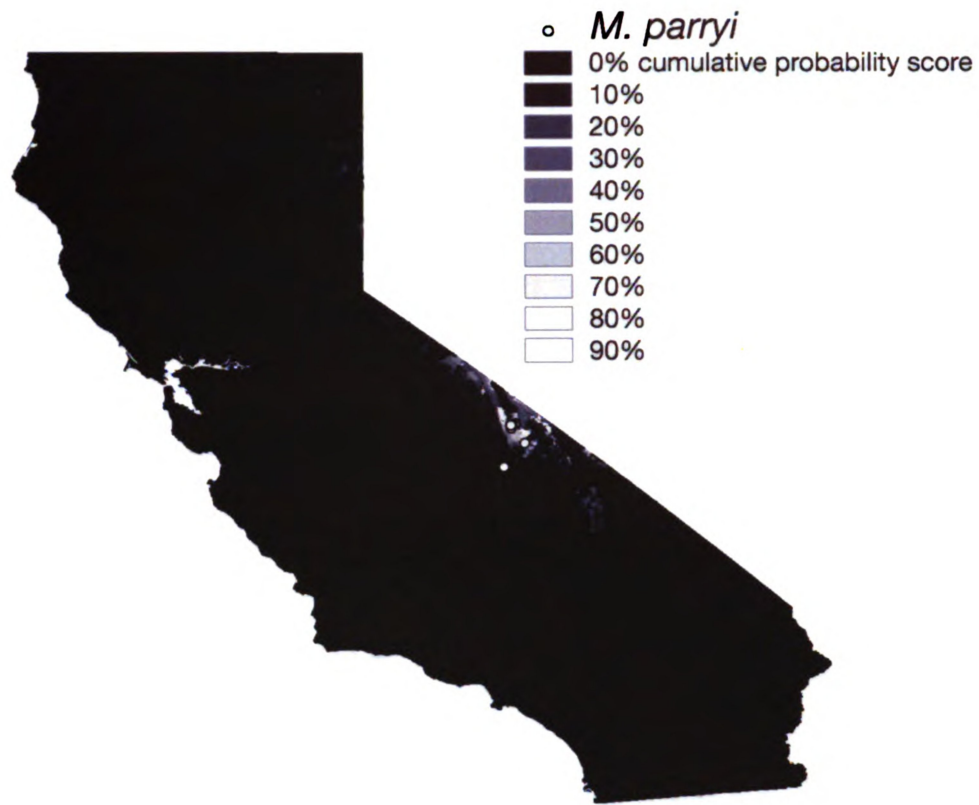


Figure 2.2W. Ecological niche model of *Mimulus parryi*. Ecological niche model generated in Maxent using the ‘jackknife minus 1’ cross validation procedure for the species, *M. parryi*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

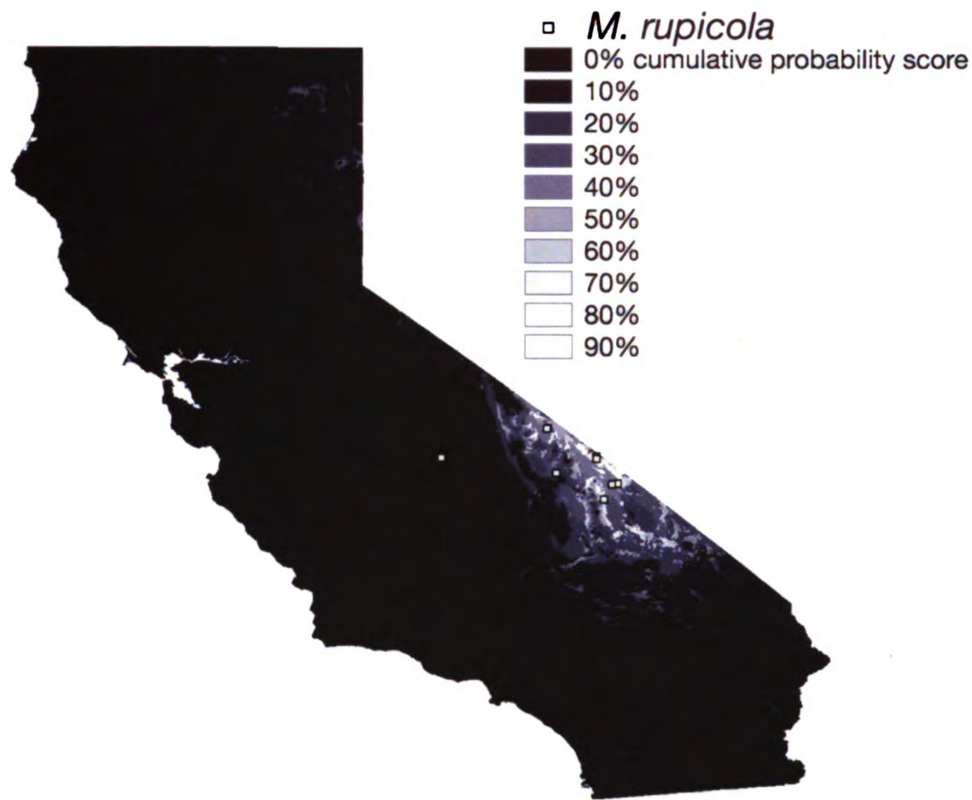


Figure 2.2X. Ecological niche model of *Mimulus rupicola*. Ecological niche model generated in Maxent using the 'jackknife minus 1' cross validation procedure for the species, *M. rupicola*. Shading corresponds to output of cumulative probability values multiplied by 100. Lighter areas indicate areas of highest habitat suitability for the species, and dots show the locations of collections used.

Figure 2.3A. Variation in environmental variables between the species pair *Mimulus androsaceus* and *M. shevockii*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

es pair *Mimulus*
 tion of each variable
 inuous variables
 values for each
 le, geology, boxes
 y within the dataset,
 ental variable are
 ed significant

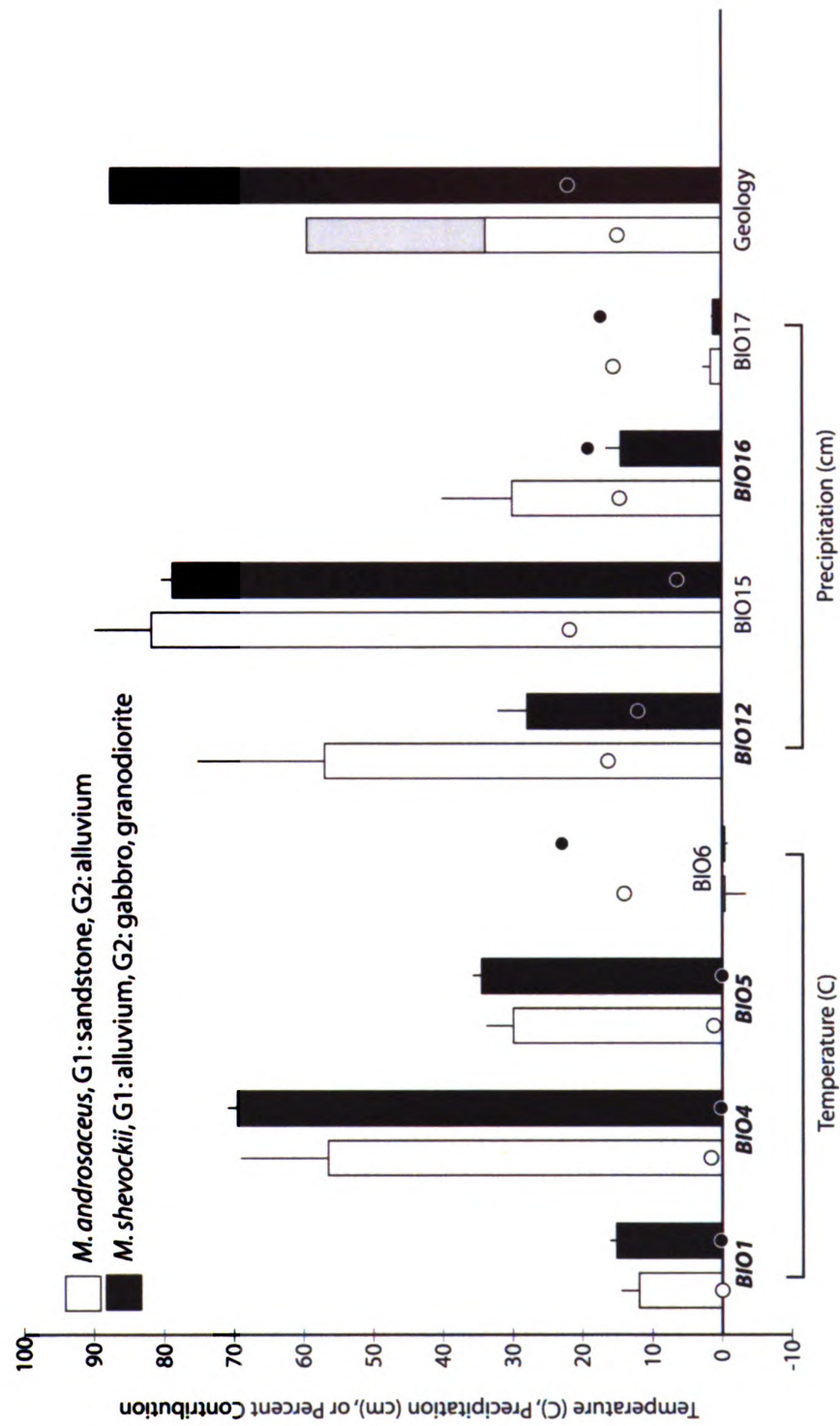


Figure 2.3B. Variation in environmental variables between the species pair *Mimulus angustatus* and *M. pulchellus*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

es pair *Mimulus*
 on of each variable
 inuous variables
 values for each
 le, geology, boxes
 y within the dataset
 ental variable are
 d significant

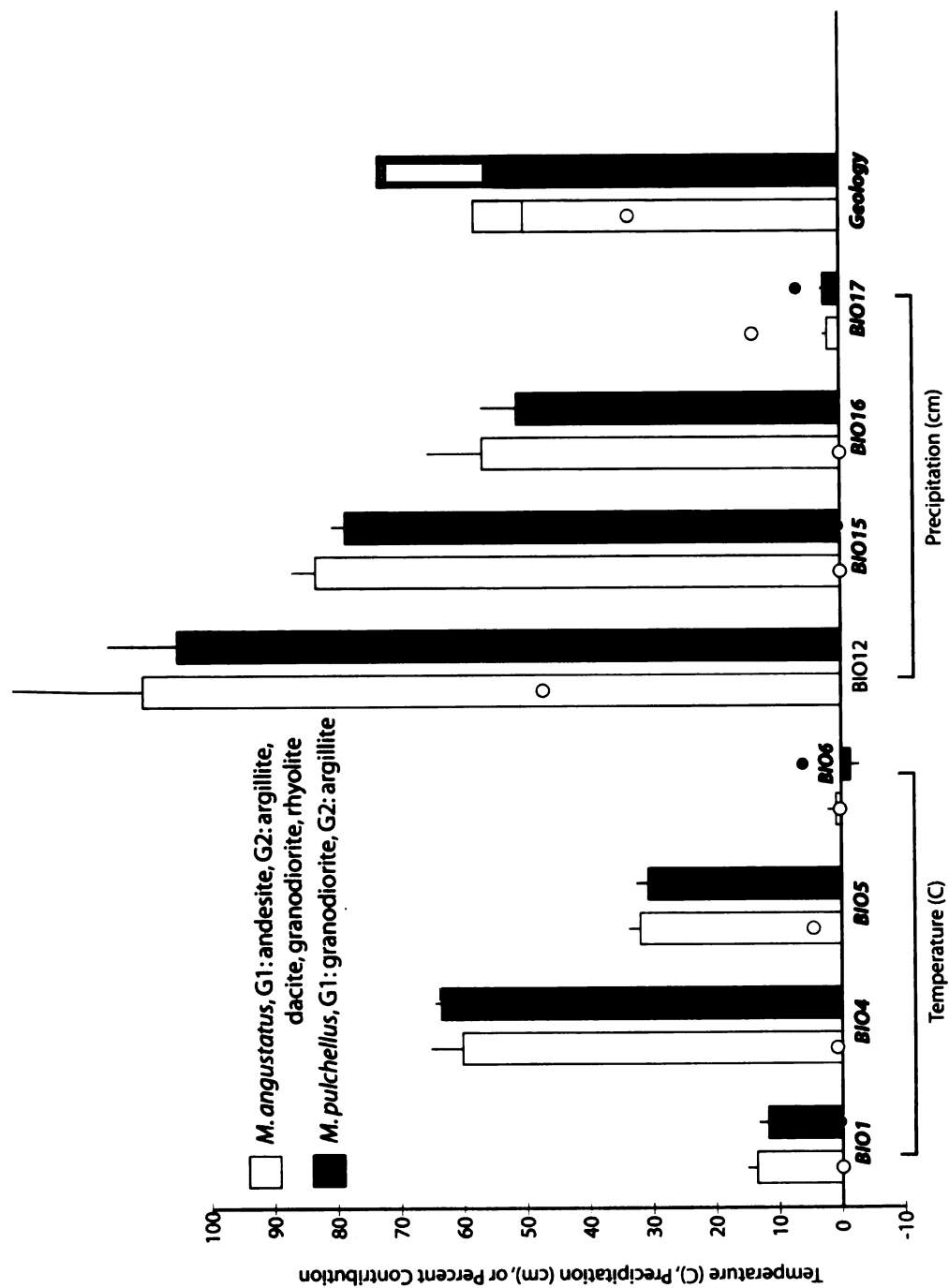


Figure 2.3C. Variation in environmental variables between the species pair *Mimulus bicolor* and *M. filicaulis*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

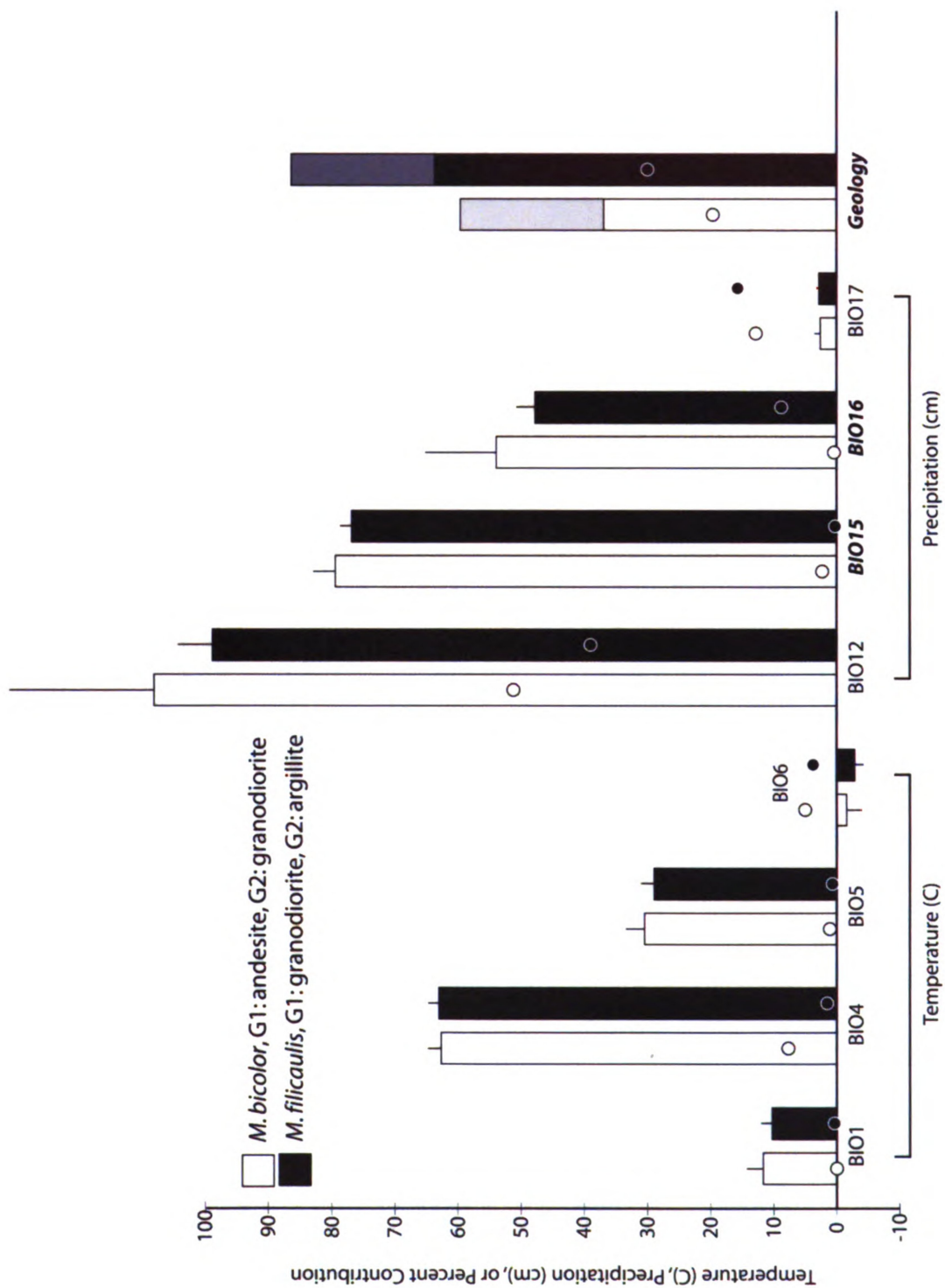


Figure 2.3D. Variation in environmental variables between the species pair *Mimulus bigelovii* and *M. bolanderi*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

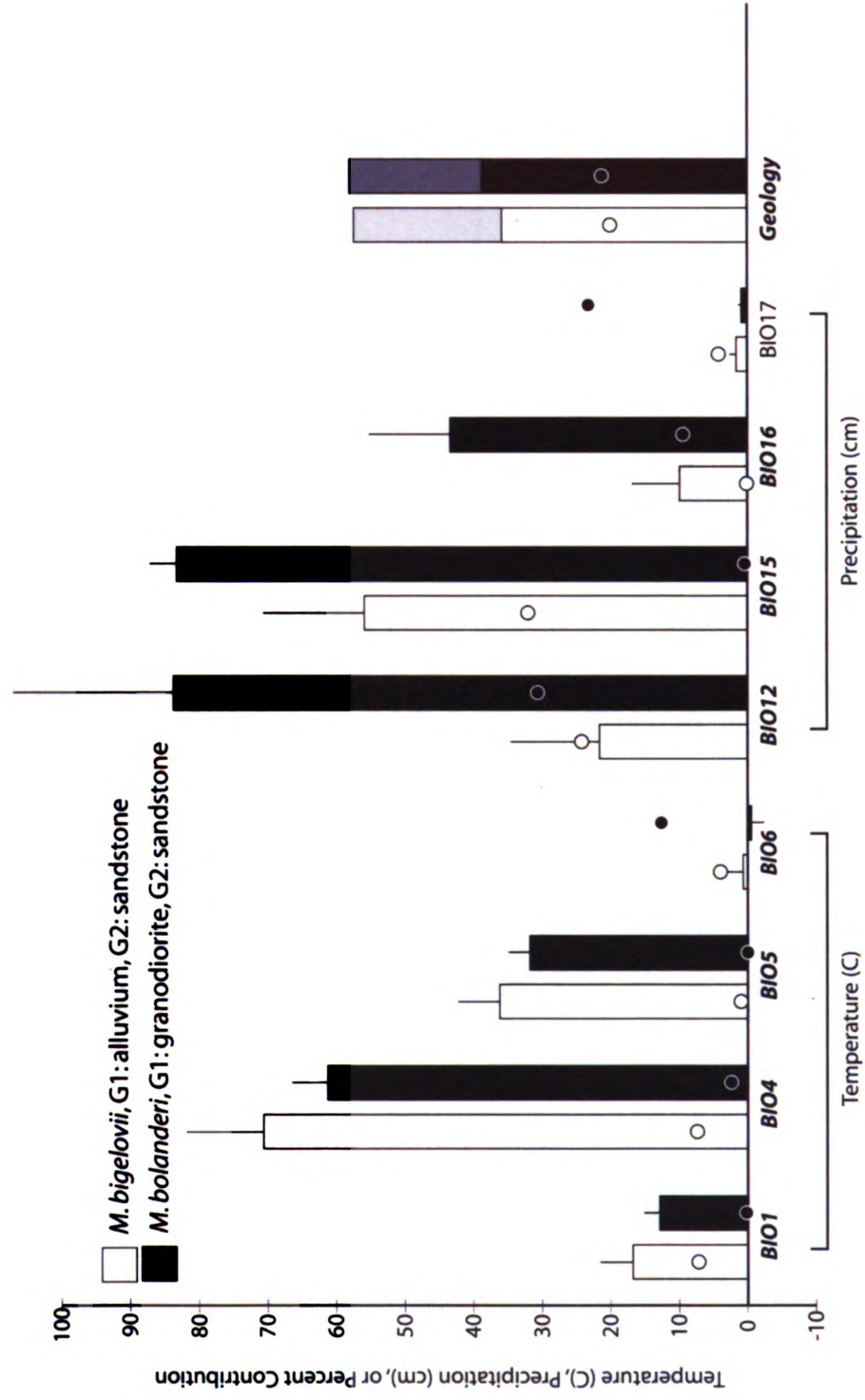


Figure 2.3E. Variation in environmental variables between the species pair *Mimulus brevipes* and *M. johnstonii*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

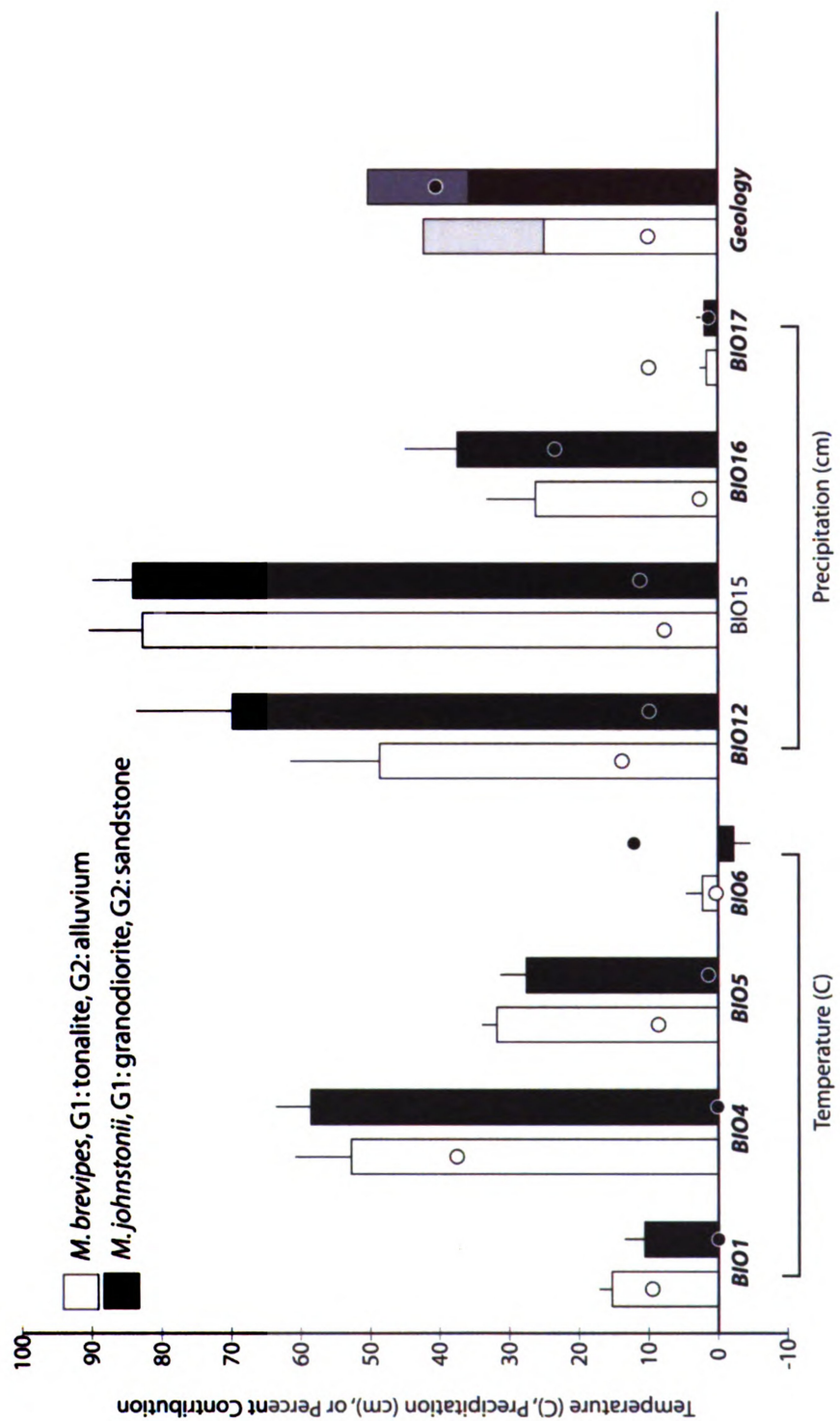


Figure 2.3F. Variation in environmental variables between the species pair *Mimulus cardinalis* and *M. lewisii*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

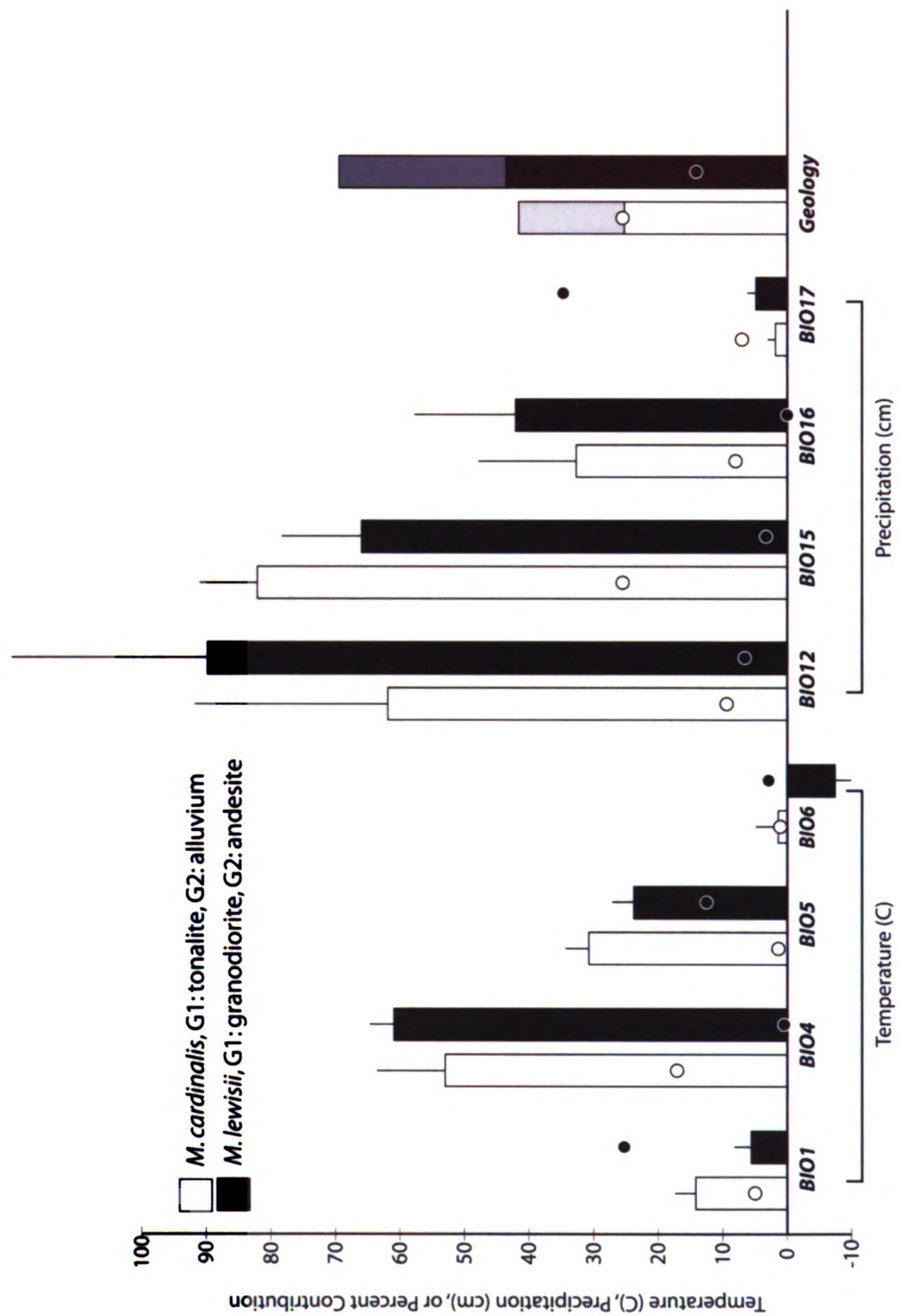


Figure 2.3G. Variation in environmental variables between the species pair *Mimulus constrictus* and *M. whitneyi*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

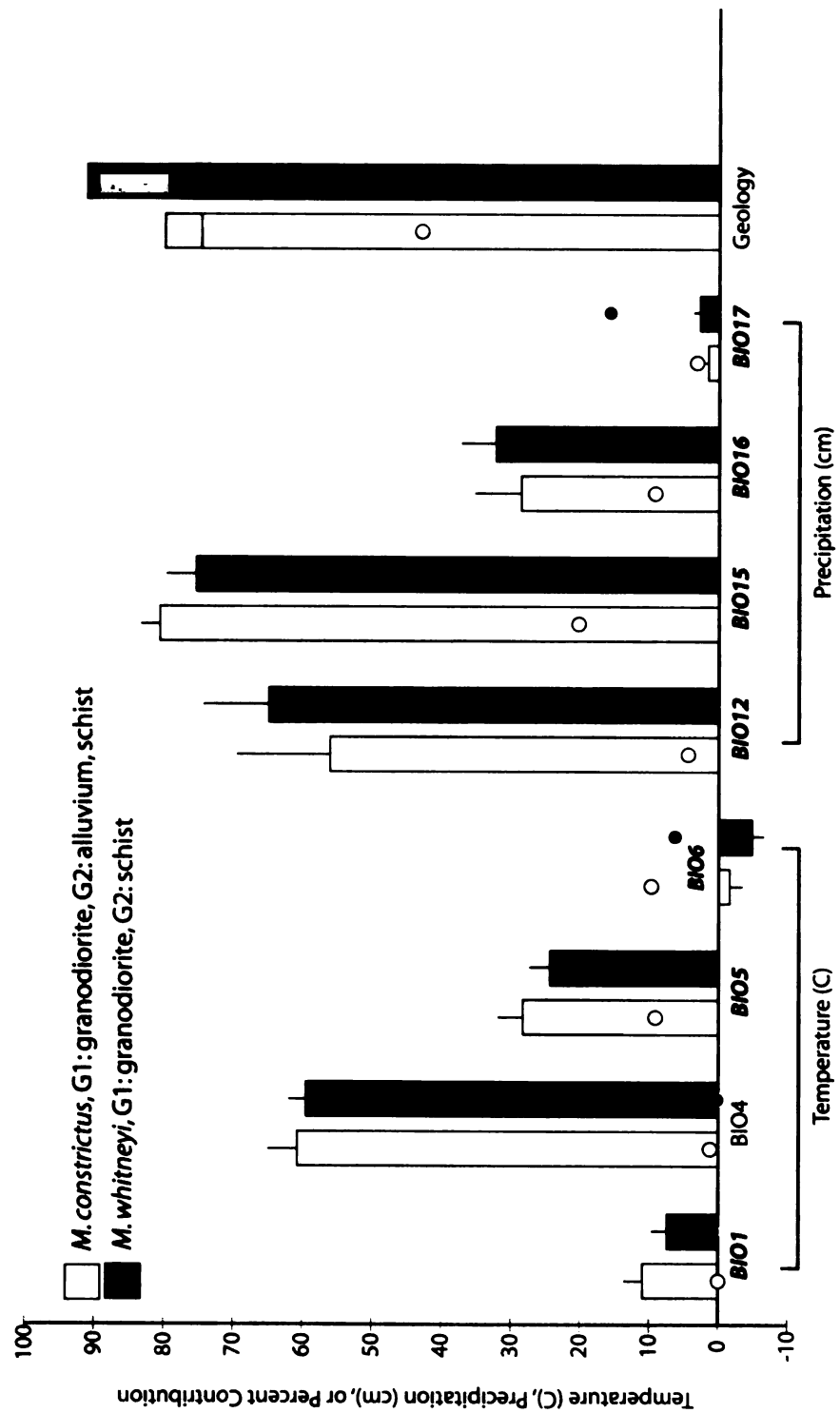


Figure 2.3H. Variation in environmental variables between the species pair *Mimulus cusickii* and *M. nanus*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

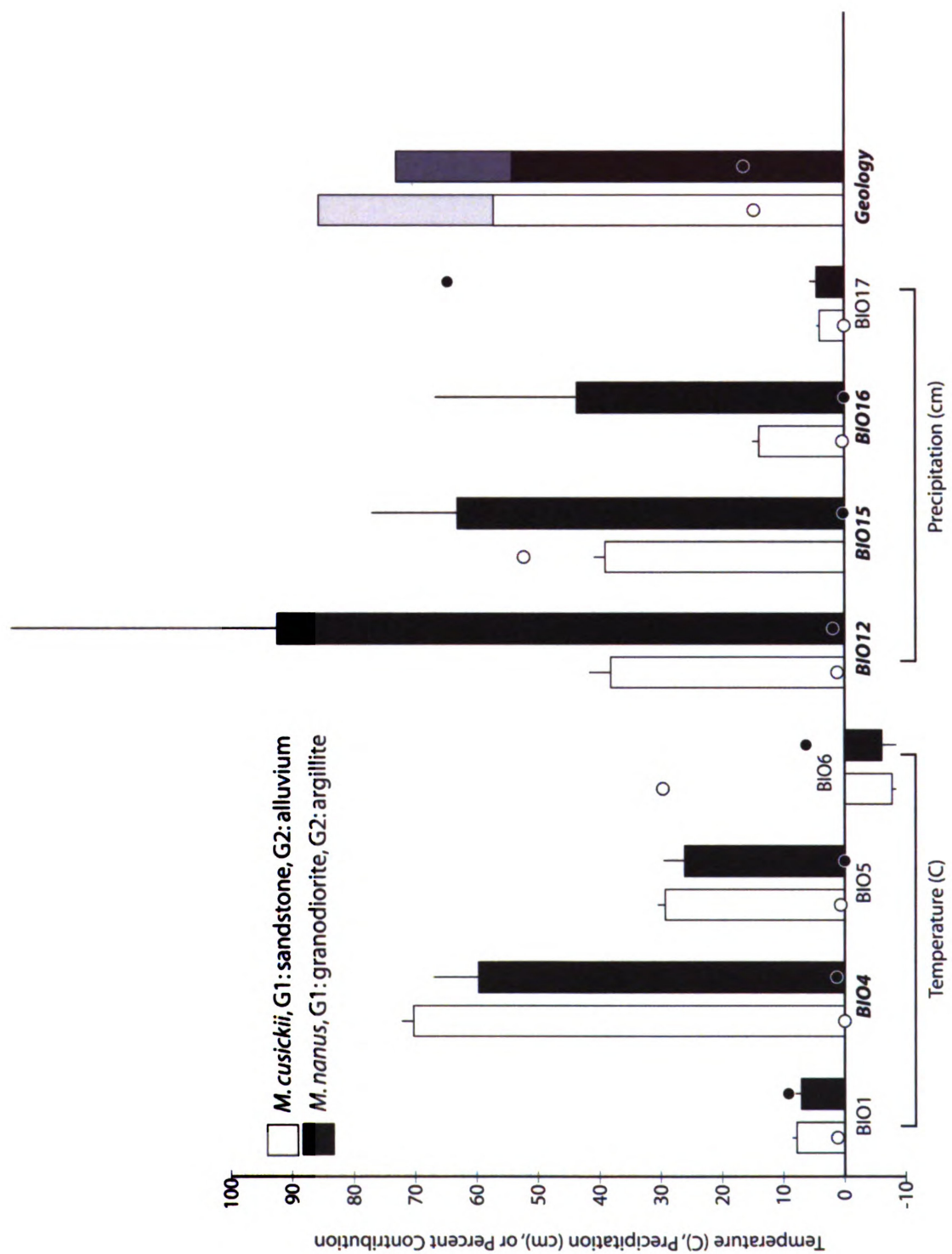


Figure 2.3I. Variation in environmental variables between the species pair *Mimulus douglasii* and *M. kelloggii*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

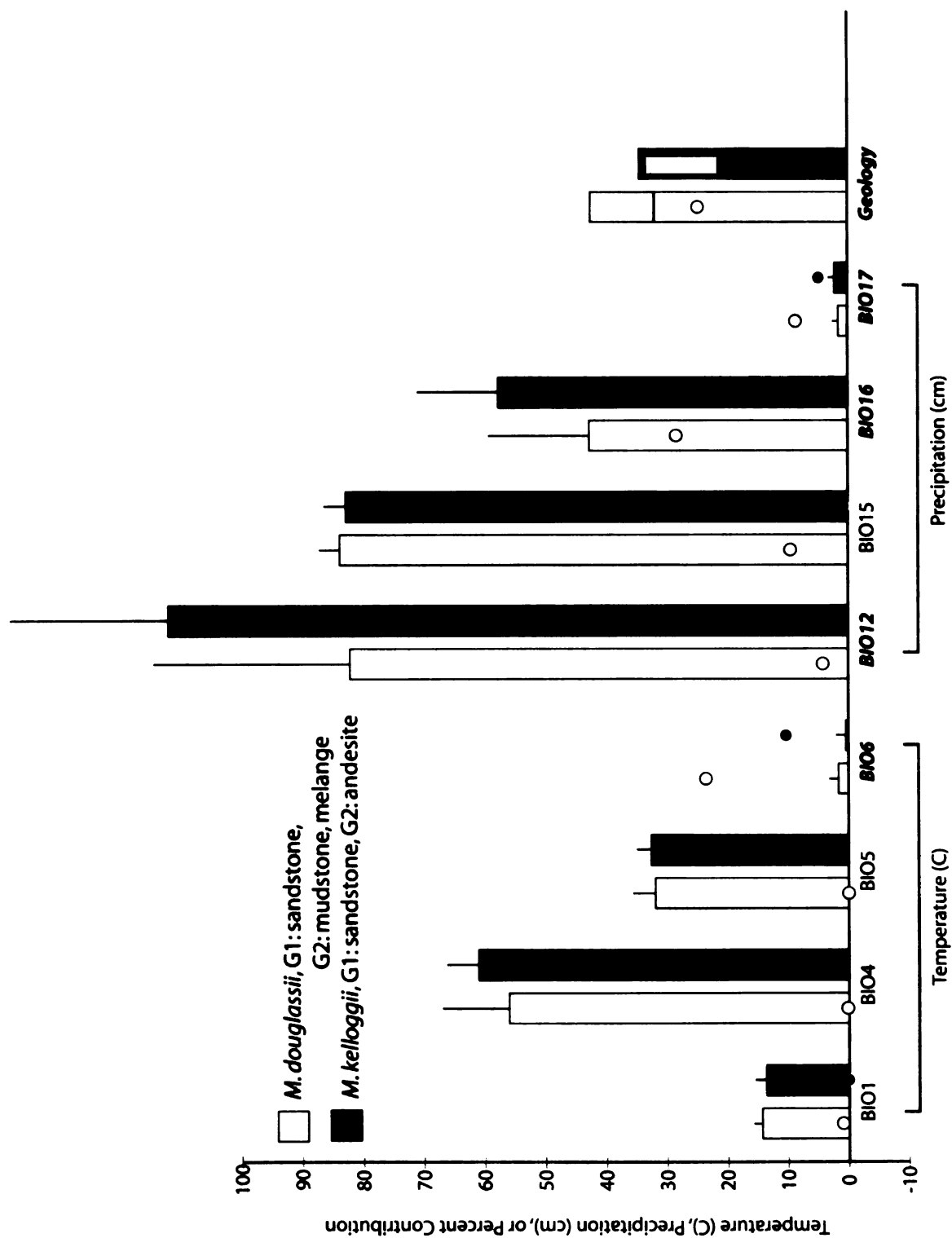


Figure 2.3J. Variation in environmental variables between the species pair *Mimulus floribundus* and *M. norrisii*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

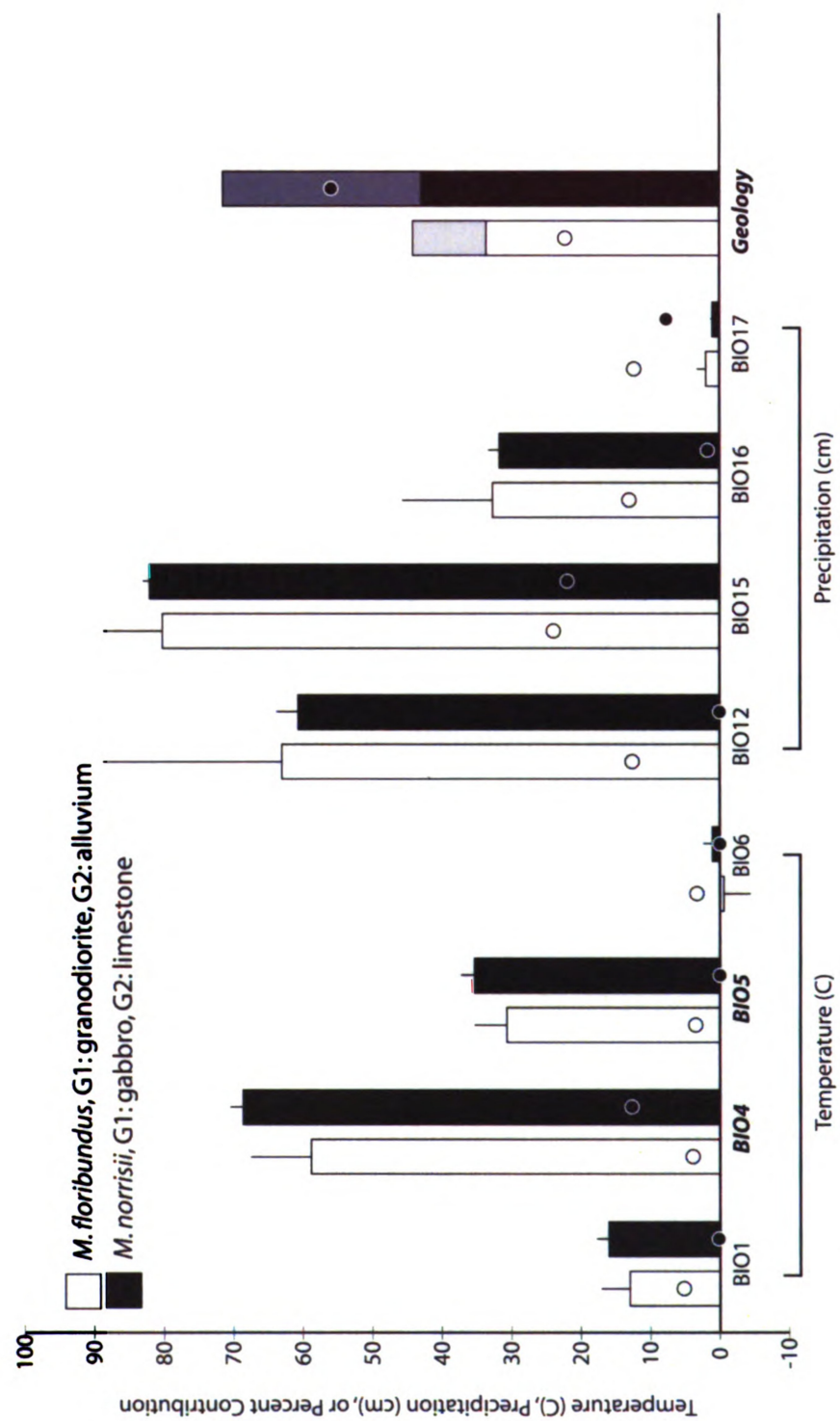


Figure 2.3K. Variation in environmental variables between the species pair *Mimulus gracilipes* and *M. palmeri*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

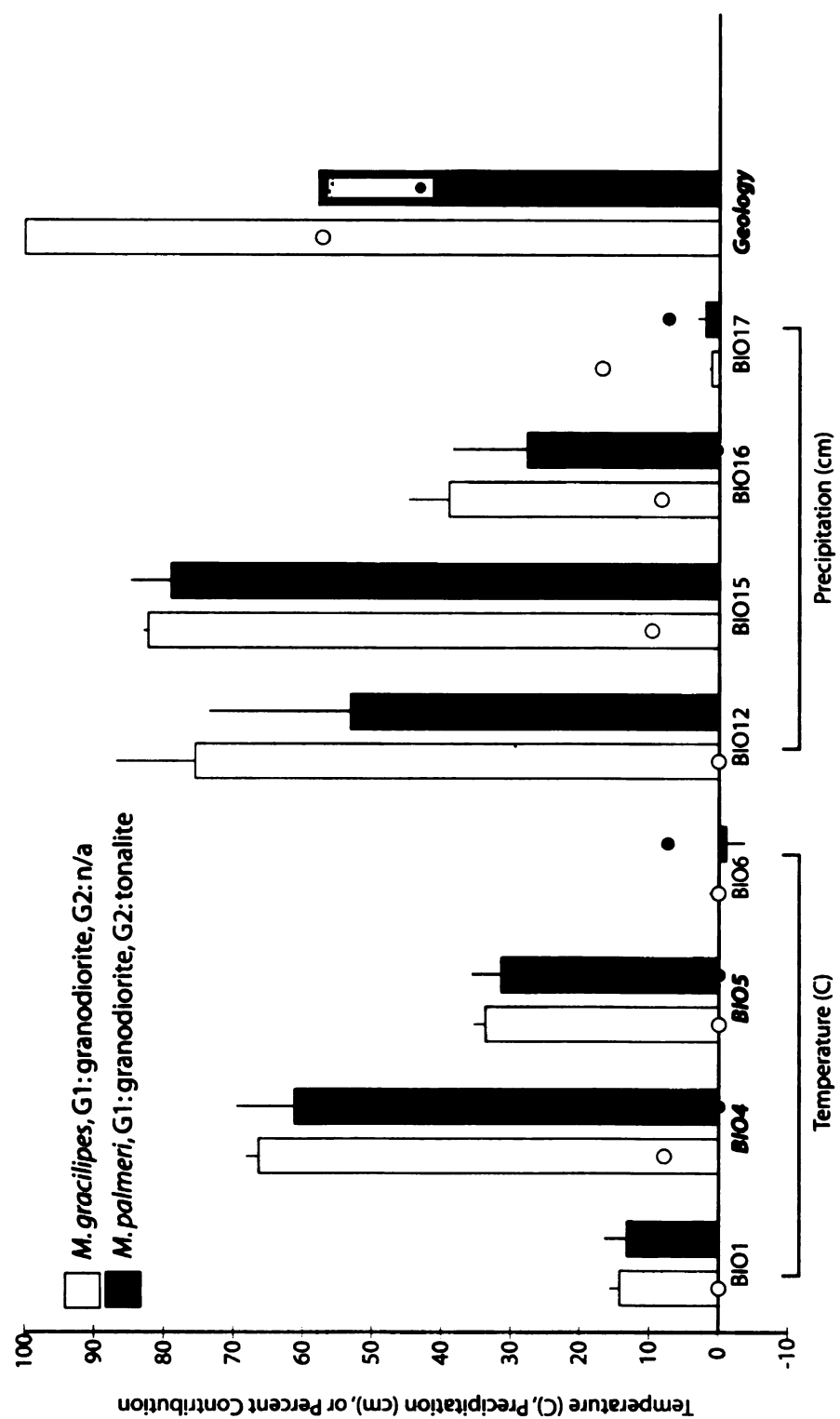


Figure 2.3L. Variation in environmental variables between the species pair *Mimulus parryi* and *M. rupicola*. Circles represent the relative contribution of each variable to the ecological niche models presented in Figures 2.2A-X. For continuous variables from the WORLDCLIM dataset (Table 2.2), boxes represent average values for each variable with standard deviation error bars. For the categorical variable, geology, boxes represent the frequency of the first and second most common category within the dataset, and their identities are noted in the legend. Labels for each environmental variable are bold/italicized when Mann-Whitney U or contingency analysis revealed significant differences between species.

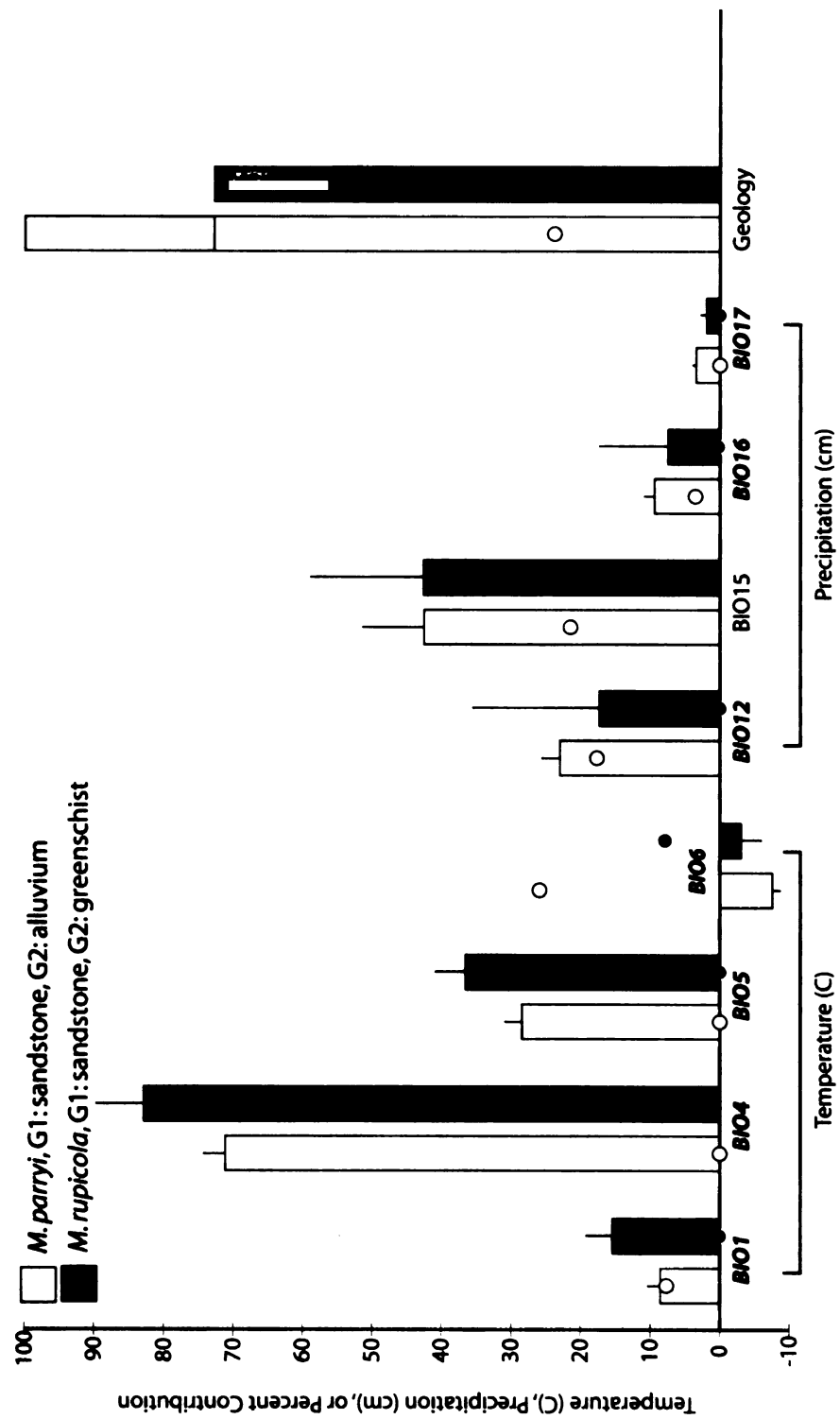


Figure 2.4A. Identity tests using Schoener's D. Results of 'Identity' tests implemented in ENMTools (<http://enmtools.blogspot.com>) (Warren et al. 2008). Histograms represent null distributions of Schoener's D values for each species pair based on randomizing species identities associated with environmental variables used in ecological niche modeling. The distributions consist of 100 pseudoreplicates each, and the arrows indicate where the actual value for Schoener's D between species pairs falls. In all cases, actual values are significantly different from a null expectation at $\alpha = 0.05$, indicating that ecological niches of these species pairs are more dissimilar than expected by chance.

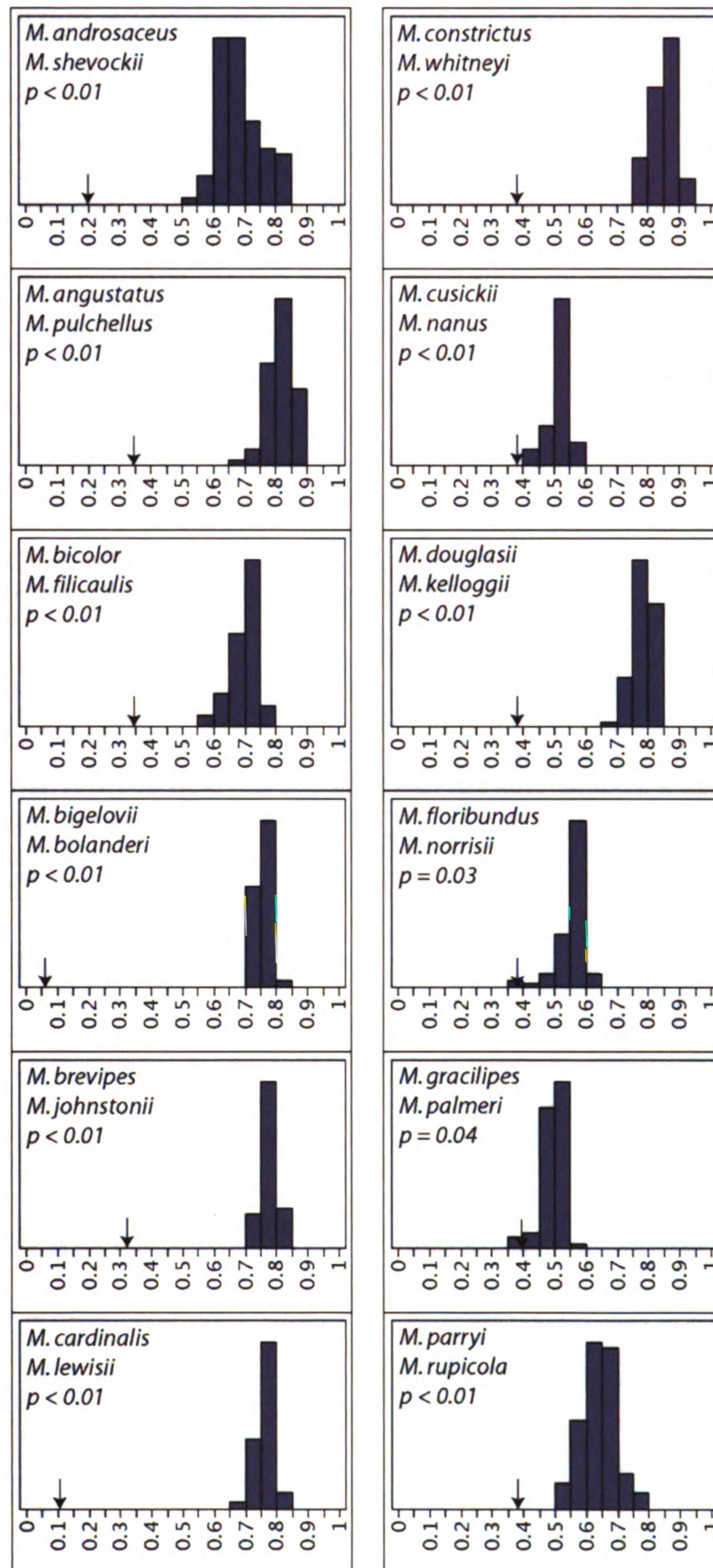
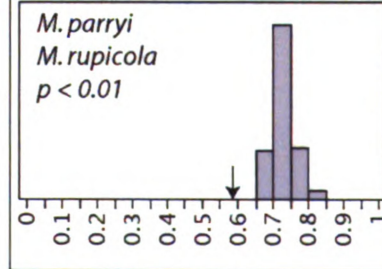
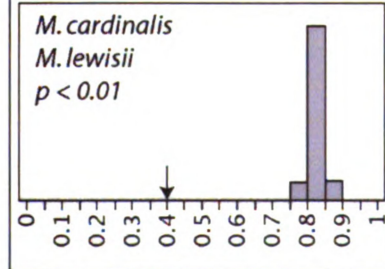
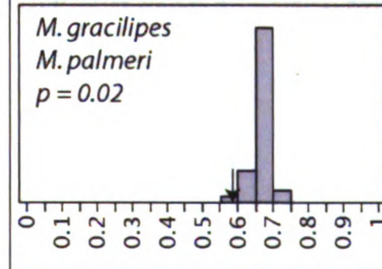
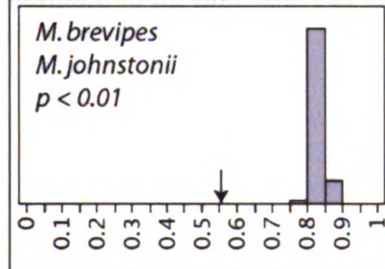
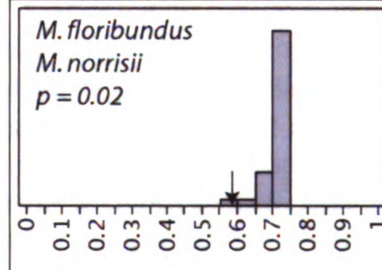
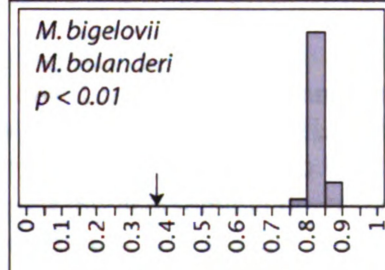
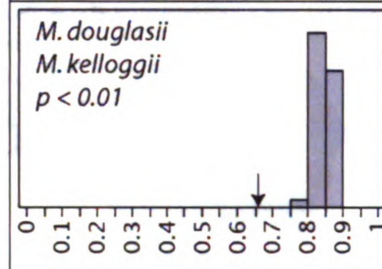
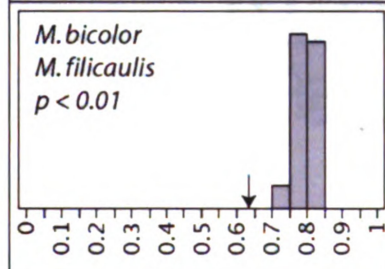
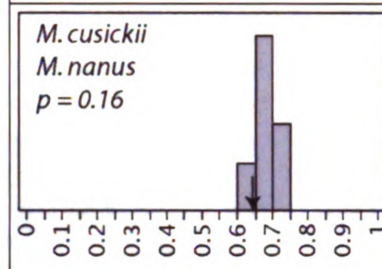
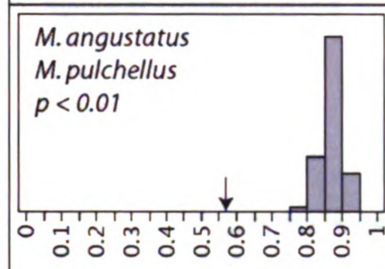
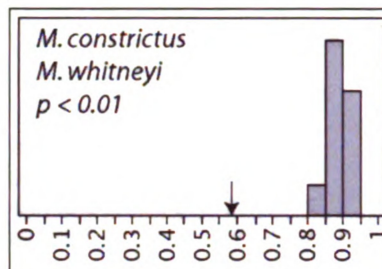
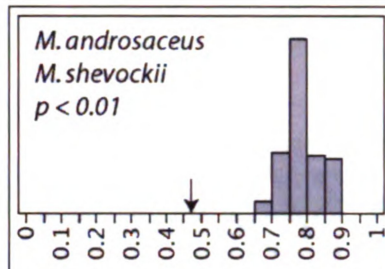


Figure 2.4B. Identity tests using *I* statistic. Results of ‘Identity’ tests implemented in ENMTools (<http://enmtools.blogspot.com>) (Warren et al. 2008). Histograms represent null distributions of *I* values (see Warren et al 2008 for details) for each species pair based on randomizing species identities associated with environmental variables used in ecological niche modeling. The distributions consist of 100 pseudoreplicates each, and the arrows indicate where the actual value for *I* between species pairs falls. In 11 of the 12 species pairs actual values were significantly different from a null expectation at $\alpha = 0.05$, indicating that ecological niches of these species pairs are more dissimilar than expected by chance.



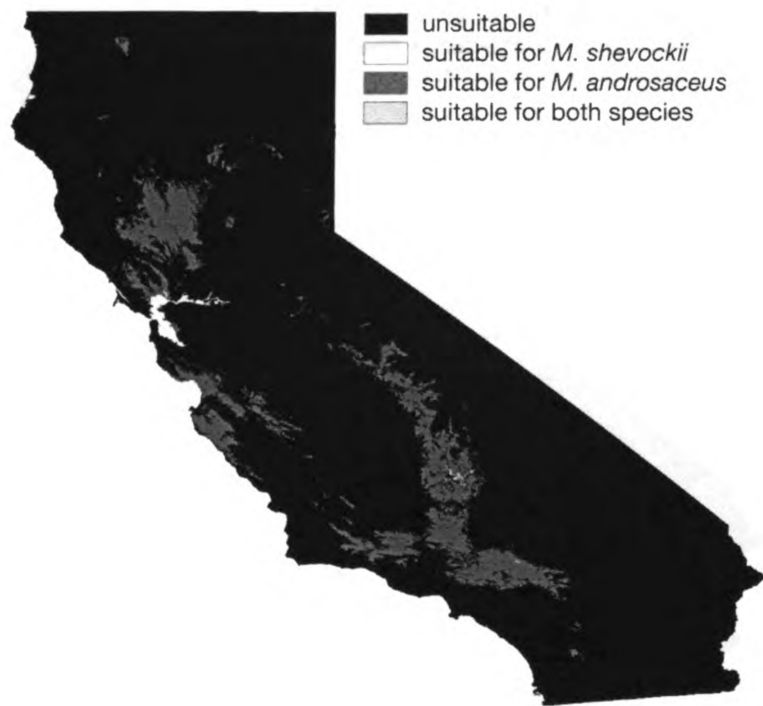


Figure 2.5A. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus androsaceus* and *M. shevockii*. The continuous ecological niche models shown in Figure 2.2A and B were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. androsaceus* alone, suitable to *M. shevockii* alone, and suitable to both species.

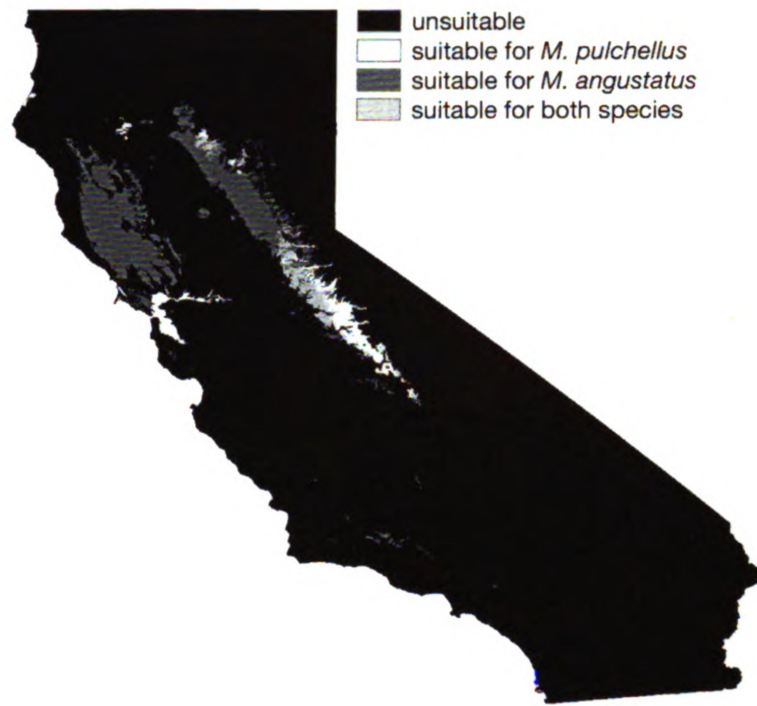


Figure 2.5B. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus angustatus* and *M. pulchellus*. The continuous ecological niche models shown in Figure 2.2C and D were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. angustatus* alone, suitable to *M. pulchellus* alone, and suitable to both species.

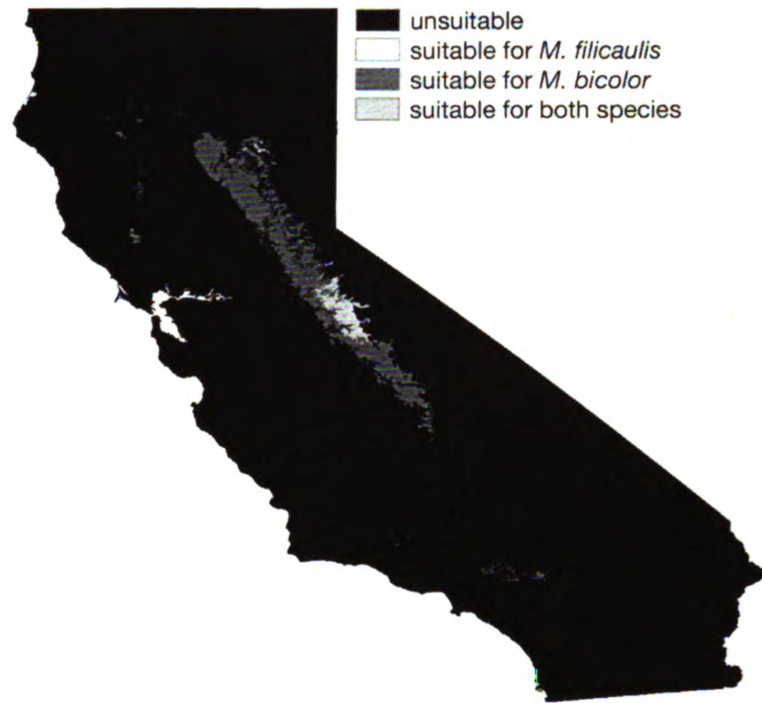


Figure 2.5C. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus bicolor* and *M. filicaulis*. The continuous ecological niche models shown in Figure 2.2E and F were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. bicolor* alone, suitable to *M. filicaulis* alone, and suitable to both species.

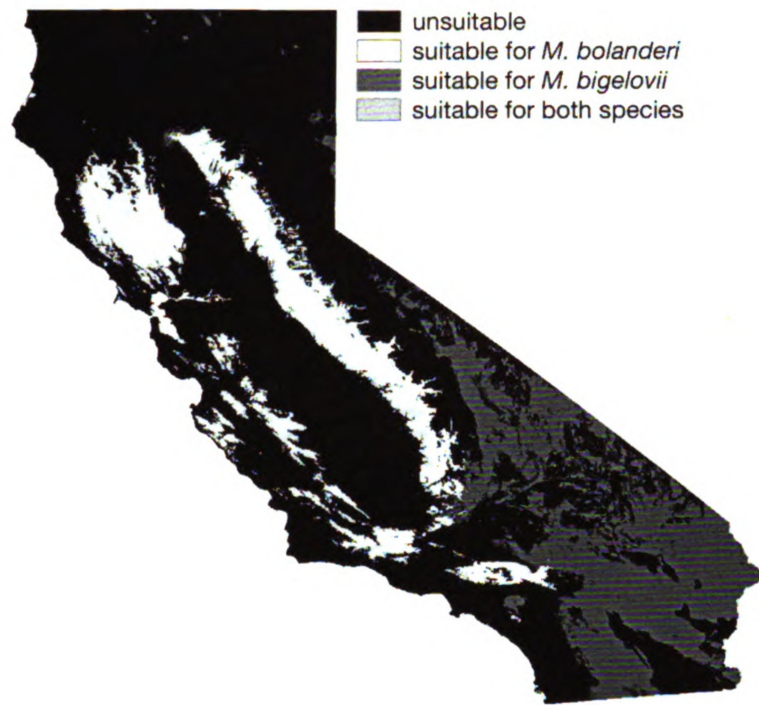


Figure 2.5D. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus bigelovii* and *M. bolanderi*. The continuous ecological niche models shown in Figure 2.2G and H were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlaid onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. bigelovii* alone, suitable to *M. bolanderi* alone, and suitable to both species.

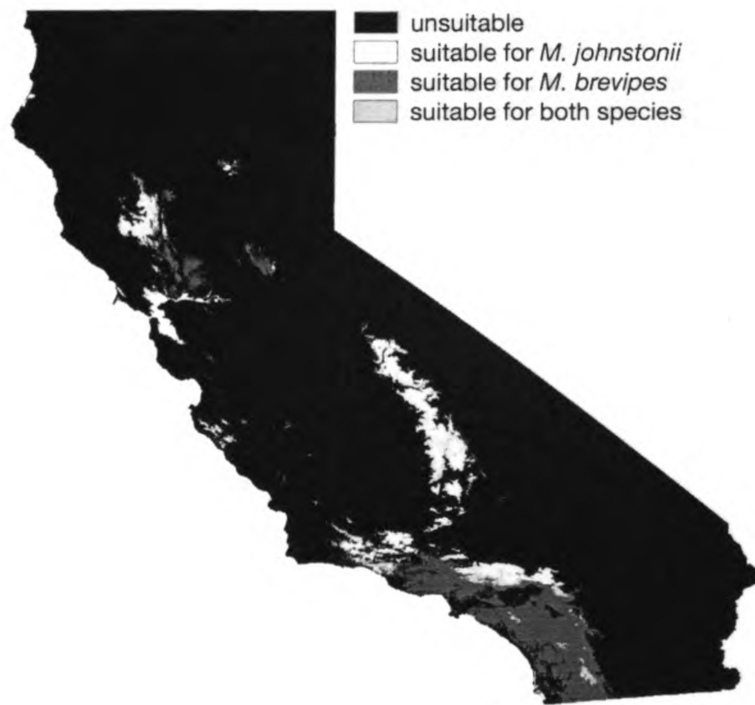


Figure 2.5E. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus brevipes* and *M. johnstonii*. The continuous ecological niche models shown in Figure 2.2I and J were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. brevipes* alone, suitable to *M. johnstonii* alone, and suitable to both species.

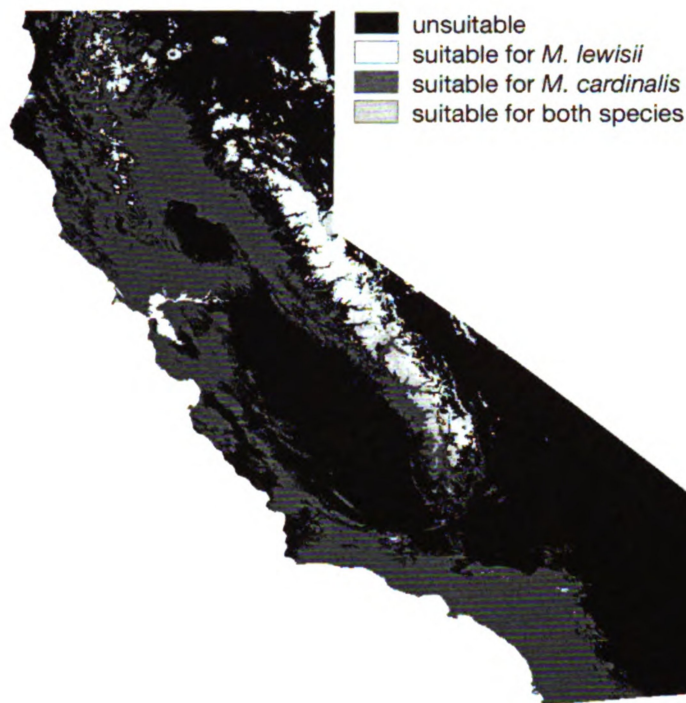


Figure 2.5F. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus cardinalis* and *M. lewisii*. The continuous ecological niche models shown in Figure 2.2K and L were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlaid onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. cardinalis* alone, suitable to *M. lewisii* alone, and suitable to both species.

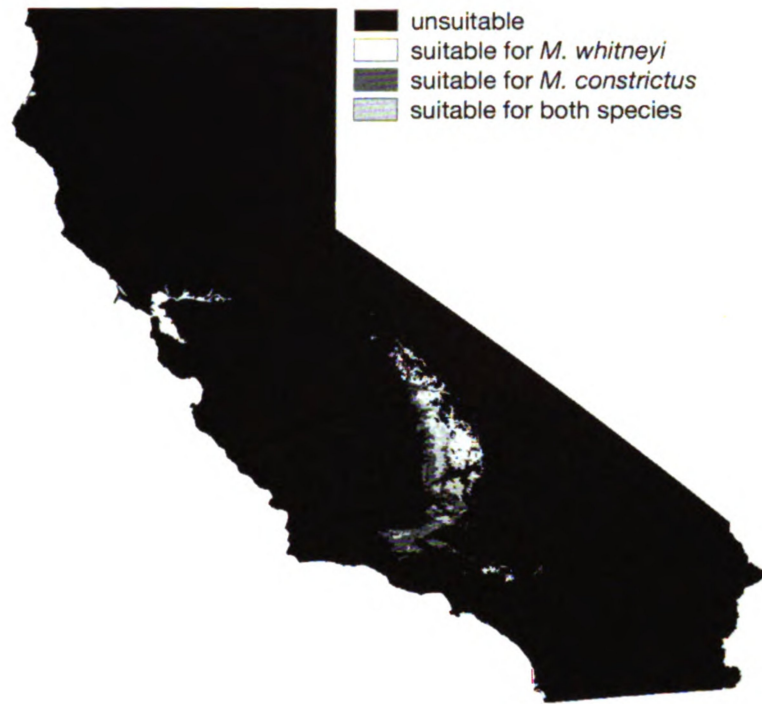


Figure 2.5G. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus constrictus* and *M. whitneyi*. The continuous ecological niche models shown in Figure 2.2M and N were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. constrictus* alone, suitable to *M. whitneyi* alone, and suitable to both species.

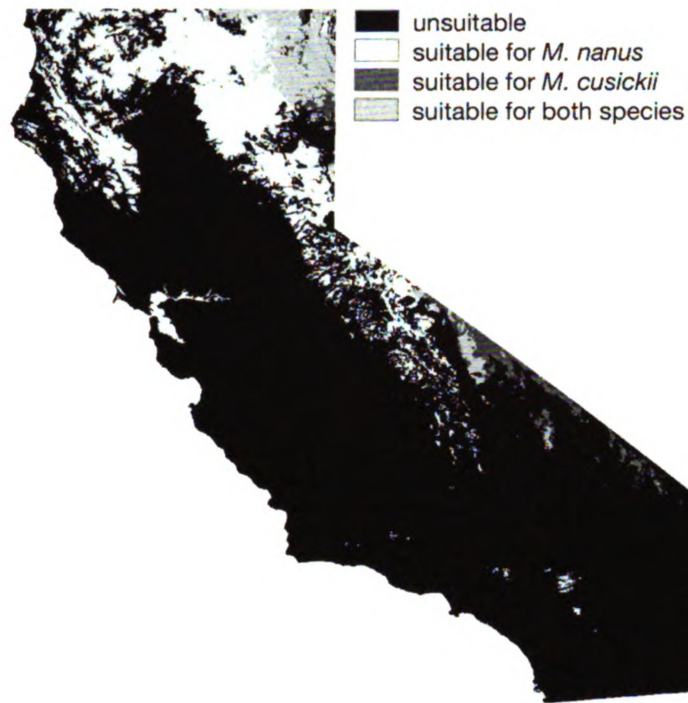


Figure 2.5H. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus cusickii* and *M. nanus*. The continuous ecological niche models shown in Figure 2.2O and P were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. cusickii* alone, suitable to *M. nanus* alone, and suitable to both species.

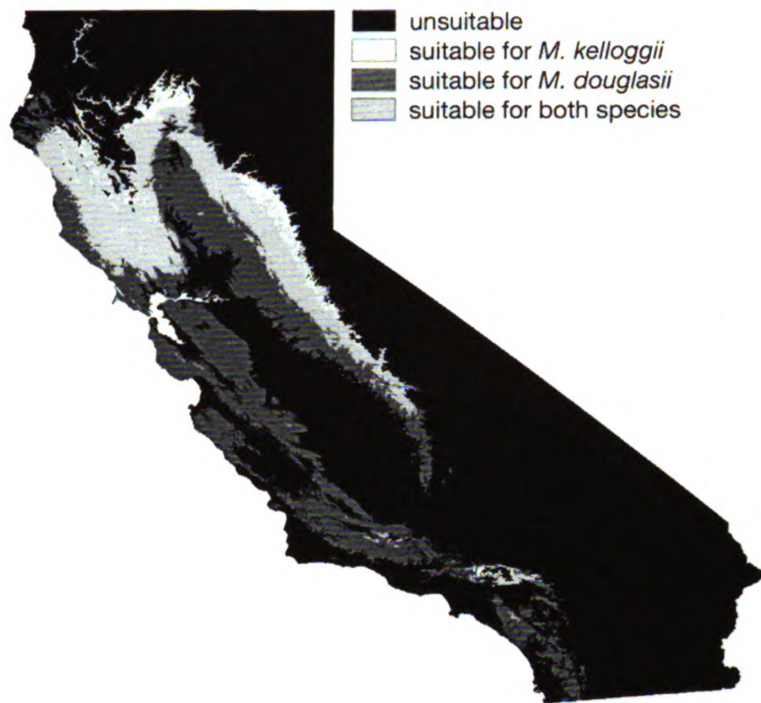


Figure 2.5I. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus douglasii* and *M. kelloggii*. The continuous ecological niche models shown in Figure 2.2Q and R were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. douglasii* alone, suitable to *M. kelloggii* alone, and suitable to both species.

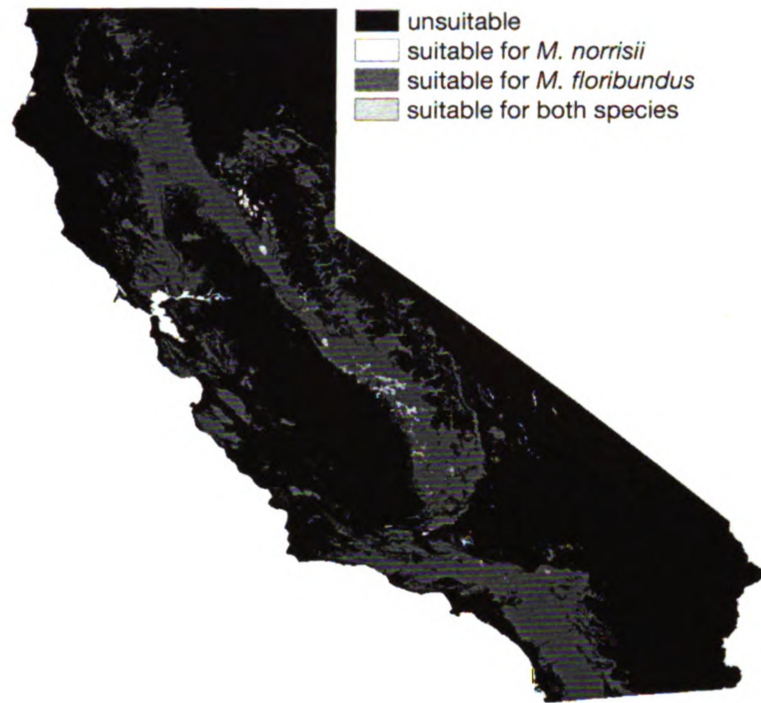


Figure 2.5J. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus floribundus* and *M. norrisii*. The continuous ecological niche models shown in Figure 2.2S and T were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. floribundus* alone, suitable to *M. norrisii* alone, and suitable to both species.

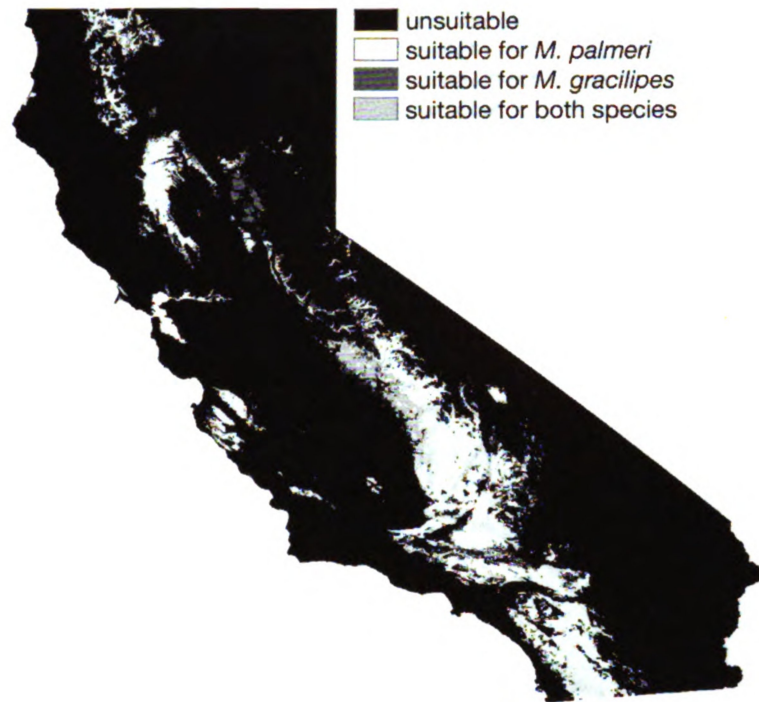


Figure 2.5K. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus gracilipes* and *M. palmeri*. The continuous ecological niche models shown in Figure 2.2U and V were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlayed onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. gracilipes* alone, suitable to *M. palmeri* alone, and suitable to both species.

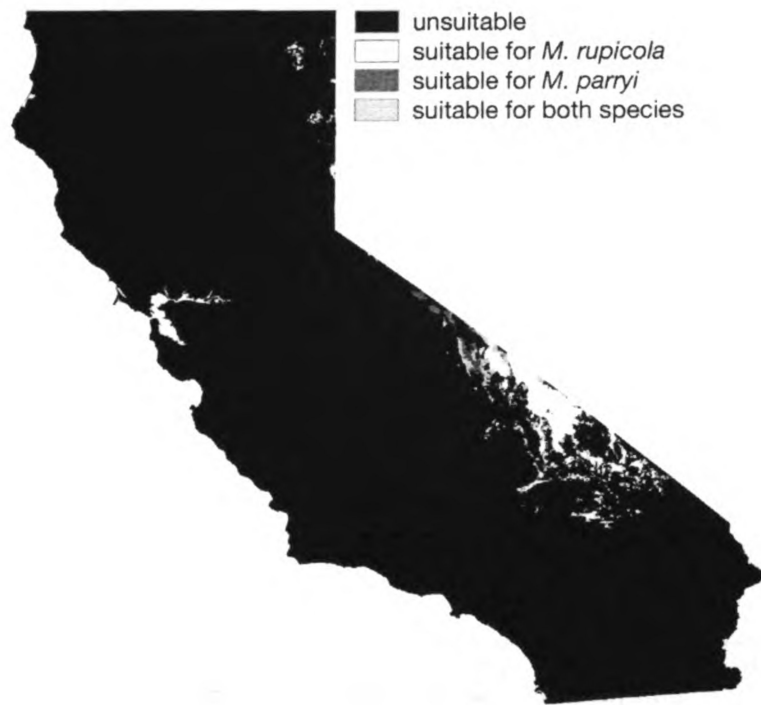


Figure 2.5L. Overlay plot showing amount of ecogeographic overlap in the species pair *Mimulus parryi* and *M. rupicola*. The continuous ecological niche models shown in Figure 2.2W and X were made into binary suitable/unsuitable maps using the 10th percentile of training data method described in the Methods section. Data from the two species were overlaid onto a single map in order to show habitat that is unsuitable to both species, suitable to *M. parryi* alone, suitable to *M. rupicola* alone, and suitable to both species.

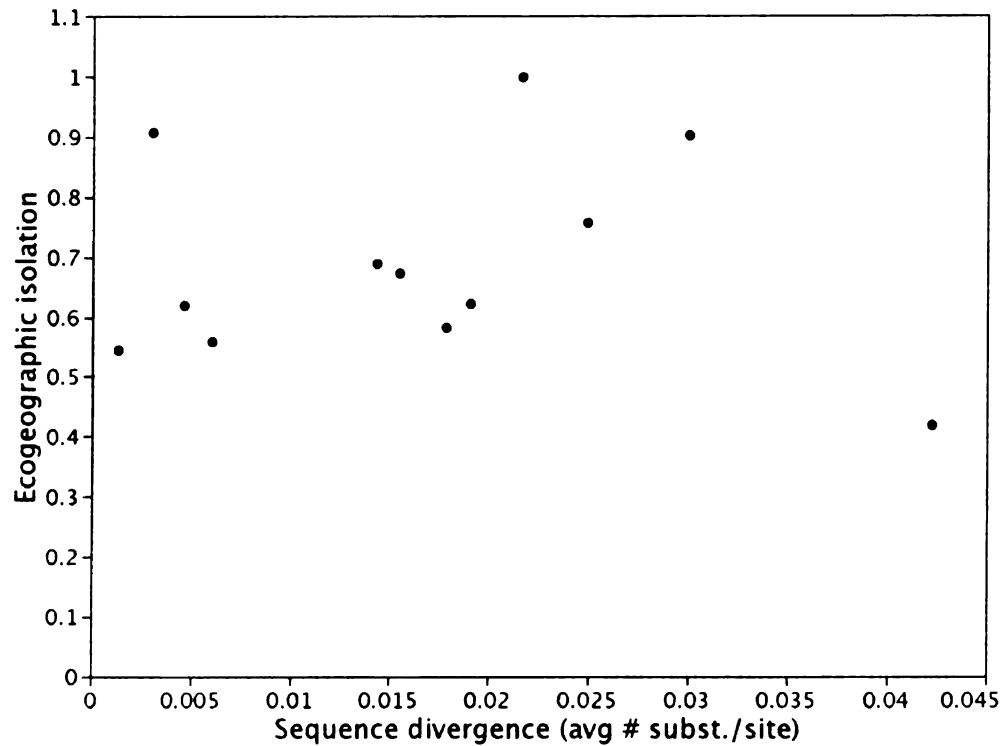


Figure 2.6. Estimates of ecogeographic isolation in *Mimulus* obtained using threshold 2 regressed on sequence divergence between the species pairs (average number of substitutions per site from phylogenetic analysis). Correlation analysis indicates no relationship between ecogeographic isolation and sequence divergence ($p = 0.9188$). Note that 11 out of 12 species pairs exhibit ecogeographic isolation above 0.5 (dashed line), the point at which this barrier will contribute more to total reproductive isolation on a relative scale than any other barrier. The lone species pair (*M. kelloggii* / *M. douglasii*) that does not exceed 0.5 is both the species pair with the highest sequence divergence, and one member of the pair, *M. douglasii* is highly selfing (J. Sobel, unpublished).

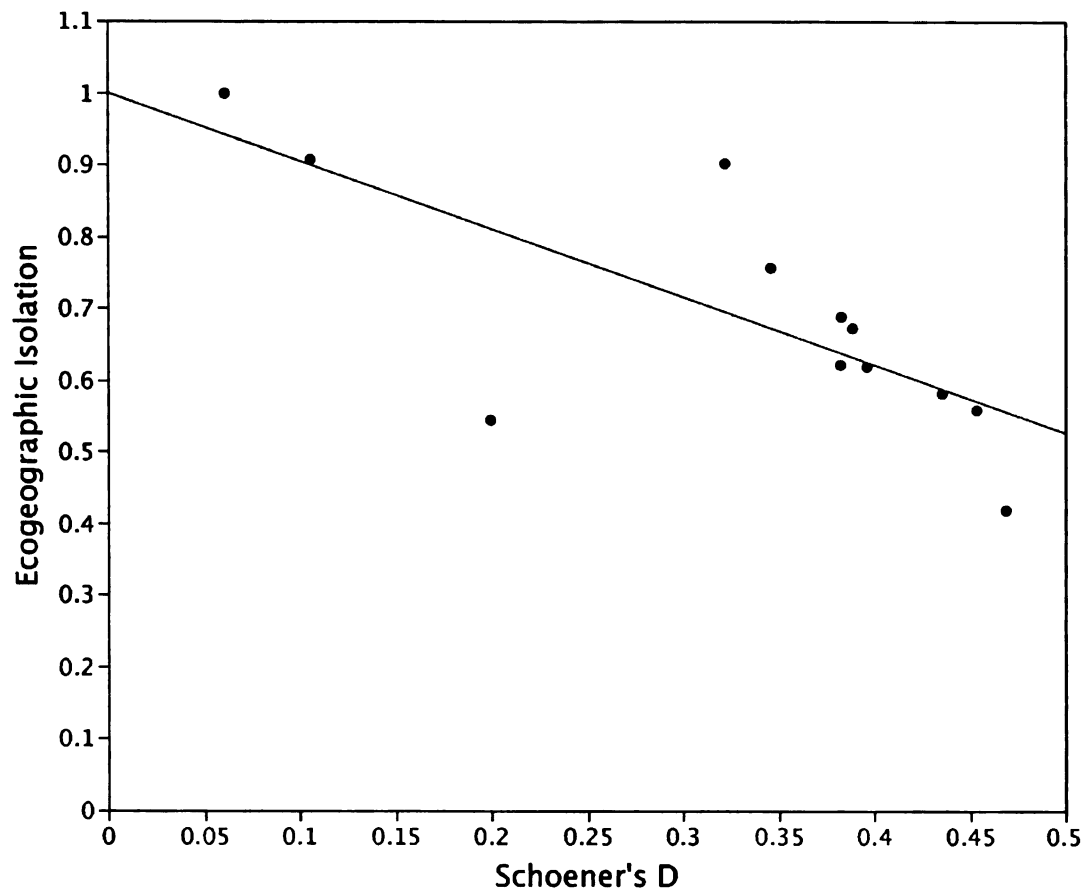


Figure 2.7. Relationship between niche similarity and ecogeographic isolation. Regression analysis of ecogeographic isolation on the measure of niche identity, Schoener's D. While a significant linear relationship exists ($r^2 = 0.547$; $p = 0.006$), the analysis illustrates that the ecogeographic consequence of niche differences can be relatively variable.

Literature Cited

- Angert, A. L. 2006. Demography of central and marginal populations of monkeyflowers (*Mimulus cardinalis* and *M. lewisii*). *Ecology* 87:2014-2025.
- Angert, A. L., H. D. Bradshaw, and D. W. Schemske. 2008. Using experimental evolution to investigate geographic range limits in monkeyflowers. *Evolution* 62:2660-2675.
- Angert, A. L., and D. W. Schemske. 2005. The evolution of species' distributions: Reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution* 59:1671-1684.
- Barracough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155:419-434.
- Beardsley, P. M., and R. G. Olmstead. 2002. Redefining Phrymaceae: The placement of *Mimulus*, tribe Mimuleae and *Phryma*. *American Journal of Botany* 89:1093-1102.
- Beardsley, P. M., S. E. Schoenig, J. B. Whittall, and R. G. Olmstead. 2004. Patterns of evolution in Western North American *Mimulus* (Phrymaceae). *American Journal of Botany* 91:474-489.
- Bolnick, D. I. 2004. Waiting for sympatric speciation. *Evolution* 58:895-899.
- Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology Evolution and Systematics* 38:459-487.
- Bradshaw, H. D., and D. W. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426:176-178.
- Bradshaw, H. D., S. M. Wilbert, K. G. Otto, and D. W. Schemske. 1995. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* 376:762-765.
- Coyne, J. A., and H. A. Orr. 2004, *Speciation*. Sunderland, MA, Sinauer Associates.
- Dobzhansky, T. 1937, *Genetics and the Origin of Species*. New York, Columbia University Press.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.

- Endler, J. A. 1982. Problems in distinguishing historical from ecological factors in biogeography. *American Zoologist* 22:441-452.
- Feder, J. L., X. F. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski et al. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America* 102:6573-6580.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.
- . 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Fishman, L., A. J. Kelly, and J. H. Willis. 2002. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* 56:2138-2155.
- Fitzpatrick, B. M., and M. Turelli. 2006. The geography of mammalian speciation: Mixed signals from phylogenies and range maps. *Evolution* 60:601-615.
- Graham, C. H., S. Ferrier, F. Huettmann, C. Moritz, and A. T. Peterson. 2004a. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19:497-503.
- Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004b. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781-1793.
- Grant, A. L. 1924. A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden* 11:99-388.
- Guisan, A., and W. Thuiller. 2005. Predicting species distributions: Offering more than simple habitat models. *Ecology Letters* 8:993-1009.
- Hiesey, W. M., M. A. Nobs, and O. Bjorkman. 1971. Experimental Studies on the Nature of Species. V. Biosystematics, genetics, and physiological ecology of the *Erythranthe* section of *Mimulus*. Washington, D.C., Carnegie Institution of Washington Publication 628.
- Jarnevich, C. S., T. J. Stohlgren, D. Barnett, and J. Kartesz. 2006. Filling in the gaps: Modelling native species richness and invasions using spatially incomplete data. *Diversity and Distributions* 12:511-520.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. *Physical Review* 106:620-630.

The Jepson Online Interchange for California Floristics 2004. University and Jepson Herbaria, University of California. Berkeley, CA.

Jordan, D. S. 1908. The law of geminate species. *American Naturalist* 42:73-80.

Jordan, D. S., and V. L. Kellogg. 1907, *Evolution and Animal Life: An Elementary Discussion of Facts, Processes, Laws and Theories Relating to the Life and Evolution of Animals*. New York, D. Appleton and Company.

Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated Neotropical gingers. *Evolution* 60:538-552.

Kiang, Y. T. 1972. Pollination study in a natural population of *Mimulus guttatus*. *Evolution* 26:308-310.

Kirkpatrick, M., and V. Ravigne. 2002. Speciation by natural and sexual selection: Models and experiments. *American Naturalist* 159:S22-S35.

Kozak, K. H., C. H. Graham, and J. J. Wiens. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* 23:141-148.

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

Liu, C. R., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385-393.

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

Macnair, M. R. 1983. The genetic control of copper tolerance in the yellow monkey flower, *Mimulus guttatus*. *Heredity* 50:283-293.

Martin, N. H., and J. H. Willis. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61:68-82.

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

—. 1947. Ecological factors in speciation. *Evolution* 1:263-288.

—. 1963, *Animal Species and Evolution*. Cambridge, MA, Harvard University Press.

- . 1984. Species concepts and their applications, Pages 531-540 in E. Sober, ed. *Conceptual Issues in Evolutionary Biology*. Cambridge, MA, MIT Press.
- Nakazato, T., M. Bogonovich, and L. C. Moyle. 2008. Environmental factors predict adaptive phenotypic differentiation within and between two wild andean tomatoes. *Evolution* 62:774-792.
- Nakazato, T., D. L. Warren, and L. C. Moyle. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* 97:680-693.
- Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17:2103-2106.
- Ortega-Huerta, M. A., and A. T. Peterson. 2008. Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Revista Mexicana De Biodiversidad* 79:205-216.
- Papes, M., and P. Gaubert. 2007. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and Distributions* 13:890-902.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34:102-117.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78:419-433.
- Peterson, A. T., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265-1267.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* 190:231-259.
- Poulton, E. G. 1908. What is a species?, Pages 46-94 *Essays on Evolution 1889-1907*. London, Oxford at the Clarendon Press.
- R Development Core Team 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57:1520-1534.

- Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we learned in 40 years? *Evolution* 47:1637-1653.
- Schemske, D. W. 2000. Understanding the origin of species. *Evolution* 54:1069-1073.
- Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences of the United States of America* 96:11910-11915.
- Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16:372-380.
- Schoener, T. W. 1968. Anolis lizards of Bimini: Resource partitioning in a complex fauna. *Ecology* 49:704-726.
- Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and ecology of species range limits. *Annual Review of Ecology Evolution and Systematics* 40:415-436.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295-315.
- Stebbins, G. L. 1950, *Variation and Evolution in Plants*. New York, Columbia University Press.
- Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59:2548-2559.
- Thorpe, R. S., Y. Surget-Groba, and H. Johansson. 2008. The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363:3071-3081.
- Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. *Trends in Ecology & Evolution* 16:330-343.
- Via, S. 2001. Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution* 16:381-390.
- Vickery, R. K. 1959. Barriers to gene exchange within *Mimulus guttatus* (Scrophulariaceae). *Evolution* 13:300-310.
- . 1964. Barriers to gene exchange between members of *Mimulus guttatus* complex (Scrophulariaceae). *Evolution* 18:52-69.
- . 1995. Speciation by aneuploidy and polyploidy in *Mimulus* (Scrophulariaceae). *Great Basin Naturalist* 55:174-176.

- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 62:2868-2883.
- Wiens, J. J. 2004. What is speciation and how should we study it? *American Naturalist* 163:914-923.
- Willis, J. H. 1993. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* 47:864-876.
- Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14:763-773.
- Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220-230.

CHAPTER 3: THE EVOLUTION OF REPRODUCTIVE ISOLATION ACROSS THE
GENUS *MIMULUS*

CHAPTER 3: THE EVOLUTION OF REPRODUCTIVE ISOLATION ACROSS THE GENUS *MIMULUS*

Abstract

Identifying the forms of reproductive isolation that contribute most to total isolation at the time of speciation is a major goal of speciation studies. Two previous methods have gathered significant data on this problem, based on either 1) exhaustive treatment of multiple forms of isolation in a single pair of recently diverged species or 2) comparative study of one or two forms of isolation across multiple species pairs. These two approaches are combined in the presented study to elucidate which forms of isolation are most commonly associated with speciation in 9 pairs of species in the genus *Mimulus*. The barriers measured include ecogeographic isolation, temporal isolation (both in nature and in common garden), seed set isolation (potentially a combination of both gametic and intrinsic postzygotic isolation), intrinsic postzygotic isolation based on relative hybrid fitness, and intrinsic postzygotic isolation based on relative hybrid fertility. Analyses reveal that many of the traits associated with these forms of isolation are significantly different in many species pairs, leading to large variation in how much isolation individual barriers confer. Among the strongest individual barriers are ecogeographic isolation with an average individual strength of 0.641 and intrinsic postzygotic isolation due to relative hybrid fitness with an average individual strength of 0.346. When barriers are considered within the framework of the linear sequential ordering method, ecogeographic isolation retains its individual strength as the first barrier to act throughout the life history of the organisms. However, the relative contribution of later acting

barriers is diminished considerably compared to their individual strength. The average relative contribution of intrinsic postzygotic isolation by fitness differences is 0.077 despite its moderate individual strength. These data confirm that early acting (prezygotic) barriers are highly important to speciation regardless of whether the linear sequential method is employed.

Introduction

As the ultimate source of biodiversity, the process of speciation has attracted the attention of biologists from the time of Darwin (1859). Under the biological species concept (Mayr 1942), species are defined as interbreeding natural populations, and it has been appreciated for many years that there are many factors that can limit gene flow in nature (Dobzhansky 1937; Mayr 1942; Poulton 1908). Much intense debate has surrounded the issue of which of these barriers should be studied (Coyne and Orr 2004; Rice and Hostert 1993; Schemske 2000), and many would consider the 'holy grail' of speciation as the identification of which of these forms of isolation operates most commonly in nature. Forms of isolation are commonly divided into those that operate before zygote formation (prezygotic) and those that act after hybrid formation (postzygotic), and the relative importance of these two broad category of isolation has received considerable attention (e.g. Coyne and Orr 1989; Coyne and Orr 1997; Mendelson 2003).

Prezygotic forms of isolation are typically based on ecological differences between species, affecting the likelihood that species will encounter each other in nature. A primary form of prezygotic isolation involves adaptation to geographically different locations, resulting in ecogeographic isolation. This form of isolation has been largely neglected in the literature (Sobel et al. 2010). However, in the cases where it has been measured, it strongly separates many recently diverged species in nature (Sobel Chapter 2; Ramsey et al. 2003; Kay 2006). A number of other prezygotic mechanisms can also act. Temporal isolation, for example, restricts gene flow between populations or species that differ in the timing of reproduction (e.g. Dopman et al. 2010; Lowry et al. 2008;

Pascarella 2007 ; Yamamoto & Sota 2010). One virtually unexplored aspect of temporal isolation is that the timing of reproduction can vary both due to intrinsic and extrinsic means. Intrinsic temporal isolation is due to differences in either mating or flowering time genes. Extrinsic temporal isolation results from differences in mating or flowering times that are mediated by other factors such as habitat. For example, spatial variation in timing of precipitation may determine that flowering occurs in species A several weeks earlier than for species B (with flowering coinciding with peak soil moisture availability in each habitat). Species A and B may therefore differ in the timing of reproduction without experiencing any differences in flowering time alleles.

Sexual isolation in animals and pollinator isolation in plants can also serve as a form of isolation based on mate selection. Sexual isolation is known to be strong in many cases of *Drosophila*, even when postzygotic barriers are absent (Coyne & Orr 1989; 1997); and pollinator isolation has been shown to generate significant isolation when plants have shifted between pollination syndromes (e.g. Cruzan and Arnold 1994; Ramsey et al. 2003; Schemske and Bradshaw 1999). Postmating, prezygotic isolation also occurs in which pollination or mating occurs, but either pollen or sperm precedence favors fertilization by conspecific gametes (Howard 1999). This form of isolation presents significant difficulties in studying, as it is often impossible to distinguish the relative impacts of gametic interactions from early zygote abortion.

Postzygotic isolation occurs when hybrids are less fit than parents, either due to intrinsic or extrinsic means (Coyne and Orr 2004). Intrinsic postzygotic isolation can occur before hybrids reproduce through differences in survival and growth relative to parents, resulting in lower viability (e.g. Tang and Presgraves 2009). Hybrids may also

experience differences in fertility relative to parents, resulting in intrinsic postzygotic isolation based on either full or partial sterility (e.g. Phadnis and Orr 2009). Hybrids may also be unfit due to a lack of suitable ecological niche to fill. This extrinsic form of postzygotic isolation has been investigated in cases of lake populations of benthic and limnetic stickleback fish that are known to hybridize (Hatfield and Schluter 1999).

One mechanism of speciation that often gets cited as a form of reproductive isolation is due to differences in mating system. For example, Martin and Willis (2007) investigated reproductive isolation between two species of *Mimulus* in which one member of the pair was highly selfing. While selfing was not measured directly, its impact on gene flow was inferred from hybridization rates. However, selfing species are a notorious difficulty for the biological species, so including it as a legitimate form of isolation is problematic. Increases in selfing isolate individuals within a population from each other just as much as from other populations or species (Coyne and Orr 2004; D. Schemske, pers. comm.), so characterizing the effect it has on gene flow is not straightforward.

Two approaches have most commonly been used to evaluate the relative importance of different forms of isolation. The first approach is to choose a pair of closely related species, and measure the strength of as many forms of isolation as possible to evaluate which is strongest (e.g. Ramsey et al. 2003; Chari and Wilson 2001; Husband and Sabara 2004). This method can reveal which forms of isolation are strongest in a particular species pair, but because species pairs are generally selected non-randomly, many biases exist in which forms of isolation are likely to be encountered. An example of this approach is the work of Ramsey et al. (2003) in which *Mimulus*

cardinalis and *M. lewisii* were found to be isolated predominantly by ecogeographic isolation and pollinator preference despite the presence of several other potential sources of isolation. An alternative approach is to measure only a limited number of isolating barriers, but to examine a large number of species pairs within a group of organisms. This approach is exemplified by Coyne & Orr's (1989; 1997) comparative study of mating and postzygotic isolation in *Drosophila*. They examined the rate at which these two forms of isolation evolved by comparing the strength of each form of isolation with the time since speciation (genetic distance). This revealed that the prezygotic barrier based on mating preferences evolved faster than the postzygotic barrier based on relative hybrid fitness and fertility. This difference was primarily due to sympatric taxa, sparking a rejuvenation of interest in the process of reinforcement.

While each of these approaches can help elucidate the relative importance of isolating barriers, neither are sufficient to answer the question of which forms of isolation are most important on their own. The single species pair study design may reveal the forms of isolation that are important in that pair of species, but generalizing about common trends in speciation is difficult from these limited examples. Alternatively, comparative studies of isolation in which multiple species pairs and more than one isolating barrier is measured (Coyne and Orr 1989; Coyne and Orr 1997; Mendelson 2003; Moyle et al. 2004) either rely on limited data from the literature or are so labor intensive that very few isolating barriers can be included. While these studies may reveal the relative rate at which the few included barriers arise, it is impossible to know which forms of isolation are most important when many are excluded from analysis.

A commonly overlooked issues in speciation research is the need to study the forms of isolation that reduce gene flow the most in nature (Rice and Hostert 1993; Schemske 2000; Sobel et al. 2010). Coyne & Orr (1989) first recognized that in calculating the strength of total reproductive isolation, postzygotic isolation can only account for any isolation that was unaccounted for by prezygotic isolation. Ramsey et al. (2003) provided an important insight that many forms of isolation can be treated sequentially because they operate at distinct stages in the life history of organisms. In this approach, late acting barriers are discounted by any barriers that affect gene flow at an earlier stage. By relating each barrier's individual strength to the amount of gene flow that it actually has the potential to interrupt, this method allows a direct calculation of the relative contribution of each barrier to total reproductive isolation.

In the presented work, the comparative approach of Coyne & Orr (1989) was combined with the single species pair approach of Ramsey et al. (2003) to ask the following:

- 1) Which forms of isolation have the highest individual strengths among recently diverged species pairs?
- 2) What are the relative contributions of different forms of isolation to total reproductive isolation? and Are there general trends in which forms of isolation are strongest?
- 3) Are early acting barriers more likely to interrupt gene flow than late acting barriers?

Materials & Methods

Study system

The genus *Mimulus* consists of approximately 120 species of wildflowers, approximately 75% of which are found in western North America (Grant 1924). The genus exhibits a broad diversity in traits interesting to ecologists and evolutionary biologists including: growth form (Lowry et al. 2008), mating systems (Fishman et al. 2002; Willis 1993), ploidy levels (Vickery 1995; Beardsley et al. 2004), pollination syndromes (Bradshaw and Schemske 2003; Bradshaw et al. 1995; Schemske and Bradshaw 1999; Streisfeld and Kohn 2005), edaphic specialization (Macnair 1983), and climatic tolerances (Angert and Schemske 2005). Previous phylogenetic work has elaborated relationships at broad taxonomic levels (Beardsley and Olmstead 2002), and a species-level phylogeny is available with nearly complete coverage of the North American species (Beardsley et al. 2004). A rich history of ecological and genetic work (Hiesey et al. 1971; Kiang 1972; Vickery 1959; Vickery 1964), combined with more recent advances in developing genomic resources (Wu et al. 2008) has elevated the genus to a model system in ecology and evolutionary biology, especially in studies of adaptation and speciation (Fishman and Willis 2001; 2006; Martin and Willis 2007; Ramsey et al. 2003; Sweigart et al. 2007).

Selection of species pairs

Species pairs were selected across the genus using the phylogenetic hypothesis of Beardsley et al. (2004) (Table 3.1). The decision to include species pairs involved a combination of factors including high bootstrap support of the relationship, practical

ability to collect and propagate specimens, and desired broad coverage of the traditional sectional delineations (Grant 1924). In practice, closely related species pairs were identified in the phylogeny, and field collections were attempted for one or both members of the pair. Because many *Mimulus* species are rare, a combination of efforts for locating species was needed, including referencing herbarium records, utilizing the expertise of local forest and park service botanists, and extensive field surveys. When one member of a pair was located, increased efforts were made to locate the alternate member of a pair. The result is selection of species pairs that is not truly random across the genus, but was somewhat haphazard, and more importantly, species were selected with no prior knowledge of reproductive barriers that were in operation.

Each species pair represents very recent divergence in order to ensure that estimates of reproductive isolation obtained are as close as possible to the isolation that was present at the moment of speciation. Therefore, in most cases, species pairs represent each other's closest extant relative. However, due to phylogenetic uncertainty or inability to collect specimens in the field, some species pairs consist of two very close relatives in which a third may be more closely related to either of the two selected. For example, *Mimulus kelloggii*, *M. douglasii*, and *M. congdonii* are a trio of closely related species in section *Oenoe*. The phylogeny of Beardsley et al (2004) suggests that *M. douglasii* and *M. congdonii* are sister species. However, *M. congdonii* proved elusive in the field, and therefore, *M. kelloggii* and *M. douglasii* were instead used as a species pair. In order to maintain phylogenetic independence, and avoid the need for a phylogenetic correction, it is only necessary for each of the included species pairs to be more closely related to each other than to any other species in the study (Felsenstein 1985; Harvey and Pagel 1991),

and this method of species selection achieves this goal (see Sobel, Chapter 2- Figure 2.1).

Nine species pairs in the genus *Mimulus* were sampled in the field between 2005 and 2008 (Figure 3.1). Populations were located in the spring and early summer, and GPS coordinates were recorded (Table 3.1). Collections of seeds were carried out in late summer or early fall.

Overview

In the following sections, methods employed for measuring each form of isolation are described. The order of isolating barriers presented corresponds with the sequence in which they act in nature, progressing from barriers that act early in the life stage of *Mimulus* (prezygotic barriers), and concluding with barriers that act at later stages (postzygotic barriers). Finally, details are provided for how the relative importance of each barrier was measured, both by differences in absolute strengths and in their relative contribution to total isolation.

Temporal isolation

Two types of data were collected to measure the strength of temporal isolation in pairs of *Mimulus* species. The first consisted of collection dates for specimens within the Jepson Online Interchange for California Floristics (UC-JEPS2004). Dates were collected from all specimens in the database (both georeferenced and non-georeferenced), and converted to the Julian date system. This represents the distribution of dates that each species flowers in nature, giving an indication of the total isolation experienced by species pairs based on flowering time. This measure of isolation can include both

intrinsic biological differences in flowering time genes, and/or differences in flowering time mediated by habitat (extrinsic temporal isolation).

Because collection records do not provide sufficient information to estimate isolation using approaches that take relative abundance into consideration, isolation was estimated using the simplified RI_7 equation presented in Chapter 1:

$$RI_{speciesA} = 1 - \left(\frac{Shared}{Shared + Unshared_A} \right).$$

In this equation, “Shared” refers to the Julian dates of co-flowering for species A and B, while “ $Unshared_A$ ” is the duration of the dates that species A flowered while species B did not. Gaps in Julian dates were assumed to be due to patchiness of the collection record rather than actual biology, so timing of flowering was essentially treated as a block from earliest to latest date of specimen collection. Temporal isolation was calculated separately for each species of a pair and then averaged together for a composite estimate of reproductive isolation. Distributions in flowering time were often non-normally distributed, so the non-parametric Mann-Whitney U test implemented in JMP version 8.0 was used to test for significant differences between species.

Plants were also grown in common garden conditions in the Michigan State greenhouses to evaluate the amount of temporal isolation due to intrinsic genetic differences between species. Assays were conducted to establish the optimal gibberellic acid concentration for each species (Sobel, unpublished). Seeds were treated with appropriate gibberellic acid concentration overnight to break dormancy, rinsed thoroughly, sterilized with 25% bleach solution, and sown on Phytoblend agar (Caisson Laboratories) plates with Gamborg B5 plus sucrose nutrient medium (BIOPLUS). Plates

were randomly assigned positions within growth chambers using settings optimized for each species pair. A minimum of 200 seeds from 6 maternal lines were sown, and each seed was given a unique identification number and followed to record day of germination and day of flowering. Upon germination, seedlings were transplanted into a mixture of Baccto High Porosity soil mix with supplemental perlite and vermiculite. After transplanting, seedlings were placed under misters for 3-4 days to acclimate to greenhouse conditions, and then were randomly assigned positions on benches. Seeds and seedlings were monitored for germination and flowering respectively three days a week until germination rates dropped precipitously.

This aspect of temporal isolation was estimated using equation RI_g from Chapter 1:

$$RI_A = 1 - 2 \sum_i \left(\frac{A_i}{A_{total}} \times \frac{B_i}{A_i + B_i} \right).$$

Many of these species can be maintained in the greenhouse indefinitely, and do not naturally senesce without being pollinated. Therefore, an arbitrary cutoff date of three weeks past the last opened flower was used as the flowering duration of the entire population. This convention for estimating temporal isolation results in a symmetrical distribution of flowering times, and therefore a single isolation estimate was calculated for each species as a pair.

Isolation due to differences in seed set

Experimental hybridizations were performed on plants from the temporal isolation growouts described above. For each species pair (Table 1), intraspecific crosses

and interspecific crosses in both directions were done. In previous experiments (Sobel, unpublished), it was determined whether each species would set seeds autogamously. For those species that do set autogamous seeds, dissections were performed to remove anthers previous to stigma receptivity. Flowers were marked as emerging buds, and only flowers that were 2-3 days past opening were used in crosses. On any given day of crossing, flowering individuals were counted to insure that a roughly equal numbers of intraspecific and interspecific crosses could be performed. Non-pollinated controls were also conducted on each day of crossing. Two independent anthers were collected from the paternal parent of each cross, and pollen was applied to the receptive stigma by directly rubbing the anther across the stigmatic surface. Because *Mimulus* has reactive bilabiate stigmas that close when perturbed, pollinations had to be performed quickly to ensure a saturating amount of pollen had been applied before the stigma closes. Therefore, for each cross, an additional identical pollination was performed on each flower after the stigma re-opened (10 minutes to 2 hours later). For intraspecific crosses, individuals were chosen from independent maternal lines to minimize issues related to inbreeding. A minimum of 20 interspecific, and 20 intraspecific crosses were performed for each species as maternal parent.

Seeds were collected prior to dehiscence and counted. Two separate analyses of seed set were conducted. First, within each class of cross, success or failure of a pollination (either resulting in some offspring or zero offspring) was treated as a categorical variable and analyzed using contingency table analysis in JMP version 8.0. Among crosses that did produce offspring, seed number was analyzed as a continuous variable in a Mann Whitney U test in which the number of seeds produced in

interspecific crosses was compared to intraspecific seed set. This comparison was made within each species such that the intraspecific seed set of species A was compared to the interspecific cross between species A and B in which species A was the maternal parent.

Individual values of reproductive isolation were calculated for both the failure analysis and the seed set analysis using equation RI_5 from Chapter 1:

$$RI = 1 - 2\left(\frac{H}{C + H}\right)$$

In the failure analysis, the success rate of interspecific cross was used as the H term and the success rate of intraspecific crosses was the C term. Average seed set from interspecific (H) and intraspecific (C) crosses were used directly in this equation. The two RI values were combined to produce a composite seed set isolation estimate. Because differences in seed set between cross types could result from either gametic interactions or from early zygote abortion, these data can not unambiguously be classified as either pre- or postzygotic. Therefore this form of isolation is treated separately from other measures of intrinsic postzygotic isolation.

Intrinsic postzygotic isolation - Fitness

Intrinsic postzygotic isolation was measured by examining hybrid fitness relative to each parent species. In the growout described above in the temporal isolation section, seeds were monitored for germination, and an overall rate of germination of hybrids versus parents was compared. For a given species pair, both parents and interspecific offspring from both directions of cross were sown on agar and placed randomly in a growth chamber. Most species were represented by at least 200 seeds per genotypic class

(mean of 210), but due to seed limitation, a few classes were represented by as little as 64 seeds. Because seed germination was highly variable among lines, germination data were pooled into each genotypic class (P1, F1(P1), F1(P2), P2) for analysis. Fitness was set relative to 1.0 within each comparison in order to facilitate analyses across groups and life history stages.

Of the seeds from the above growouts that germinated, each seedling was transplanted and monitored for flowering (described in temporal isolation section). Within each species pair, survival rate was calculated for each genotypic class by comparing the number of individuals that flowered to the total number of seeds that germinated. As before, fitness was set relative to 1.0. Of the individuals that survived until flowering, the total number of flowers produced throughout the lifetime of each individual was recorded. This measure of relative hybrid fitness was also relativized to 1.0. A combined metric of relative hybrid fitness was obtained by multiplying across each of these three life history stages (Ramsey et al. 2003). This combined hybrid fitness estimate was used in equation RI_5 to obtain estimates of reproductive isolation based on the relative fitness of hybrids to parents. As before, all comparisons were made between a parent species, and the fitness of hybrids produced when that species was the maternal parent in the interspecific crosses.

Intrinsic postzygotic isolation - Fertility

Intrinsic postzygotic isolation was also estimated through its effect on male fertility with pollen viability assays. For each genotypic class (P1, F1(P1), F1(P2), P2), a minimum of 8 individuals from independent maternal lines were assayed. Each individual

was measured 4 times, and these repeated measures were averaged to increase confidence in individual estimates. Flowers were marked in the bud stage, and anthers were collected on the first day of dehiscence. Pollen fertility was measured using a modified standard viability assay (Kearns and Inouye 1993). Anthers were placed in a microcentrifuge tube containing a solution with an optimized sucrose concentration specific to each species (ranging from 10-25% sucrose; Sobel, unpublished). In addition, the solution contained 100mg/L boric acid, 300mg/L calcium nitrate, 200mg/L magnesium sulfate heptahydrate, and 100mg/L potassium nitrate. The tube was vortexed briefly to dislodge dehiscent pollen from the anthers and the solution was incubated at room temperature for approximately 2 hours. Two hundred pollen grains were counted per sample, and the number of pollen grains that had germinated was used as an indication of pollen viability.

As with previous analyses of intrinsic postzygotic isolation, pollen viability was compared between a parent of a cross and the interspecific offspring from crosses where that parent was maternal. Mann-Whitney U tests were performed for all parent hybrid combination, and the average number of viable pollen grains from hybrids was used as the H term while the average number of viable pollen grains from parents was used as the C term in equation RI_5 from Chapter 1.

Calculation of total reproductive isolation

The linear sequential method of estimating total reproductive isolation that was introduced in Coyne and Orr (1989) and expanded in Ramsey et al. (2003) was employed to calculate both the total reproductive isolation experienced by each pair of species and the relative contribution of each barrier to total reproductive isolation. By this method,

the strength of each individual barrier is essentially discounted by the gene flow that has already been interrupted by any barriers that act previously within the life history of the organism. The relative contribution of any barrier is therefore the amount of gene flow that is actually prevented by the action of that barrier. For the first barrier that acts (ecogeographic isolation), the individual strength is equivalent to its relative contribution. For the second barrier to act (temporal isolation), the relative contribution is the individual strength of that barrier multiplied by the remaining potential gene flow after the first barrier has acted $(1 - RI_1) * RI_2$. This process is repeated through all barriers in the analysis in the following order: 1) ecogeographic isolation, 2) temporal isolation (from collection records), 3) seed set isolation, 4) intrinsic postzygotic isolation – fitness, 5) intrinsic postzygotic isolation – fertility.

Relationship between genetic distance and strength of isolation

Genetic distance was calculated from molecular sequence data used to construct the phylogeny in Beardsley et al. (2004). Loci included in the analysis were *trnL/F*, ITS, and ETS, and the distance matrix was extracted from the maximum likelihood phylogenetic analysis conducted in PHYLIP (Felsenstein 2005) presented in Chapter 2 (Figure 2.1). Genetic distance was used as a proxy for time to evaluate the rate at which various forms of isolation and total reproductive isolation evolved. While there are too few data points to provide significant relationships by linear or logistic regression, examining the few data available may still be informative to patterns of reproductive isolation evolution.

Results

Temporal isolation

Differences in flowering time as calculated by differences in collection dates in the Jepson collection indicate that flowering times typically differ moderately between species of *Mimulus* (Figure 3.2). Only 2 of the 9 species pairs did not show significant differences by Mann-Whitney U test- *M. cusickii* / *M. nanus* (Figure 3.2 E; $p = 0.074$) and *M. gracilipes* / *M. palmeri* (Figure 3.2 H; $p = 0.188$). In both of these species pairs, one member is extremely rare, and therefore, the analysis was conducted on a relatively small sample size. The magnitudes of these differences in flowering time results in moderately strong reproductive isolation across the species pairs, with an average of 0.246 (Table 3.2).

Common garden experiments reveal that much of the temporal isolation experienced by species pairs does not arise due to intrinsic genetic differences between species. Differences in flowering time between species are presented in Figure 3.3. Several species pairs exhibit significantly different average days to flowering (*M. brevipes/johnstonii*, Figure 3.3C; *M. constrictus/whitneyi*; Figure 3.3D; *M. cusickii/nanus*, Figure 3.3E; *M. floribundus/norrisii*, Figure 3.3G; *M. gracilipes/palmeri*, Figure 3.3H; and *M. jungermannioides/washingtonensis*, Figure 3.3I). While these differences were statistically significant, their biological significance is minimal. Even when a member of a species pair initiated flowering earlier than the alternate species, the two populations were co-flowering in the greenhouse for the majority of their lifespan. This resulted in very modest estimates of reproductive isolation for each of these species pairs, ranging from 0.004 to 0.092, with a mean of 0.052 (Table 3.2).

Several species pairs showed reversed patterns of flowering times in the greenhouse and collection data. For example, in the Jepson collection records, *M. constrictus* flowers significantly earlier (average of June 8th) than *M. whitneyi* (average of July 17) (Figure 3.2D). This difference results in a composite strength of isolation between the two species of 0.37. However, in the greenhouse, *M. whitneyi* flowers significantly earlier (mean days to flower 46.5 ± 12.0) than *M. constrictus* (63.6 ± 11.0) (Figure 3.3D). This pattern was also seen in *M. brevipes/johnstonii*.

Postmating (seed set)

Patterns of seed set differences between experimental intra- and interspecific crosses varied highly across species pairs (Figure 3.4). Failure rates (instances of pollinations resulting in zero offspring) did not differ significantly in most crosses (Figure 3.4). However, in the species pairs *M. constrictus/whitneyi* and *M. cusickii/nanus*, interspecific crosses were more likely to fail than intraspecific crosses. This was consistent in both directions of the cross for each species pair. *M. constrictus* had a failure rate of 56.5% in interspecific crosses, while only 20.6% of intraspecific pollinations failed (Figure 3.4A; $N=80$, $\chi^2=10.4$, $p=0.0012$). Similarly, in the reciprocal cross, *M. whitneyi* showed significantly higher failure rates in interspecific crosses (79.8%) versus intraspecific crosses (37.8%) ($N=194$, $\chi^2=35.6$, $p<0.0001$). This results in moderately strong reproductive isolation for this species pair with a composite metric of 0.40. In the other species pair with significant differences in failure rates, *M. cusickii* sets fails in interspecific crosses with *M. nanus* 61.0% of the time, while intraspecific crosses fail only 12.2% of the time (Figure 3.4 E; $N=82$, $\chi^2=21.0$, $p<0.0001$). In the

reciprocal cross, *M. nanus* interspecific crosses fail 88.2% of the time, while they fail 10.3% of the time in intraspecific crosses ($N=90$, $\chi^2=54.2$, $p<0.0001$).

Quantitative differences among fruits that bore seeds were also highly variable across species pairs (Figure 3.4). For every species pair at least one member of the pair exhibited significant differences in seed set between inter- and intraspecific crosses. 6 of the 18 species showed no differences between the cross types: *M. angustatus* (Figure 3.4A), *M. filicaulis* (Figure 3.4B), *M. johnstonii* (Figure 3.4C), *M. nanus* (Figure 3.4E), *M. gracilipes* (Figure 3.4H), and *M. washingtonensis* (Figure 3.4I). *M. nanus* should be viewed with caution however, as the high failure rate of interspecific crosses led to very few fruits with viable seeds to count. The remaining species showed significant differences in seed set between the cross types falling into two categories: A) interspecific crosses yield fewer seeds than intraspecific (isolation) and B) interspecific crosses yielding more seeds than intraspecific (heterosis). Species falling into the isolation category include: *M. angustatus*, *M. bicolor*, *M. brevipes*, *M. constrictus*, *M. whitneyi*, *M. cusickii*, *M. douglasii*, *M. kelloggii*, *M. floribundus*, *M. norrisii*, *M. palmeri*, and *M. jungermannioides* (Figure 3.4). Only *M. pulchellus* exhibited heterosis (Figure 3.4A); however, *M. filicaulis* and *M. washingtonensis* also trended in this direction.

Reproductive isolation estimated from the combined effects of pollination failure and quantitative differences in seed set are presented in Table 3.2. While many species pairs exhibited differences in seed set, overall isolation was relatively low, with an average RI value of 0.181 across all pairs. Two species pairs exhibited strong isolation by seed set differences: *M. constrictus/whitneyi* (RI composite = 0.664) and *M. cusickii/nanus* (RI composite = 0.708).

Intrinsic postzygotic isolation – Viability

Three components of relative hybrid fitness were measured to assess intrinsic postzygotic isolation due to viability. Table 3.3 gives provides germination rates, survival rates, and numbers of flowers produced set relative to 1 for each direction of a cross. In most cases, offspring of intraspecific crosses germinated at higher rates than offspring of interspecific crosses. The exceptions included *M. angustatus*, in which intraspecific offspring germinated at a rate of 23.9% of hybrids and *M. floribundus* in which intraspecific offspring germinated at a rate of 23.6% relative to hybrid offspring. Several instances of moderately strong reproductive isolation by germination rate differences exist, including *M. pulchellus*, *M. bicolor*, *M. constrictus*, *M. cusickii*, *M. norrisii*, *M. palmeri*, and *M. jungermannioides* (Table 3.3). This form of isolation also gives several instances of complete reproductive isolation, as some classes of seeds did not germinate at all. These include hybrid progeny of: *M. brevipes*, *M. johnstonii*, *M. whitneyi*, and *M. nanus*. For these 4 species, further measures of reproductive isolation are therefore absent because no hybrids were produced to measure.

Survival and flowering rates also differed considerably among the remaining classes of offspring (Table 3.3). In particular, despite reasonably high germination rates, no hybrid offspring *M. cusickii* maternal crosses survived to adulthood. *M. norrisii* and *M. washingtonensis* hybrids also showed significant decreases in F1 survival to flowering despite reasonable germination rates. Each of these three fitness components were multiplied together (as in Ramsey et al. 2003) to produce a composite metric of relative hybrid fitness, and reproductive isolation was calculated from this value. A broad range

of reproductive isolation values were obtained, ranging from the above-mentioned instances of complete isolation to instances of strong heterosis. In an example of the latter, *M. angustatus* hybrids consistently outperformed conspecific offspring in all three categories, resulting in an isolation coefficient of -0.826. *M. floribundus* also showed evidence of heterosis, with a total combined metric of -0.708 for isolation. Both of these values indicate that the relative fitness of hybrids would actually facilitate gene flow rather than limit it.

Intrinsic postzygotic isolation - Fertility

Relative fertility of pollen reveals that several species experience reduced F1 fertility relative to parent fertility (Figure 3.5). *M. angustatus*, *M. bicolor*, *M. constrictus*, *M. floribundus*, and *M. jungermannioides* all show evidence of significant declines in the fertility of their interspecific offspring. In contrast, only one species, *M. pulchellus*, exhibits a strong significant increase in pollen viability in hybrids. One species pair, *M. gracilipes* / *M. palmeri* shows marginally significant differences with *M. gracilipes* showing a decline in hybrid fertility, and *M. palmeri* showing an increase. The isolation that results from these differences in fertility mostly results in insignificant amounts of reproductive isolation, with an average strength of 0.082 (Table 3.2).

Total reproductive isolation

Ecogeographic isolation data presented in Chapter 2 were added to these measures of individual barriers to compare the strengths of individual barriers. Due to a lack of collection records available, ecogeographic isolation and temporal isolation were

not calculated for *M. jungermannioides* / *M. washingtonensis*. Therefore, only the later acting barriers can be included in comparisons for this species pair. Figure 3.6 shows the distribution of estimates for each form of reproductive isolation. One-way ANOVA and posthoc comparisons reveal that ecogeographic isolation is significantly stronger than either form of temporal isolation, seed set isolation, and intrinsic postzygotic isolation by fertility. It is indistinguishable from gametic isolation (due to small sample size of pollen competition experiments) and intrinsic postzygotic isolation due to relative fitness. No other barriers are significantly different from each other by this analysis.

Calculations of total reproductive isolation are given in Table 3.2. Three species show particularly low values of isolation: *M. kelloggii* has an isolation value of 0.539 from *M. douglasii*, *M. filicaulis* has an isolation of 0.248 from *M. bicolor*, and *M. gracilipes* has isolation of 0.274 from *M. palmeri*. The remaining have an isolation value of at least 0.8 (overall mean is $0.847 \pm .26$), suggesting that the majority of isolation has been accounted for in most species pairs.

Figure 3.7 demonstrates the difference between the measured individual barrier strength and the relative contribution of each barrier to total isolation. Due to the sequential nature of isolating barriers, early acting barriers have an increased potential to affect gene flow between species. Figure 3.7A shows the individual barrier strengths broken down by species pair. The individual isolation values for each species has been averaged together to create a single value of isolation per pair. There is considerable variation in which barriers are strongest in particular pairs (as also shown in Table 3.2); however, ecogeographic isolation and intrinsic postzygotic isolation due to relative hybrid fitness commonly have the highest values.

In calculating the relative contributions of each, the sequential nature of barriers leads to a reduction in the relative strength of late acting barriers across all species (Table 3.4). In all species pairs where ecogeographic isolation was measured (all but *M. jungermannioides/washingtonensis*), it is the strongest relative contributor to isolation. Even cases where individual barriers were complete, such as intrinsic postzygotic isolation due to fitness in *M. brevipes/johnstonii* or *M. cusickii/nanus*, relatively little gene flow ever reaches those barriers to give them the opportunity to act (Figure 3.7B)

Relationship between isolation and genetic distance

While the species pairs presented in this study all represent recent cases of speciation, sufficient variation in genetic distance occurs to examine the relationship between genetic distance and values of reproductive isolation. Figure 3.8 shows the relationship between sequence divergence and total reproductive isolation, total prezygotic, and total postzygotic isolation. While only suggestive, it appears that prezygotic isolation mediated through differences in habitat (both ecogeographic isolation and temporal isolation) are strongest in early stages of divergence, while postzygotic isolation evolves after more time has passed.

Discussion

Uncovering the relative importance of different forms of isolation is a major goal of speciation research (Coyne and Orr 2004); however, relatively few data are available to speculate on this important topic. The presented study offers a unique approach to answer the question of which forms of isolation are responsible for isolating species by

combining the comparative approach of Coyne & Orr (1989; 1997) with a detailed study of isolation between species pairs known to be recently speciated (e.g. Kay 2006; Matsubayashi and Katakura 2009; Ramsey et al. 2003; Chari and Wilson 2001). By selecting recently diverged species pairs randomly from across a phylogenetic hypothesis, much of the bias that plagues studies of reproductive isolation was avoided (Sobel and Randle 2009). In addition, choosing recently diverged species pairs helps assure that the strength of barriers measured today are as similar as possible to the strength at the moment of speciation. This sampling method also avoids the need for phylogenetic corrections on data (Felsenstein 1985), which can obscure patterns of isolation evolution.

By broadly sampling across the genus and placing these results into the linear sequential framework of Coyne and Orr (1989; 1997) and Ramsey et al. (2003), a general picture of the evolution of reproductive isolation in the genus *Mimulus* emerges. Geographic isolation clearly plays a prominent role in speciation (Barracough and Vogler 2000; Fitzpatrick and Turelli 2006), and its sequence within life history (first to act) virtually assures that geographic differences will play a large role in limiting gene flow. However, by considering only the fraction of geographic isolation that is based on intrinsic genetic differences (Sobel et al. 2010), this work confirms that closely related species are separated not only by simple geography (Jordan 1908), but are also commonly separated by ecological differences (Stebbins 1950). This confirms that geographic differences observed in these species are truly ecogeographic differences that can legitimately be included as a barrier within the biological species concept.

Previous work in the genus compares favorably with the findings presented here. For example, Ramsey et al. (2003) studied the closely related species pair *Mimulus cardinalis* and *M. lewisii*. This species pair was used in Chapter 2 in order to apply the ecological niche modeling technique described there to the problem of estimating the strength of ecogeographic isolation. This verified that ecogeographic isolation was indeed very strong between these two species, and similarly to many species pairs in this study, this was by far the biggest contributor to total reproductive isolation despite some low levels of intrinsic postzygotic isolation that were detected. This study also revealed that this species pair experiences substantial reproductive isolation by differences in pollinators, resulting in isolation of 0.976 in sympatry and 0.403 relative contribution to total isolation. While substantial pollinator data for this study were not collected due to unfortunately timed drought events in the Sierra, anecdotal evidence suggests that many of the species pairs in the present study share primary pollinators (Sobel, unpublished). Therefore, despite abundant work on the genetics and ecology of pollinator isolation in this species pair (Bradshaw and Schemske 2003; Bradshaw et al. 1995; Ramsey et al. 2003; Schemske and Bradshaw 1999), this form of isolation is most likely only mildly important in the majority of speciation events in the genus.

Unfortunately, very few other studies of reproductive isolation include an ecogeographic isolation aspect, limiting the potential to compare these findings with other measures. For other forms of isolation, previous studies have also shown similar findings to those presented here. The work of Martin & Willis (2007) provides an example of reproductive isolation in a different pair of *Mimulus* species, *M. guttatus* and *M. nasutus*. In this study, among other factors, these two species are moderately isolated

by differences in flowering time, which the authors speculate is driven mainly by habitat differences. Similarly, Lowry et al. (2008) found that inland and coastal races of *Mimulus guttatus* are strongly isolated by a habitat mediated temporal isolation.

The temporal isolation measured in this study was almost completely mediated by differences in habitat. For example, *Mimulus constrictus* flowers significantly earlier in the field than *M. whitneyi*, leading to moderately strong reproductive isolation. Interestingly, the pattern of flowering is reversed in the common garden conditions of the greenhouse, with *M. whitneyi* flowering earlier than *M. constrictus* (though the isolation that results from this difference is modest). This pattern of flowering time is consistent with the biology of species that occupy distinct altitudes. *M. constrictus* is found at mid-range elevations in the southern Sierra between approximately 800-2000m (2600-6500ft), and *M. whitneyi* grows between 1500-3000m (4900-9800ft). The difference in flowering times observed in the collection records (Figure 3.2D) occurs because suitable habitat for the high elevation *M. whitneyi* is often still under snow cover when *M. constrictus* starts to germinate and flower. However, at high altitudes, growing seasons are constricted, and many species are expected to be under selection to reproduce earlier. Therefore, it is perhaps unsurprising that the later flowering *M. whitneyi* has a shorter duration between germination and flowering than *M. constrictus* when grown in the greenhouse (Figure 3.3D).

In other systems, Kay (2006) found that the highest contributions to total isolation in two species of Neotropical spiral ginger came from geographic isolation and mechanical floral isolation. While mechanical isolation was not measured in this study, the similarity of floral shape and size between closely related species suggests this too is

unimportant in *Mimulus*. In a study of the isolation between diploid and tetraploid fireweeds in the genus *Chamerion*, Husband & Sabara (2004) found that geography also plays a significant role despite the strong postzygotic isolation involved in the ploidy shift.

Previous comparative work is especially difficult to relate to the presented work because none of the studies published thus far have included an estimate of how ecogeographic isolation affects speciation, and very few include more than a single form of isolation. Comparative studies that include at least one prezygotic barrier and one postzygotic barrier include studies of *Drosophila* by Coyne & Orr (1989; 1997), *Etheostoma* fish by Mendelson (2003), and three genera of plants by Moyle et al. (2004). Both Coyne & Orr (1989; 1997) and Mendelson (2003) found that prezygotic isolation evolves faster than postzygotic, while Moyle et al. (2004) showed no significant difference between the two types of isolation. In addition, the genetic distances covered by these three studies extends far beyond recent speciation events, limiting their applicability to the presented work. Given that ecogeographic isolation has been very strong in nearly every study that has ever included it, adding some estimate of this form of isolation to these comparative studies would greatly enhance our understanding of speciation.

Definitions of importance

While in the presented work, the working definition of the most important barriers to speciation has been the forms that prevent the most gene flow at the time of speciation, other definitions of importance could clearly be used. For example, many would argue

that intrinsic postzygotic isolation deserves its historical prominence (Coyne and Orr 2004) because it has the potential to be more permanent than prezygotic forms of reproductive isolation such as ecogeographic. This may be true once isolation is complete; once an isolation of 1 is achieved by intrinsic postzygotic means, it is unlikely that this form of isolation will decay and be reduced in the future. On the other hand, ecogeographic isolation may go to completion, but then be reduced as species ranges change through climate changes and adaptation. Therefore, complete isolation by these two forms differ in their permanence, but incomplete isolation may have the opposite impact. Consider an example where two allopatric populations evolve some incomplete amount of intrinsic postzygotic isolation without any adaptive divergence due to habitat. If the two populations come back into contact and do not differ in ecological niche, selection may favor the elimination of segregating alleles involved in postzygotic isolation in order to increase the number of available mates rather than promoting enhancement of prezygotic isolation through reinforcement. Under this scenario, postzygotic isolation could be just as ephemeral as ecogeographic isolation if not more so.

Another definition of importance of isolating barriers may hinge on considering only those barriers that completely prevent gene flow on an individual basis. For example, in the presented work, intrinsic postzygotic isolation based on relative hybrid fitness showed complete reproductive isolation for two species pairs, *M. cusickii/nanus* and *M. brevipes/johnstonii*. One somewhat counterintuitive result of using the linear sequential combination approach to estimating the strength of total isolation is that complete reproductive isolation ($RI=1$) can only fully be achieved if a single individual

barrier is equal to 1. Therefore, while the relative contribution of postzygotic isolation to total isolation in *M. cusickii/nanus* and *M. brevipes/johnstonii* is relatively low due to previously acting barriers, only these two species pairs achieved complete total isolation (Table 3.2) due to the action of intrinsic postzygotic isolation in preventing all gene flow.

Unmeasured forms of isolation and error in estimation

While most of the isolation potentially experienced by species pairs was captured by this study (mean total RI of 0.85 ± 0.26 , Table 3.2), several important barriers were not measured, and could account for the remaining isolation. For example, microspatial habitat isolation can play an important role in limiting gene flow (e.g. Lynch 1978; Howard et al. 1997; Rand and Harrison 1989). The ecogeographic isolation estimates presented in Chapter 2 measure habitat based isolation on a broad geographic scale, but will miss substantial local differences in habitat that vary at spatial scales below the grid size of the analysis. In the field, it is extremely uncommon to find two closely related species of *Mimulus* at the same site, unless one member of the pair is highly selfing (Sobel, personal observation). Despite extensive searching in the field, none of the species in this study were ever found at a sympatric site despite some broad overlap in geographic ranges in the ecogeographic analysis. Therefore, it is quite possible that this form of isolation both acts in nature and has a large magnitude.

Another important form of isolation related to habitat that was not measured is extrinsic postzygotic isolation. Extrinsic postzygotic isolation is thought to be important in stickleback fish (Hatfield and Schluter 1999), and could play a significant role in limiting gene flow between species that can and do hybridize. The same traits that confer

ecogeographic isolation could also lead to extrinsic postzygotic isolation if/when hybrids are formed. In the course of performing this study, several thousand field collected seeds were grown, and no instances of obvious F1 hybrids were ever recorded. Given that these seeds were grown under the benign common garden environment of the greenhouse, the lack of hybrids suggests that extrinsic postzygotic isolation is probably not commonly important, and lends support to the idea that there is some portion of prezygotic isolation that has not been measured.

One possible explanation is that the ecogeographic isolation estimated by ecological niche modeling is underestimating the true strength. Species that might be susceptible to underestimation would be those whose taxonomy is easily confused. For example, specimens of *Mimulus whitneyi* are commonly misidentified as its sister species *M. constrictus* (Sobel. pers. obs.). While efforts were made to correct for this problem, it is possible that the some of the herbaria records used to construct the niche models created more overlap than species actual experience because of these misidentifications. This limitation of herbarium/museum records can hinder efforts to reconstruct species ranges, and points to the need for competent systematic work in maintaining collections (Graham et al. 2004).

Another source of error in estimates of total reproductive isolation may arise from difficulty in obtaining accurate estimates of isolation. One form of isolation could conceivably have significant error is the germination rate component of relative hybrid fitness. Many species of *Mimulus*, especially those in sections *Oenoe* and *Eunanus*, have very strong seed dormancy, which requires treatment with gibberellic acid to break. For each species pair, an appropriate concentration was identified such that both parents and

both classes of hybrids could be germinated at a common concentration of gibberellic acid. However, in some species, especially in the *Eunanus* section (*M. brevipes/johnstonii*, *M. cusickii/nanus*, and *M. constrictus/whitneyi*) seed germination rates were chronically low and often failed unexpectedly. Germination is also a difficult measure of hybrid fitness because high germination rates in hybrids (such as in *M. angustatus* hybrids; Table 3.3) may be an indication of short-circuiting an adaptive dormancy trait rather than indicative of higher fitness.

Three species showed markedly low total postzygotic isolation compared to the remaining species. Interestingly, biological aspects may be involved in at least 2 of the 3. *M. kelloggii* was estimated as being isolated from *M. douglasii* at a strength of 0.539. Isolation in the opposite direction (*M. douglasii* in reference to *M. kelloggii*) is also not complete at 0.879. One possible explanation for this species pair's low value of isolation despite very clear morphological differentiation is that *M. douglasii* is highly selfing, producing both autonomous seeds when unmanipulated and producing fully cleistogamous flowers under certain conditions. While selfing is often treated as a form of isolation (e.g. Martin and Willis 2007), there are considerable difficulties in using this as a form of isolation under the biological species concept because selfing individuals can be just as isolated from other members of their own population as they are from any other population (Coyne and Orr 2004). Therefore, estimating an isolation coefficient that has any meaning for interruption of gene flow is problematic. While this was probably only a problem in this species pair, many small-flowered and presumably highly selfing *Mimulus* species exist, and methods for dealing with these species is necessary to gain a complete picture of speciation in the genus.

In another example of low isolation, *M. gracilipes* showed relatively low isolation from its closest relative, *M. palmeri*. One possible explanation is that temporal isolation may occur between these two species on a different time scale than what was measured here. *M. gracilipes* is a very rare inhabitant of mid-elevations of the Sierra foothills, and is extremely sensitive to variation in precipitation. The population that was used in this study was located in 2005, an El Niño year. In 2005 it bloomed densely, forming a dense carpet of flowers (Sobel, personal observation). However, the same site was visited every year between 2006 and 2009, and not a single flowering individual was found in any of those years. If this species is restricted to only surviving to flowering in years of extremely high precipitation, temporal isolation may actually be much higher than estimated.

An alternative to the possibility of unmeasured barriers limiting gene flow is that some species pairs may not actually be biological species. For example, *Mimulus filicaulis* appears to have very limited isolation from *M. bicolor* despite strikingly different coloration of the corolla. The habitat of *M. filicaulis* appears to be reasonably suitable to *M. bicolor* according to the ecogeographic isolation estimate (Chapter 2, Figure 2.5C), and the two species are highly interfertile. While it is possible that aspects of habitat could have been missed (such as micro-edaphic differences) by the ecological niche modeling technique, it is also possible that these two species are separated due to completely historical processes. Geographic ranges are the result of both ecological and historical processes (Endler 1982; Sobel et al. 2010), and the ecogeographic isolation estimated here only measures the portion of geographic isolation that is based on ecological differences between species. Therefore, these two species may occupy

different geographic ranges not due to biological reasons, but mainly due to stochastic processes of history. If true, these two taxonomic species may collapse if dispersal of the more common *M. bicolor* occurs into the range of *M. filicaulis*. These results clearly have ramifications for appropriate conservation strategies to ensure that the *M. filicaulis* taxonomic species is not lost to gene flow.

Summary

The divergence of species often requires the formation of multiple forms of reproductive isolation, and revealing which forms of isolation contribute most to cessation of gene flow is a major goal of speciation research. The presented study examined reproductive isolation in recently diverged species across the wildflower genus *Mimulus*. The source of isolation with the highest relative contribution to total isolation was ecogeographic isolation, owing mainly to the sequential nature of isolating barriers and that it is the first barrier to act on gene flow. Ecogeographic isolation was also commonly strong on an individual barrier basis, suggesting that regardless of whether one ascribes to the sequential nature of barriers or not, this barrier deserves to attract further attention from speciation biologists. Other forms of isolation were strong on an individual basis, especially intrinsic postzygotic isolation due to These findings suggest that, consistent with Darwin's (1859) views, understanding adaptive divergence and its direct effects on reproductive isolation is the key to understanding the origin of species.

Table 3.1. *Mimulus* species pairs used in comparative study of reproductive isolation.

Species (arranged in pairs)	Section	Collection location general description	GPS coordinates
<i>M. angustatus</i>	Oenoe	Sierra National Forest, Fresno County, CA	37.038 N, 119.258 W
<i>M. pulchellus</i>		Stanislaus National Forest, Tuolumne County, CA	38 5.915, 120 7.488
<i>M. bicolor</i>	Paradanthus	Sierra National Forest, Fresno County, CA	37.085 N, 119.387 W
<i>M. filicaulis</i>		Stanislaus National Forest, Tuolumne County, CA	37 50.162, 119 50.857
<i>M. brevipes</i>	Eunanus	Los Padres National Forest, Ventura County, CA	34 35.094, 119 15.877
<i>M. johnstonii</i>		Angeles National Forest, Los Angeles County, CA	34 24.117, 117 48.948
<i>M. constrictus</i>	Eunanus	Sequoia National Park, Tulare County, CA	36 26.493, 118 42.974
<i>M. whitneyi</i>		Sequoia National Forest, Fresno County, CA	36 47.864, 118 54.368
<i>M. cusickii</i>	Eunanus	Modoc National Forest, Modoc County, CA	41 28.447, 120 11.171
<i>M. nanus</i>		Lassen National Forest, Lassen County, CA	40 33.701, 121 3.647
<i>M. douglasii</i>	Oenoe	Snell Valley Reserve, Napa County, CA	38 42.002, 122 24.604
<i>M. kelloggii</i>		El Dorado National Forest, El Dorado County, CA	38 52.548, 120 48.956
<i>M. floribundus</i>	Paradanthus	Sequoia National Forest, Fresno County, CA	36 47.864, 118 54.368
<i>M. norrisii</i>		Sequoia National Park, Tulare County, CA	36 29.905, 118 49.241
<i>M. gracilipes</i>	Paradanthus	Sierra National Forest, Fresno County, CA	37 5.083, 119 23.261
<i>M. palmeri</i>		Sequoia National Park, Tulare County, CA	36 26.491, 118 42.974
<i>M. jungermannioides</i>	Paradanthus	Columbia River, Gilliam County, OR	45 44.411, 120 36.587
<i>M. washingtonensis</i>		Ochoco National Forest, Grant County, OR	44 14.090, 119 31.795

Table 3.2. Estimates of the strength of individual reproductive barriers in pairs of species in *Mimulus*. Species for which measurements could not be made are indicated with N/A.

Species	Ecogeographic isolation*	Temporal isolation (collections)	Temporal isolation (common garden)	Seed set isolation	Gametic isolation	Intrinsic postzygotic isolation (fitness)	Intrinsic postzygotic isolation (fertility)	Total
<i>M. angustatus</i>	0.891	0.137	0.004	0.177	0.228	-0.826	0.525	0.933
<i>M. pulchellus</i>	0.620	0.317	0.004	-0.327	0.030	0.583	-0.473	0.789
<i>M. bicolor</i>	0.898	0.458	0.005	0.084	0.241	0.065	0.184	0.961
<i>M. filicaulis</i>	0.218	0.000	0.005	-0.076	0.120	0.247	-0.187	0.248
<i>M. brevipes</i>	0.916	0.317	0.098	0.417	N/A	1.000	N/A	1.000
<i>M. johnstonii</i>	0.886	0.145	0.098	0.106	N/A	1.000	N/A	1.000
<i>M. constrictus</i>	0.654	0.274	0.073	0.601	N/A	0.197	0.197	0.935
<i>M. whitneyi</i>	0.588	0.461	0.073	0.727	N/A	1.000	N/A	1.000
<i>M. cusickii</i>	0.347	0.000	0.073	0.608	N/A	1.000	N/A	1.000
<i>M. nanus</i>	0.815	0.301	0.073	0.808	N/A	1.000	N/A	1.000
<i>M. douglasii</i>	0.631	0.130	0.017	0.459	N/A	0.250	0.068	0.879
<i>M. kelloggii</i>	0.205	0.116	0.017	0.345	N/A	0.020	-0.022	0.539
<i>M. floribundus</i>	0.988	0.682	0.092	0.248	N/A	-0.708	0.753	0.999
<i>M. norrisii</i>	0.356	0.000	0.092	0.176	N/A	1.000	N/A	1.000
<i>M. gracilipes</i>	0.305	0.000	0.091	0.124	N/A	-0.375	0.134	0.274
<i>M. palmeri</i>	0.932	0.593	0.091	0.244	N/A	0.572	-0.159	0.990
<i>M. jungermannioides</i>	N/A	N/A	0.018	0.137	N/A	0.407	0.107	0.543
<i>M. washingtonensis</i>	N/A	N/A	0.018	-0.260	N/A	-0.201	-0.137	-0.722
Mean (st. dev.)	0.641 (0.28)	0.246 (0.22)	0.052 (0.04)	0.181 (0.20)	0.155 (0.099)	0.346 (0.60)	0.082 (0.33)	0.847** (0.26)

* from chapter 2

** mean total does not include *M. jungermannioides* / *M. washingtonensis* due to unmeasured prezygotic barriers in this pair

Table 3.3. Relative parent and F1 viability in *Mimulus* species pairs. For each stage in life history (germination, survival to flowering, and number of lifetime flowers produced), fitness values were set relative to 1 in each direction of a cross between species. Total relative viability was calculated as the product of the independent stages. Instances where earlier stages showed zero hybrid fitness prevented further estimates and N/A is listed.

Species	Germination rate	Survival rate	Flowers produced	Total viability
<i>M. angustatus</i>	0.239	0.751	0.531	0.096
<i>F1 (angu)</i>	1.000	1.000	1.000	1.000
<i>M. pulchellus</i>	1.000	1.000	1.000	1.000
<i>F1 (pulch)</i>	0.556	0.888	0.533	0.263
<i>M. bicolor</i>	1.000	1.000	0.754	0.754
<i>F1 (bico)</i>	0.672	0.985	1.000	0.662
<i>M. filicaulis</i>	1.000	0.936	1.000	0.936
<i>F1 (fili)</i>	0.809	1.000	0.699	0.565
<i>M. brevipes</i>	1.000	1.000	N/A	1.000
<i>F1 (brev)</i>	0.000	0.000	N/A	0.000
<i>M. johnstonii</i>	1.000	1.000	N/A	1.000
<i>F1 (john)</i>	0.000	0.000	N/A	0.000
<i>M. constrictus</i>	1.000	0.694	1.000	0.694
<i>F1 (const)</i>	0.522	1.000	0.891	0.465
<i>M. whitneyi</i>	1.000	1.000	1.000	1.000
<i>F1 (whit)</i>	0.000	0.000	0.000	0.000
<i>M. cusickii</i>	1.000	1.000	N/A	1.000
<i>F1 (cusi)</i>	0.542	0.000	N/A	0.000
<i>M. nanus</i>	1.000	N/A	N/A	1.000
<i>F1 (nanus)</i>	0.000	N/A	N/A	0.000
<i>M. douglasii</i>	1.000	1.000	1.000	1.000
<i>F1 (doug)</i>	0.898	0.866	0.772	0.600
<i>M. kelloggii</i>	0.911	1.000	1.000	0.911
<i>F1 (kell)</i>	1.000	0.967	0.906	0.876
<i>M. floribundus</i>	0.236	1.000	0.702	0.166
<i>F1 (flor)</i>	1.000	0.972	1.000	0.972
<i>M. norrisii</i>	1.000	1.000	1.000	1.000
<i>F1 (nor)</i>	0.195	0.302	0.000	0.000
<i>M. gracilipes</i>	0.907	0.819	0.612	0.454
<i>F1 (grac)</i>	1.000	1.000	1.000	1.000
<i>M. palmeri</i>	1.000	1.000	1.000	1.000
<i>F1 (palm)</i>	0.325	1.000	0.840	0.273
<i>M. jungermannioides</i>	1.000	1.000	0.870	0.870
<i>F1 (junger)</i>	0.460	0.797	1.000	0.367
<i>M. washingtonensis</i>	1.000	0.493	1.000	0.493
<i>F1 (wash)</i>	0.971	1.000	0.763	0.741

Table 3.4. Comparison of individual barrier strength and relative contributions to total reproductive isolation in recently diverged species of *Mimulus*.

	Ecogeographic Isolation	Temporal Isolation	Relative seed set isolation	Postzygotic isolation (fitness)	Postzygotic isolation (fertility)
Average individual strength of barrier	0.641	0.246	0.181	0.346	0.004
Average relative contribution of barrier	0.641	0.088	0.049	0.077	0.001

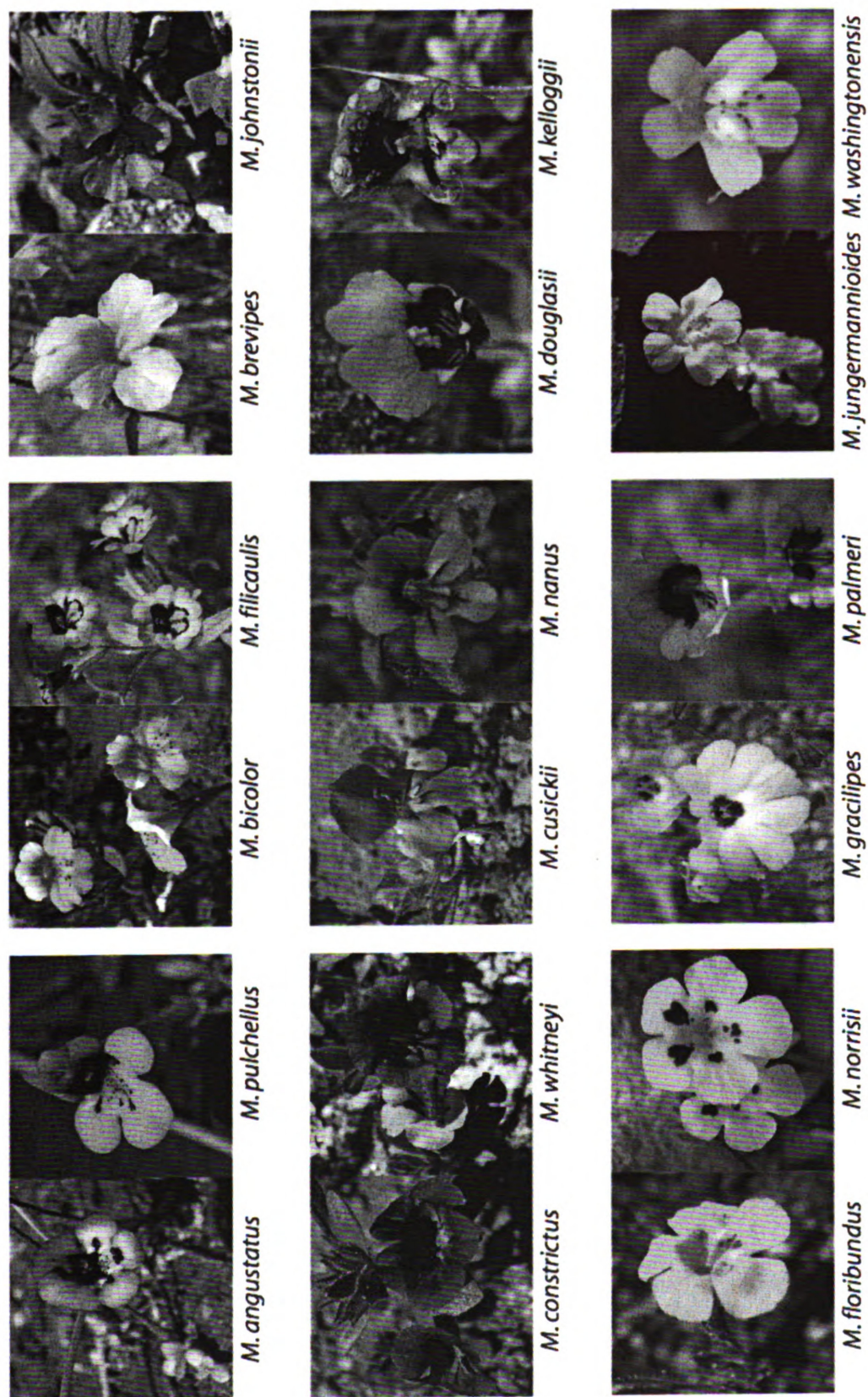


Figure 3.1. Recently diverged pairs of species in the genus *Mimulus* used in this study.

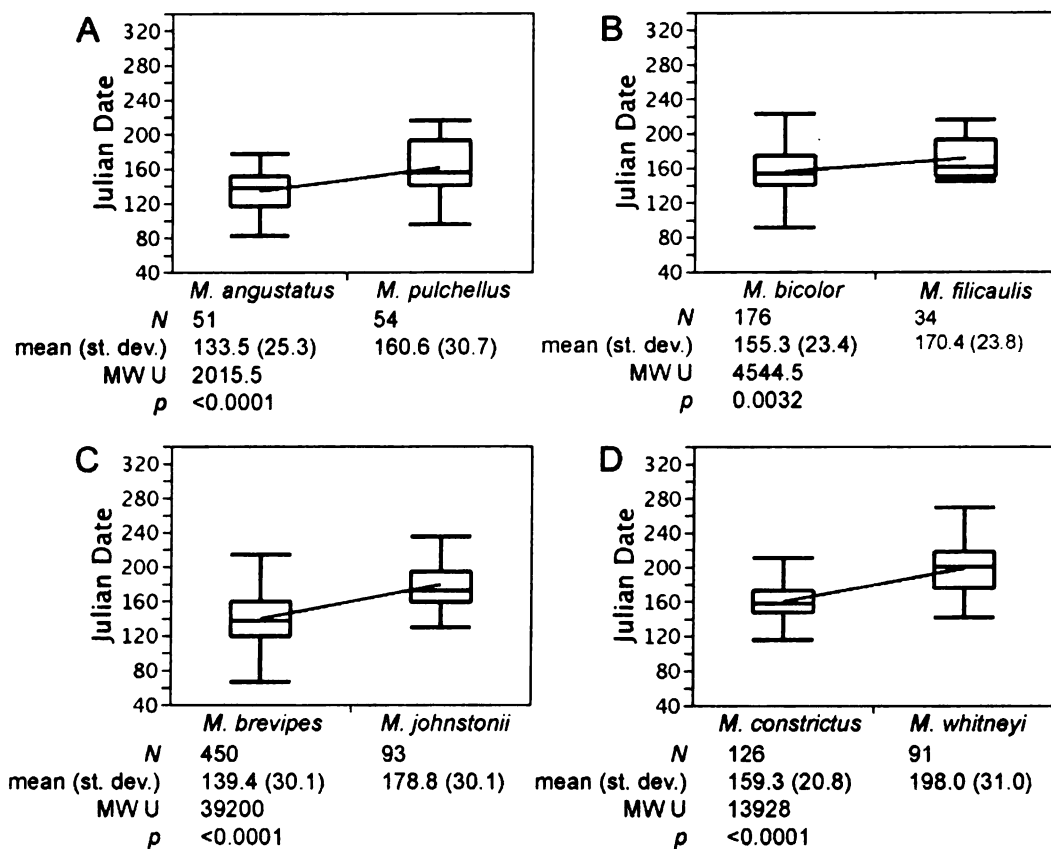


Figure 3.2A-D. Differences in flowering time based on herbarium collection for species pairs *Mimulus angustatus* / *pulchellus*, *M. bicolor* / *filicaulis*, *M. brevipes* / *johnstonii*, *M. constrictus* / *whitneyi*. Specimen collection dates from the Jepson Online Interchange (UC-JEPS2004) were compiled, and differences between species pairs were analyzed by Mann Whitney U test. Boxes represent the 25th, 50th, and 75th percentiles, while the stems show the 10th and 90th percentiles. Means are shown with the connected line between species in a pair. All species pairs show significant differences in average flowering time except *M. gracilipes* / *M. palmeri* (H).

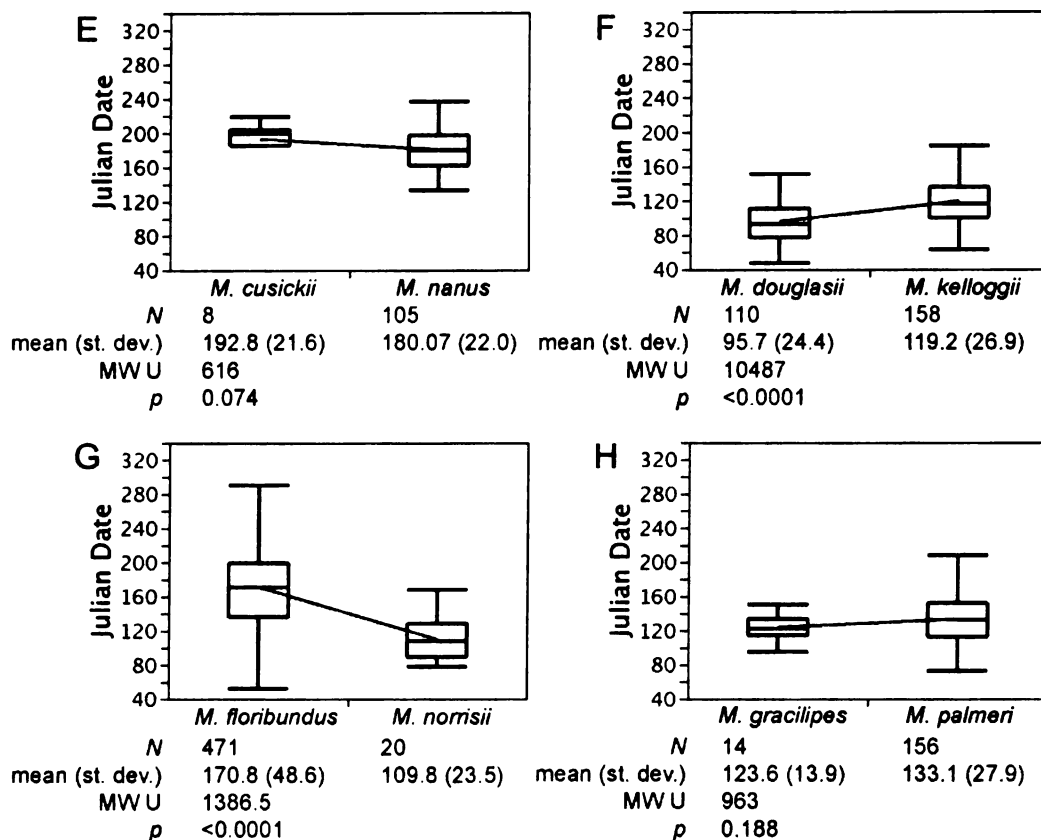


Figure 3.2E-H. Differences in flowering time based on herbarium collection in the genus *Mimulus*. Specimen collection dates from the Jepson Online Interchange (UC-JEPS2004) were compiled, and differences between species pairs were analyzed by Mann Whitney U test. Boxes represent the 25th, 50th, and 75th percentiles, while the stems show the 10th and 90th percentiles. Means are shown with the connected line between species in a pair. All species pairs show significant differences in average flowering time except *M. gracilipes* / *M. palmeri* (H).

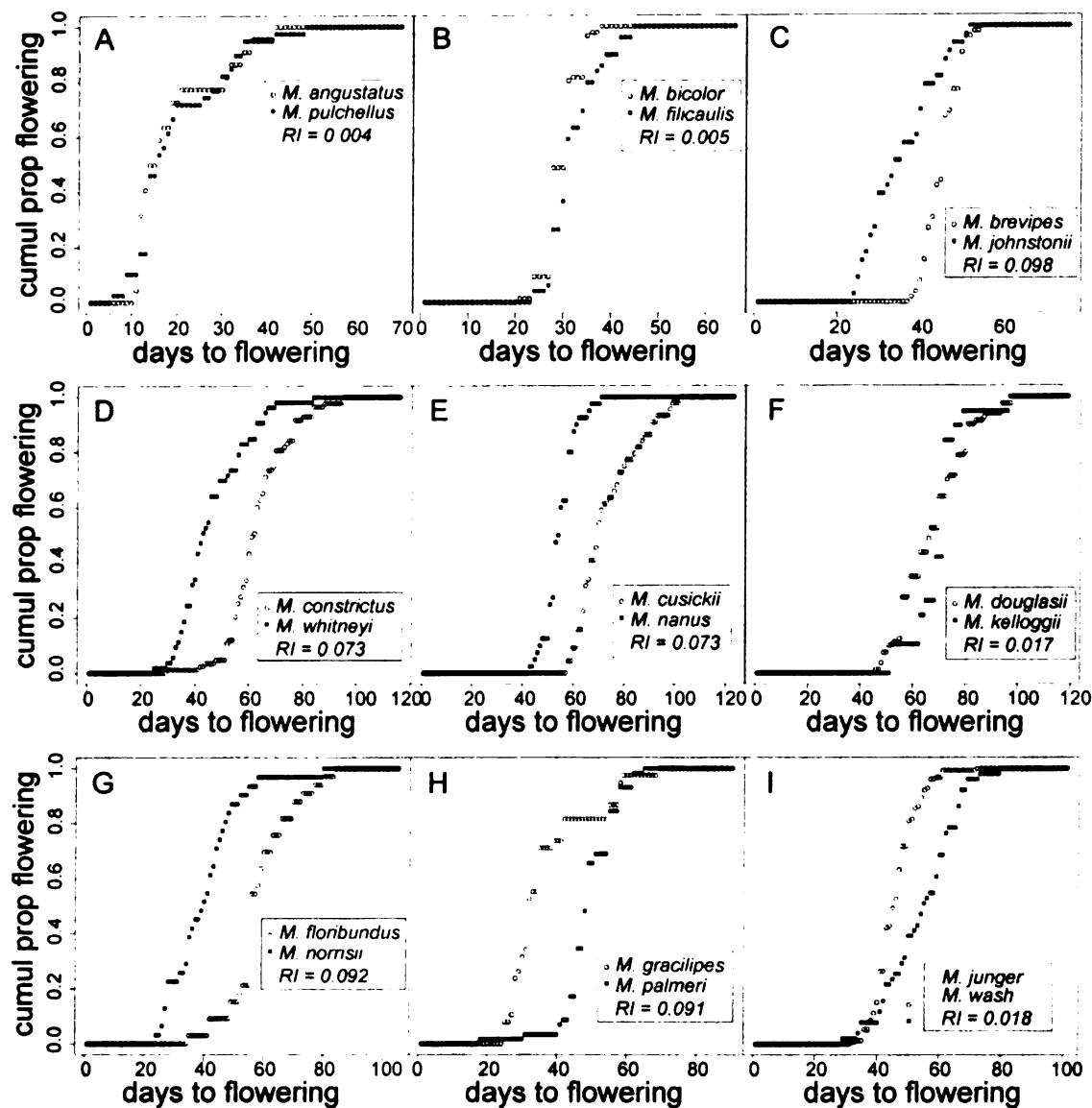


Figure 3.3. Comparisons of flowering time differences between *Mimulus* species under common garden conditions in the greenhouse. For each pair of species in the study, the cumulative proportion of plants flowering is plotted against days from germination to flowering. Statistical comparisons by Mann Whitney U reveals that many of these comparisons are statistically different ($p < 0.05$; C, D, E, G, H, I); however, reproductive isolation that results from these differences is relatively modest.

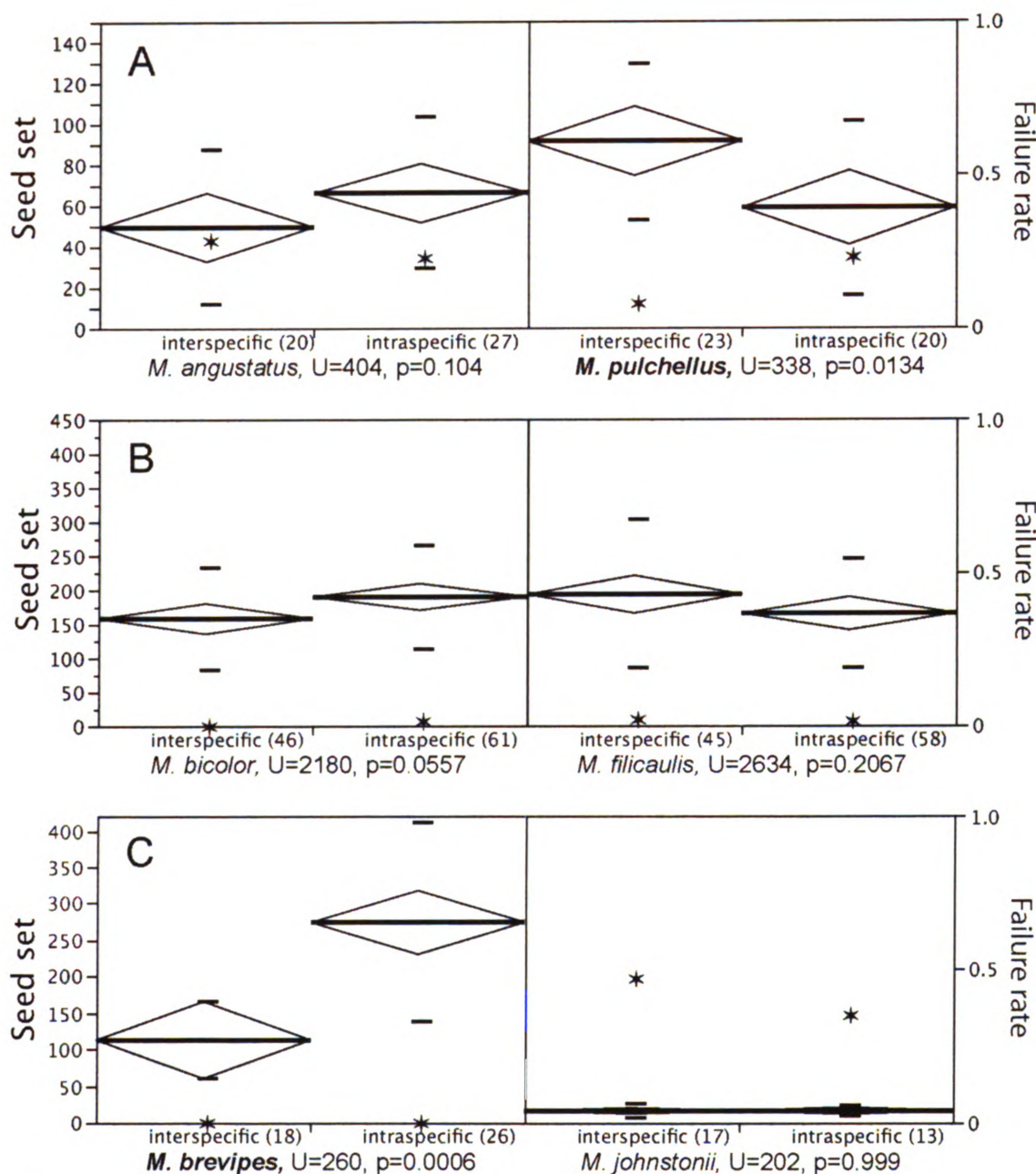


Figure 3.4A-C. Postmating reproductive isolation in the species pairs *Mimulus angustatus* / *pulchellus*, *M. bicolor* / *filicaulis*, and *M. brevipes* / *johnstonii*. Wide horizontal lines represent the average seed set from crosses resulting in progeny for both inter- and intraspecific crosses (left vertical axis). Sample sizes are in parentheses. Narrow horizontal lines indicate standard deviation, and the vertices of diamonds represent 95% confidence intervals. Species names in bold indicate instances of significant differences between inter- and intraspecific seed set, and Mann-Whitney U statistic and p-value are given next to the species name. Failure rate (proportion of crosses resulting in zero progeny) is shown by a star for each cross category (right vertical axis).

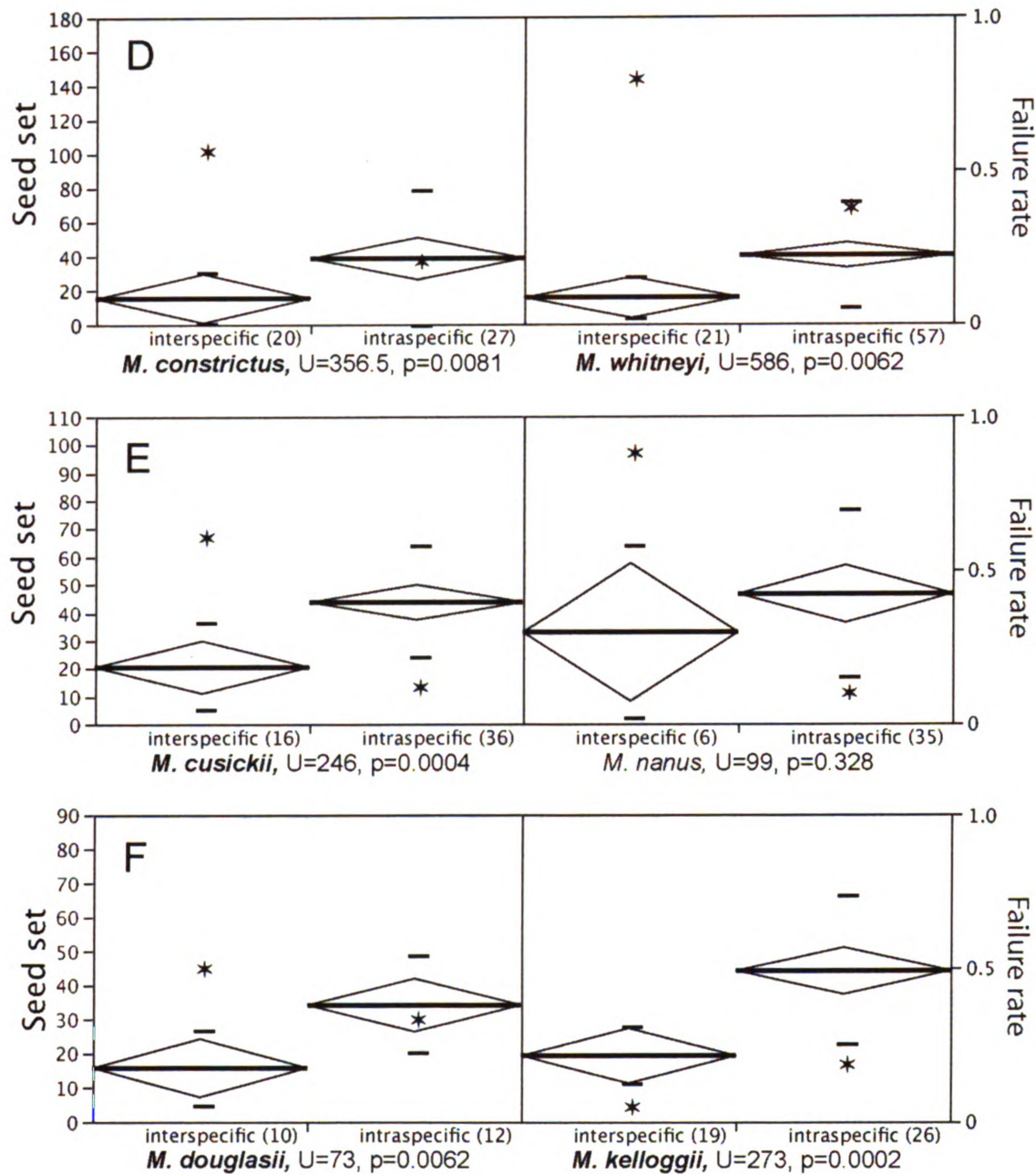


Figure 3.4D-F. Postmating reproductive isolation in the species pairs *Mimulus constrictus* / *whitneyi*, *M. cusickii* / *nanus*, and *M. douglasii* / *kelloggii*. Wide horizontal lines represent the average seed set from crosses resulting in progeny for both inter- and intraspecific crosses (left vertical axis). Sample sizes are in parentheses. Narrow horizontal lines indicate standard deviation, and the vertices of diamonds represent 95% confidence intervals. Species names in bold indicate instances of significant differences between inter- and intraspecific seed set, and Mann-Whitney U statistic and p-value are given next to the species name. Failure rate (proportion of crosses resulting in zero progeny) is shown by a star for each cross category (right vertical axis).

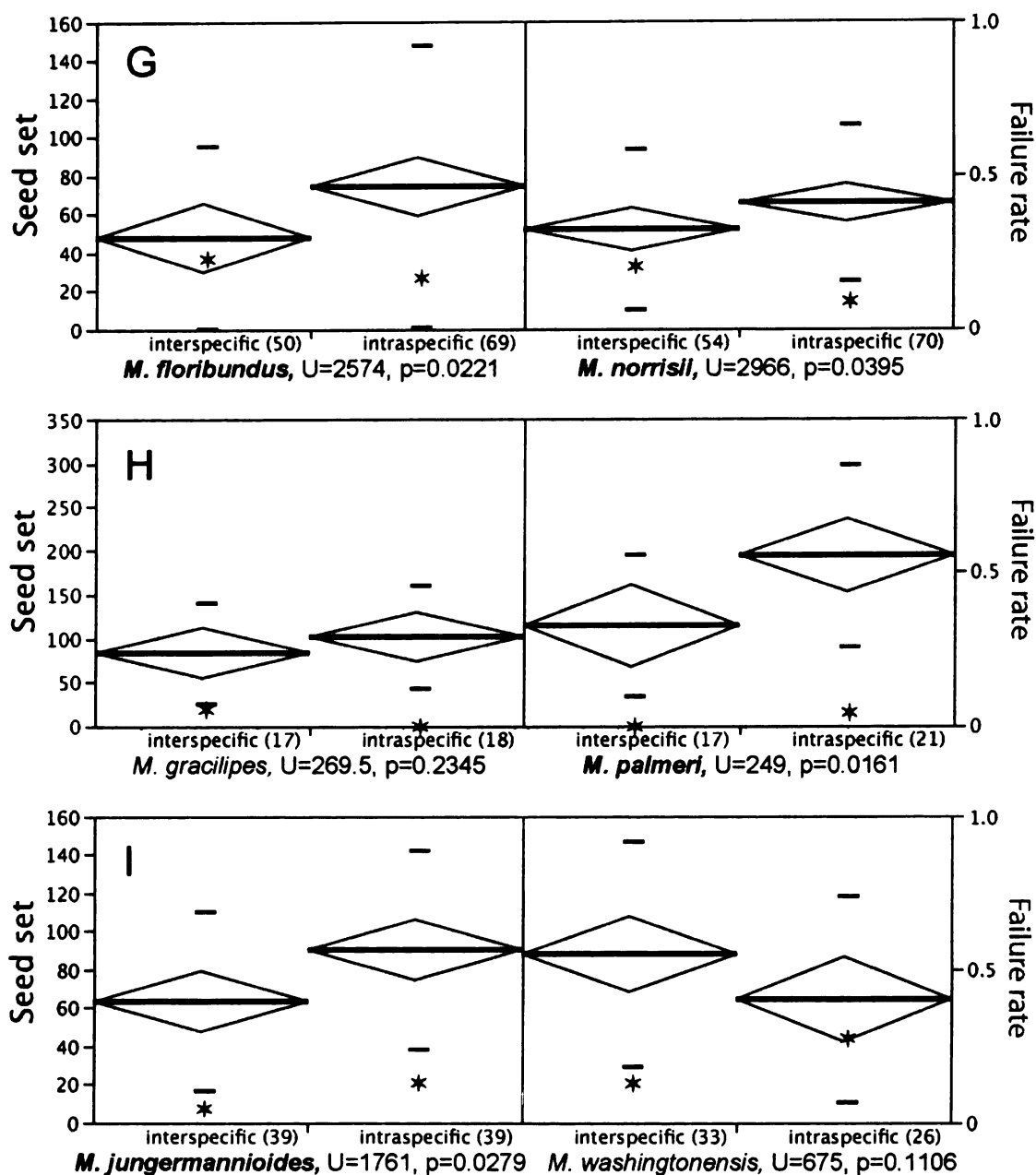


Figure 3.4G-I. Postmating reproductive isolation in the species pairs *Mimulus floribundus* / *norrisii*, *M. gracilipes* / *palmeri*, and *M. jungermannioides* / *washingtonensis*. Wide horizontal lines represent the average seed set from crosses resulting in progeny for both inter- and intraspecific crosses (left vertical axis). Sample sizes are in parentheses. Narrow horizontal lines indicate standard deviation, and the vertices of diamonds represent 95% confidence intervals. Species names in bold indicate instances of significant differences between inter- and intraspecific seed set, and Mann-Whitney U statistic and p-value are given next to the species name. Failure rate (proportion of crosses resulting in zero progeny) is shown by a star for each cross category (right vertical axis).

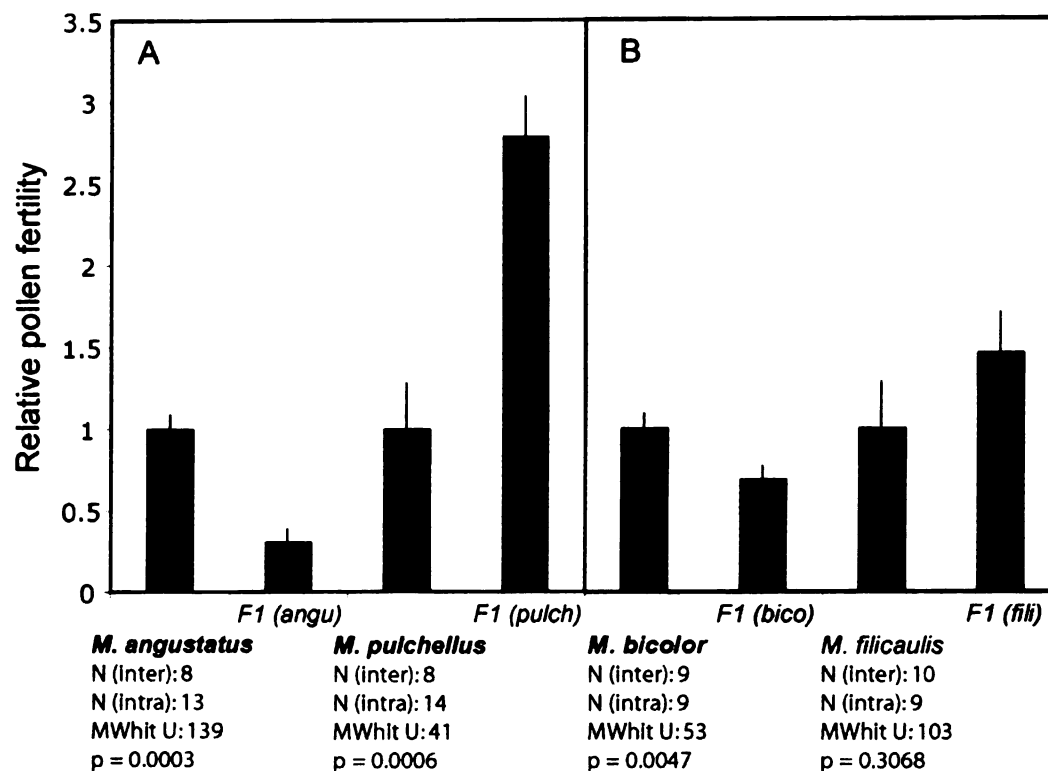


Figure 3.5A-B. Relative pollen fertility measures in species pairs *Mimulus angustatus* / *pulchellus* and *M. bicolor* / *filicaulis*. Each box contains data for a given species pair. Black bars indicate parent values while gray bars show hybrids. Estimates of pollen fertility were relativized to the value of each parent; therefore, the amount a hybrid is under or over 1 gives an estimate of the standardized change in pollen fertility. Species names in bold represent comparisons in which the intraspecific offspring differed significantly from interspecific offspring.

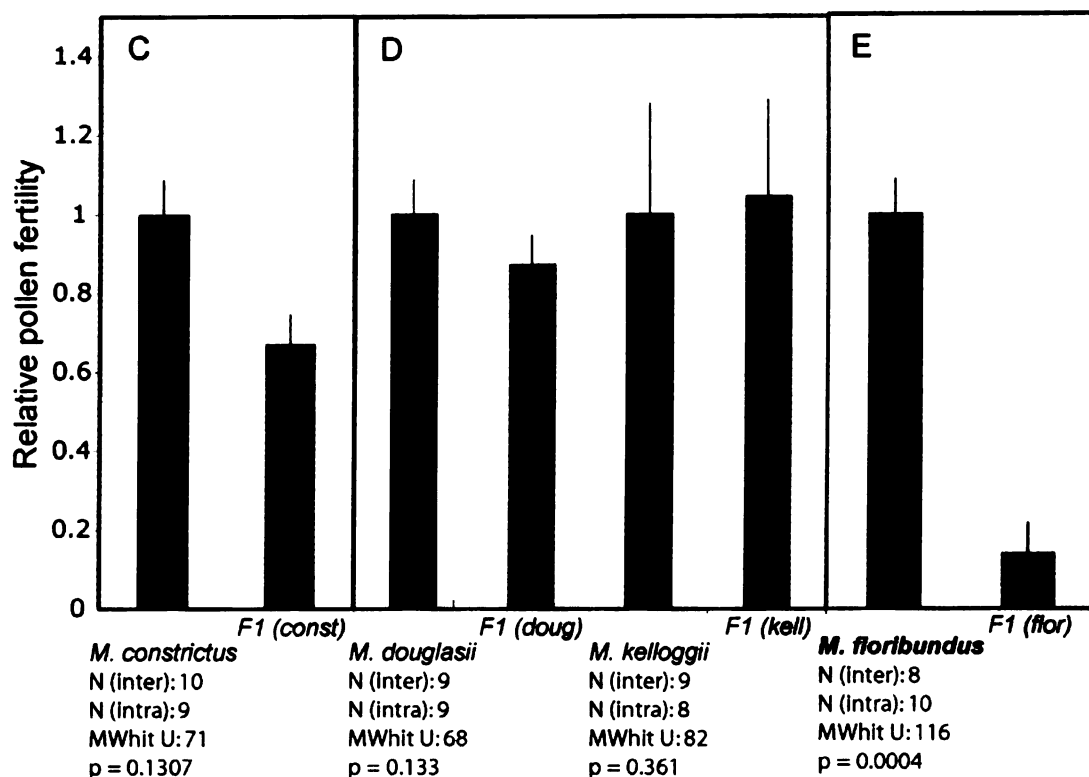


Figure 3.5C-E. Relative pollen fertility measures in species pairs *Mimulus douglasii* / *kelloggii* and individual species *M. constrictus* and *M. floribundus*. Each box contains data for a given species pair. Black bars indicate parent values while gray bars show hybrids. Estimates of pollen fertility were relativized to the value of each parent; therefore, the amount a hybrid is under or over 1 gives an estimate of the standardized change in pollen fertility. Species names in bold represent comparisons in which the intraspecific offspring differed significantly from interspecific offspring.

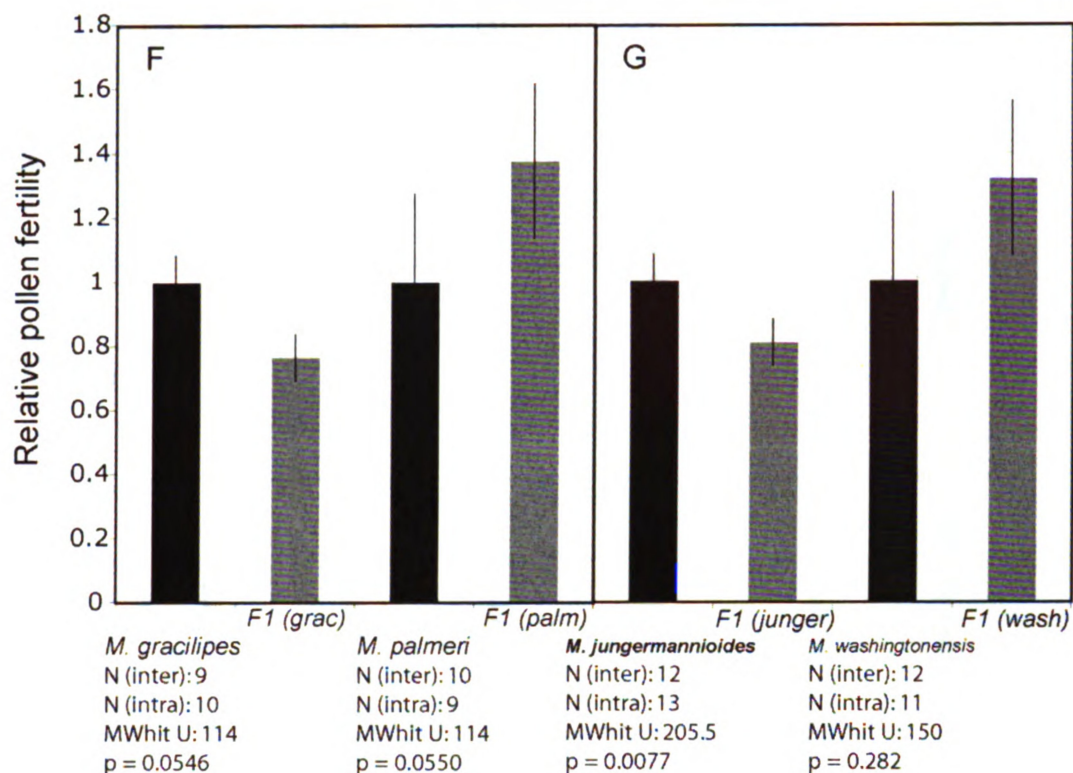


Figure 3.5F-G. Relative pollen fertility measures in species pairs *Mimulus gracilipes* / *palmeri* and *M. jungermannioides* / *washingtonensis*. Each box contains data for a given species pair. Black bars indicate parent values while gray bars show hybrids. Estimates of pollen fertility were relativized to the value of each parent; therefore, the amount a hybrid is under or over 1 gives an estimate of the standardized change in pollen fertility. Species names in bold represent comparisons in which the intraspecific offspring differed significantly from interspecific offspring.

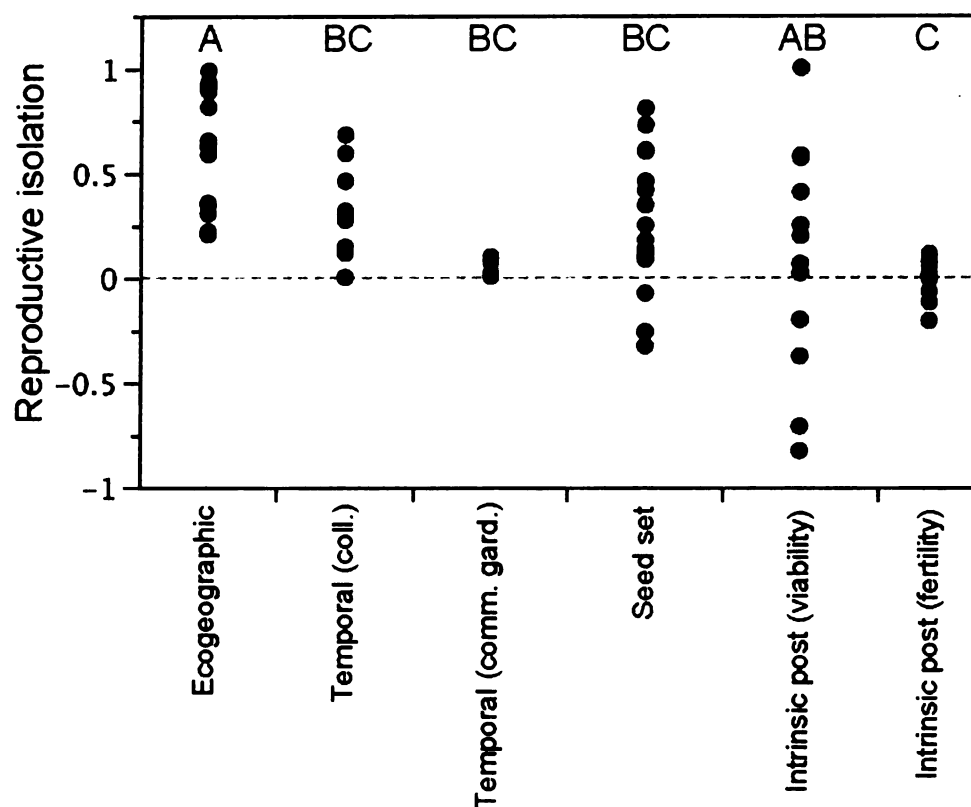


Figure 3.6. Individual estimates of reproductive isolation across the genus *Mimulus*. Each dot represents the barrier strength for a species at the given component of isolation. The barriers pictures are ecogeographic isolation, temporal isolation based on collection records (both intrinsic and extrinsic), temporal based on common garden (intrinsic only), seed set (both gametic and intrinsic postzygotic isolation), intrinsic postzygotic isolation based on viability, and intrinsic postzygotic based on fertility. Barriers differed significantly by one-way ANOVA ($F=7.686$, $df=5$, 92 , $p<0.0001$). Posthoc analysis of pairwise barrier comparisons using Tukey-Kramer technique provides separation of means indicated with A's, B's, and C's.

Figure 3.7A. Individual reproductive barrier strengths in species pairs of *Mimulus*. Reproductive isolation data for each species was averaged to create a composite estimate for each pair. Relative seed set is treated as the first postmating form of isolation due to the inability to separate gametic from early intrinsic postzygotic isolation within this type of data. Intrinsic postzygotic isolation is separated into a viability component (relative germination rates, survival rates, and numbers of flowers produced) and a fertility component (pollen fertility assays).

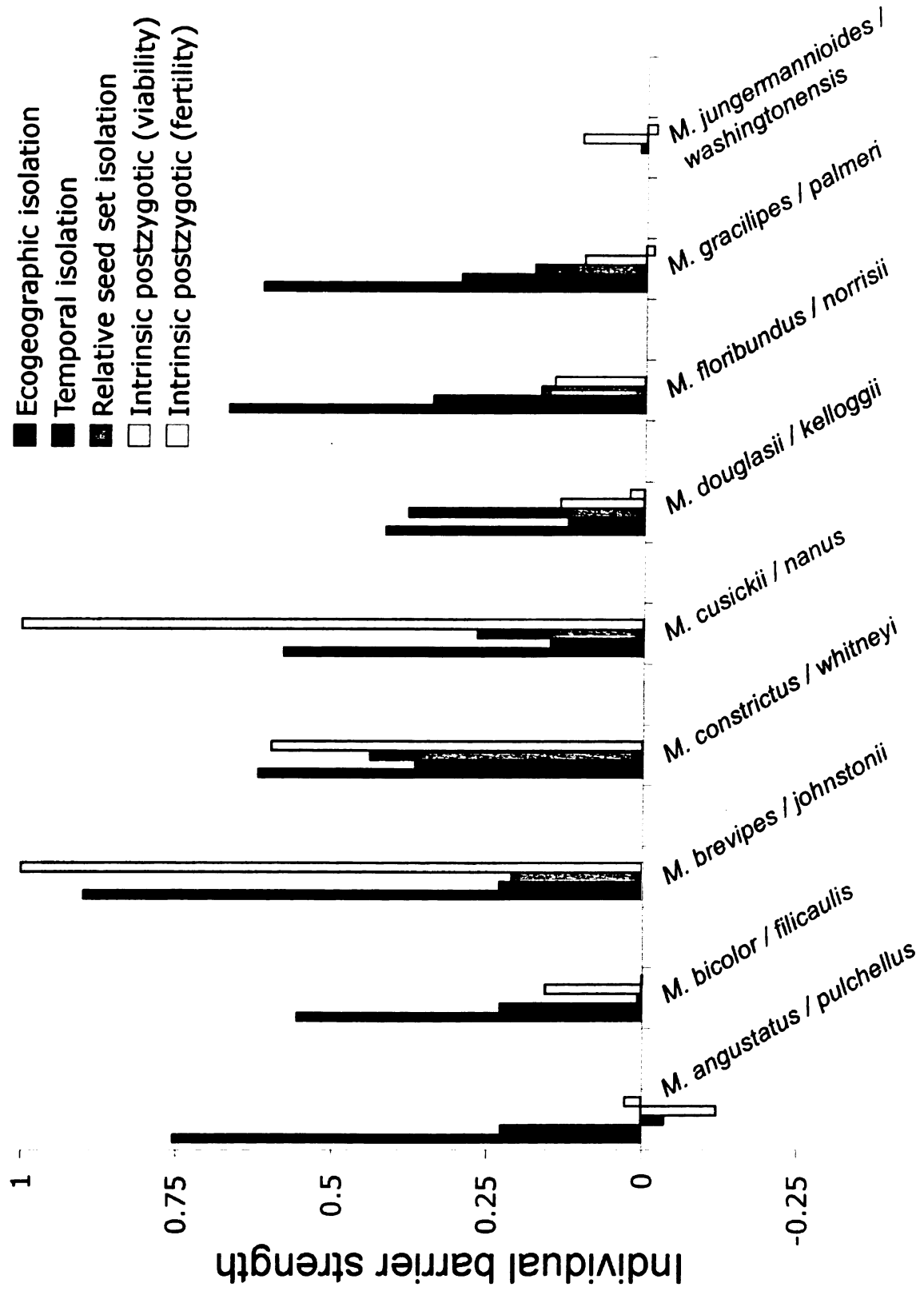
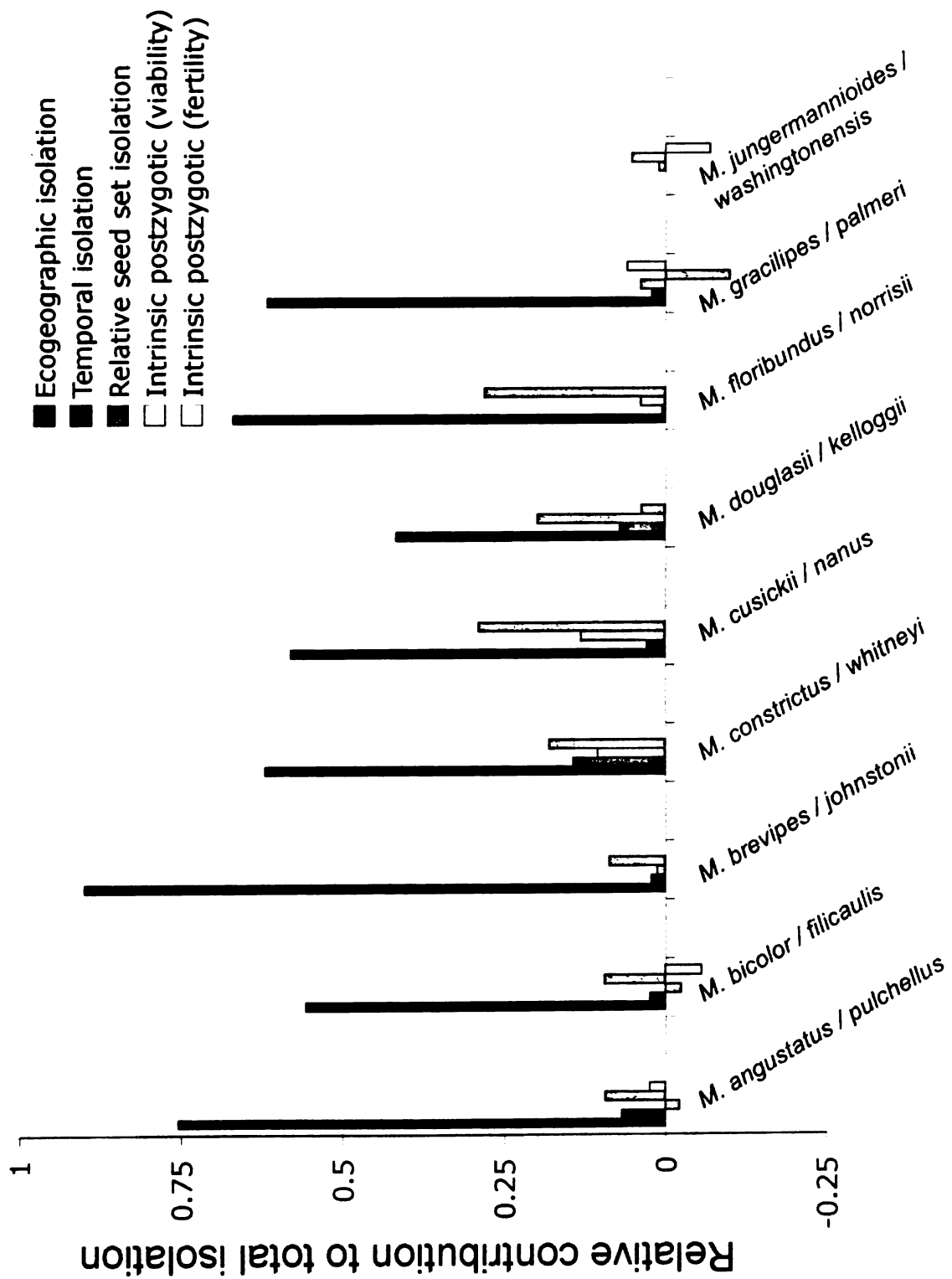


Figure 3.7B. Relative contribution of reproductive barriers to total reproductive isolation in species pairs of *Mimulus*. Reproductive isolation data for each species was averaged to create a composite estimate for each pair. Relative seed set is treated as the first postmating form of isolation due to the inability to separate gametic from early intrinsic postzygotic isolation within this type of data. Intrinsic postzygotic isolation is separated into a viability component (relative germination rates, survival rates, and numbers of flowers produced) and a fertility component (pollen fertility assays). Relative barrier contributions were calculated using the method of Coyne & Orr (1989) and Ramsey et al. (2003) in which sequentially acting barriers are discounted by the amount of isolation acting in previous barriers.



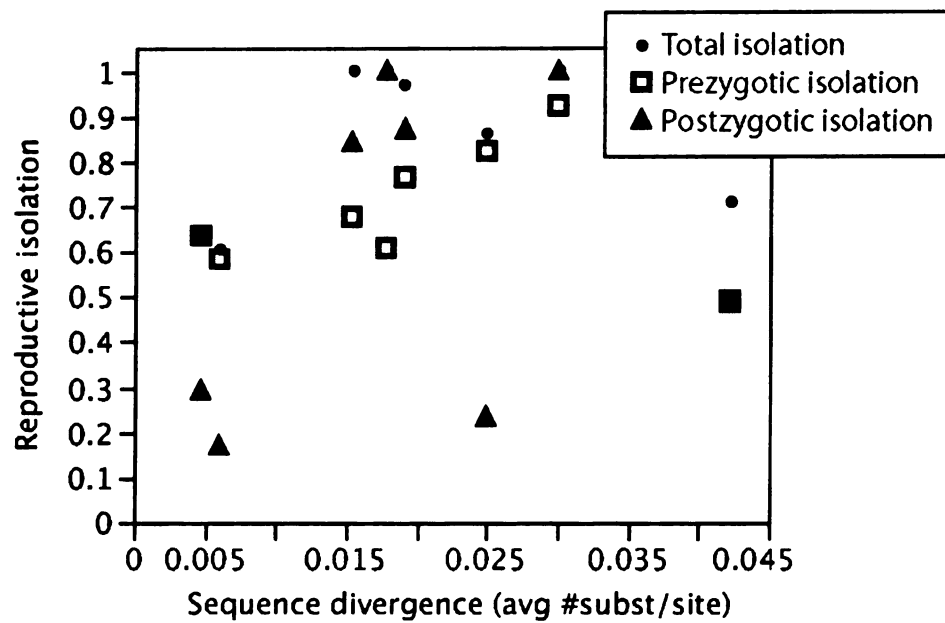


Figure 3.8. Relationship between total reproductive isolation of species pairs in *Mimulus* and genetic distance. Linear and logistic regression show no significant relationship between the variables. However, these species pairs represent recent speciation events, so very little variation in genetic distance is available.

Literature Cited

- Angert, A. L., and D. W. Schemske. 2005. The evolution of species' distributions: Reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution* 59:1671-1684.
- Barracough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155:419-434.
- Beardsley, P. M., and R. G. Olmstead. 2002. Redefining Phrymaceae: The placement of *Mimulus*, tribe Mimuleae and *Phryma*. *American Journal of Botany* 89:1093-1102.
- Beardsley, P. M., S. E. Schoenig, J. B. Whittall, and R. G. Olmstead. 2004. Patterns of evolution in Western North American *Mimulus* (Phrymaceae). *American Journal of Botany* 91:474-489.
- Bradshaw, H. D., and D. W. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426:176-178.
- Bradshaw, H. D., S. M. Wilbert, K. G. Otto, and D. W. Schemske. 1995. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* 376:762-765.
- Chari, J., and P. Wilson. 2001. Factors limiting hybridization between *Penstemon spectabilis* and *Penstemon centranthifolius*. *Canadian Journal of Botany* 79:1439-1448.
- Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43:362-381.
- Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. *Evolution* 51:295-303.
- Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.
- Cruzan, M. B., and M. L. Arnold. 1994. Assortative mating and natural selection in an iris hybrid zone. *Evolution* 48:1946-1958.
- Darwin, C. R. 1859. On the Origin of Species. John Murray, London.
- Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.
- Endler, J. A. 1982. Problems in distinguishing historical from ecological factors in biogeography. *American Zoologist* 22:441-452.

- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.
- Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Fishman, L., and J. H. Willis. 2006. A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* 60:1372-1381.
- . 2001. Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution* 55:1932-1942.
- Fishman, L., A. J. Kelly, and J. H. Willis. 2002. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* 56:2138-2155.
- Fitzpatrick, B. M., and M. Turelli. 2006. The geography of mammalian speciation: Mixed signals from phylogenies and range maps. *Evolution* 60:601-615.
- Graham, C. H., S. Ferrier, F. Huettmann, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19:497-503.
- Grant, A. L. 1924. A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden* 11:99-388.
- Harvey, P. H., and M. D. Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, New York.
- Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution* 53:866-873.
- Hiesey, W. M., M. A. Nobs, and O. Bjorkman. 1971. Experimental Studies on the Nature of Species. V. Biosystematics, genetics, and physiological ecology of the *Erythranthe* section of *Mimulus*. Carnegie Institution of Washington Publication 628, Washington, D.C.
- Howard, D. J. 1999. Conspecific sperm and pollen precedence and speciation. *Annual Review of Ecology and Systematics* 30:109-132.
- Howard, D. J., R. W. Preszler, J. Williams, S. Fenchel, and W. J. Boecklen. 1997. How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* 51:747-755.

- Husband, B. C., and H. A. Sabara. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium*. *New Phytologist* 161:713-713.
- Jepson Online Interchange for California Floristics 2004. University and Jepson Herbaria, University of California. Berkeley, CA.
- Jordan, D. S. 1908. The law of geminate species. *American Naturalist* 42:73-80.
- Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated Neotropical gingers. *Evolution* 60:538-552.
- Kearns, C. A., and D. W. Inouye. 1993. *Techniques for Pollination Biologists*. University Press of Colorado, Niwot, CO.
- Kiang, Y. T. 1972. Pollination study in a natural population of *Mimulus guttatus*. *Evolution* 26:308-310.
- Lowry, D. B., R. C. Rockwood, and J. H. Willis. 2008. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* 62:2196-2214.
- Lynch, J. D. 1978. Distribution of leopard frogs (*Rana blairi* and *Rana pipiens*) (Amphibia, Anura, Ranidae) in Nebraska. *Journal of Herpetology* 12:157-162.
- Macnair, M. R. 1983. The genetic control of copper tolerance in the yellow monkey flower, *Mimulus guttatus*. *Heredity* 50:283-293.
- Martin, N. H., and J. H. Willis. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61:68-82.
- Matsubayashi, K. W., and H. Katakura. 2009. Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles. *Evolution* 63:2563-2580.
- Mayr, E. 1942. *Systematics and the origin of species*. Columbia University Press, New York.
- Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma*). *Evolution* 57:317-327.
- Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58:1195-1208.

- Pascarella, J. B. 2007. Mechanisms of prezygotic reproductive isolation between two sympatric species, *Gelsemium rankinii* and *G. sempervirens* (Gelsemiaceae), in the southeastern United States. *American Journal of Botany* 94:468-476.
- Phadnis, N., and H. A. Orr. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323:376-379.
- Poulton, E. G. 1908. What is a species? Pp. 46-94. *Essays on Evolution 1889-1907*. Oxford at the Clarendon Press, London.
- Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57:1520-1534.
- Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: Mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43:432-449.
- Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we learned in 40 years? *Evolution* 47:1637-1653.
- Schemske, D. W. 2000. Understanding the origin of species. *Evolution* 54:1069-1073.
- Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences of the United States of America* 96:11910-11915.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295-315.
- Sobel, J. M., and A. M. Randle. 2009. Comparative approaches to the evolution of reproductive isolation: A comment on Scopece et al. 2007. *Evolution* 63:2201-2204.
- Stebbins, G. L. 1950, *Variation and Evolution in Plants*. New York, Columbia University Press.
- Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59:2548-2559.
- Sweigart, A. L., A. R. Mason, and J. H. Willis. 2007. Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution* 61:141-151.
- Tang, S. W., and D. C. Presgraves. 2009. Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323:779-782.

- Vickery, R. K. 1959. Barriers to gene exchange within *Mimulus guttatus* (Scrophulariaceae). *Evolution* 13:300-310.
- Vickery, R. K. 1964. Barriers to gene exchange between members of *Mimulus guttatus* complex (Scrophulariaceae). *Evolution* 18:52-69.
- Vickery, R. K. 1995. Speciation by aneuploidy and polyploidy in *Mimulus* (Scrophulariaceae). *Great Basin Naturalist* 55:174-176.
- Willis, J. H. 1993. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* 47:864-876.
- Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220-230.
- Yamamoto, S., and T. Sota. 2009. Incipient allochronic speciation by climatic disruption of the reproductive period. *Proceedings of the Royal Society B-Biological Sciences* 276:2711-2719.

CHAPTER 4: CONTRASTING PATTERNS OF INTROGRESSION IN TWO PAIRS
OF *MIMULUS* SPECIES

CHAPTER 4: CONTRASTING PATTERNS OF INTROGRESSION IN TWO PAIRS OF *MIMULUS* SPECIES

Abstract

While estimates of reproductive isolation obtained through laboratory and field work provide valuable information regarding the origin of species, few studies have corroborated that the strength of isolation obtained corresponds to the amount of gene flow experienced by natural populations of closely related species. Two species pairs were selected representing a species pair with strong total reproductive isolation, *Mimulus constrictus* and *M. whitneyi* and a species pair with incomplete reproductive isolation *M. bicolor* and *M. filicaulis*. Sequencing of neutral genetic markers was performed on multiple populations of each species, from populations that were both nearly sympatric and clearly allopatric with regards to the alternate species. Basic polymorphism analysis revealed that the completely isolated species pair (*M. constrictus*/*M. whitneyi*) has at least a single fixed difference between the two species and multiple cases of exclusive polymorphisms. Alternatively, the incompletely isolated pair (*M. bicolor*/*M. filicaulis*) shows a general pattern of many shared polymorphisms, fewer exclusive polymorphisms, and no fixed differences between species. Pairwise F_{ST} values confirm that the loci of *M. constrictus*/*M. whitneyi* are highly differentiated, while the F_{ST} values between *M. bicolor*/*M. filicaulis* are much lower, comparable with differentiation between populations rather than species. Taken together, these data suggest that the low value of reproductive isolation for *M. bicolor*/*M. filicaulis* obtained previously is probably accurate, and that rampant gene exchange is currently occurring (or has recently occurred. The high values of isolation obtained for *M. constrictus*/*M.*

whitneyi are also substantiated by the genetic analysis, suggesting that the high values of isolation are indeed resulting in a nearly complete restriction of gene flow.

Introduction

A long-standing debate in speciation research is how often the process occurs in the face of gene flow (Bolnick 2004; Bolnick and Fitzpatrick 2007; Coyne and Orr 2004; Feder et al. 2005; Mallet 2008; Nosil 2008; Via 2001). Sympatric and parapatric models of speciation assume that some amount of gene flow occurs throughout divergence, and determining how often speciation occurs while experiencing this constraint is a subject receiving considerable attention (Bolnick and Fitzpatrick 2007). While under most circumstances speciation should be slowed or halted by introgression, some believe that introgression may even act to facilitate speciation through the formation of novel trait combinations capable of assisting a population in traversing an adaptive valley (Arnold and Hodges 1995; Hodges et al. 1996; Rieseberg et al. 2003; Rieseberg et al. 1999).

These factors have resulted in a focus on studying speciation in populations that are incompletely isolated so that the effects of gene flow on divergence can be approached (e.g. Egan et al. 2008; Nosil et al. 2008; Turner et al. 2005). At the extreme, some advocate the abandonment of organismal species concepts altogether in favor of a 'genic' view of speciation (Lexer and Widmer 2008). While these approaches have indeed improved our understanding of how intraspecific divergence can occur in the face of introgression, an alternative point of view is that not all diverging populations will become species (Magurran 1998). Under this framework, it is important to appreciate that it may only be possible to understand the nature of speciation by studying pairs of taxa that have recently become *completely* isolated. For example, if the goal is to determine which forms of isolation are most important at the time of speciation, only taxa that have

completed speciation will be fully informative, and it is important to be able to identify the point at which species have ceased to exchange genes.

Estimates of total reproductive isolation using the linear sequential ordering of barriers estimated by laboratory or field experiments of multiple isolating barriers (e.g. Ramsey et al. 2003; Sobel, Chapter 3) can offer an excellent means of identifying the endpoint of speciation. If all of the barriers likely to be important to a group are measured, this approach can be thought of as a test of species status under the Biological Species Concept (Mayr 1942). If pairs of taxa are found to have complete (or nearly so) isolation by this method, they are diagnosed as species, and if they fall short of complete isolation, they are not yet full biological species (and there is no guarantee they will ever become so). Unfortunately, high estimates of isolation measured in the lab can be misleading. For example, Llopart et al. (2005) found an example of a hybrid zone in two species of *Drosophila* in which a class of hybrid never produced under laboratory conditions was the primary hybrid present in nature. Similarly, Yatabe et al. (2007) found evidence for introgression in two species of *Helianthus* despite strong reproductive barriers. These examples indicate that in addition to estimates of reproductive isolation, some genetic corroboration of reproductive isolation is necessary to fully assess whether speciation is complete.

Until recently, diagnosing speciation completion using molecular tools was difficult. Shared polymorphisms between populations were often detected in nature, but it is difficult to separate shared ancestral polymorphisms from those due to introgression. However, recent advances using coalescent simulations provides an opportunity to identify patterns of polymorphisms that either depart from or are consistent with neutral

expectations (Wakeley and Hey 1997; Wang et al. 1997). While these methods do not necessarily detect which loci have experienced gene flow, they can help determine if the overall pattern of polymorphism suggests that gene flow occurs at all. These methods are being increasingly utilized to corroborate measures of reproductive isolation and to estimate other population parameters important at the time of speciation (e.g. Counterman and Noor 2006; Kliman et al. 2000; Strasburg and Rieseberg 2008).

Mimulus bicolor/filicaulis and *M. constrictus/whitneyi* are two excellent species pairs in which measured strength of reproductive isolation can be compared with patterns of introgression through coalescent approaches. *M. bicolor* is a relatively widespread species throughout the Sierra Nevada, while *M. filicaulis* is restricted to a very small region outside of Yosemite National Park. While both have similar growth form, the flowers of the two species are distinct. The lower half of the corolla in *M. bicolor* is yellow and white on the upper (though some populations are dominated by an all yellow morph). In contrast, *M. filicaulis* is strikingly marked with purple markings and yellow nectar guides in the throat of the corolla. Despite the morphological differentiation, reproductive isolation seems to be relatively low between this species compared to other closely related species of *Mimulus* (Sobel, Chapter 3). There are no reproductive barriers that completely isolate the two species, and it appears that historical factors may be the key limit to species maintenance. Ecogeographic isolation is only moderately strong (average 0.56; Chapter 3), with indications that the habitat of *M. filicaulis* would be hospitable to *M. bicolor*. Other forms of isolation are almost completely missing, and a total composite value for isolation between the two species is found to be 0.60 (Sobel, Chapter 3).

In contrast, *Mimulus constrictus* and *M. whitneyi* are much closer to complete isolation. The two species live in very different habitats, with *M. constrictus* residing at mid elevations of the Coastal and Sierra ranges, and *M. whitneyi* typically growing at higher altitudes. Therefore, the two species experience limits to gene flow based on moderate to strong ecogeographic isolation (0.62), moderate habitat-mediated temporal isolation (0.37), and moderately strong isolation due to relative interspecific seed set (0.66) and hybrid viability (0.60) (Sobel, Chapter 3). A total composite index of approximately 0.97 makes this species pair much more completely isolated than *M. bicolor/filicaulis*.

These two species pairs present an opportunity to corroborate or falsify the estimates of isolation obtained in the lab and greenhouse (Sobel, Chapter 2 and 3). My hypothesis is that, while distinct in nature, the lower values of isolation suggest that *M. bicolor/filicaulis* are much more likely to experience gene flow with each other than the alternate species pair. Given their high value of reproductive isolation *M. constrictus/whitneyi* should exhibit much lower levels of introgression. I here provide a multilocus test of introgression for both of these species pairs to assess whether estimates of isolation provided in Chapter 3 are supported by genetic evidence.

Materials & Methods

Collections & population sampling

Two species pairs were selected for this study, *Mimulus bicolor/filicaulis* from section *Paradanthus*, and *M. constrictus/whitneyi* from section *Eunanus*. Each species in a pair represents each others' closest extant relative, with limited time since speciation

(Sobel, Chapter 2). Collection of plant material for DNA isolation was conducted in 2005 and 2006. For the first species pair, seven populations of *Mimulus bicolor* was collected from across a majority of its natural distribution, spanning latitudes of approximately 37.0°N to 40.2°N (populations B1-B7; Figure 4.1A). *M. filicaulis* occurs over a much more limited extent than *M. bicolor*, occurring in only a few populations near the western border of Yosemite National Park. Attempts were made to collect as many isolated populations of this species as possible, but given the close proximity among sites, they may all behave like a single population. Five populations of *M. filicaulis* were collected (Figure 4.1A). For the purposes of this analysis, the two *M. bicolor* populations flanking the range of *M. filicaulis* were considered the most likely to exchange genes and therefore were classified as 'sympatric' (B4 & B5; Figure 4.1A), while the remaining populations were classified as 'allopatric' (B1-3; B6-7; Figure 4.1A). This classification may be somewhat misleading, as these two species never co-occur at a single site (Sobel, pers. obs.), and may instead be thought of as 'near' and 'far' populations.

In the second species pair, *Mimulus constrictus* was collected from three sites across its range, nearly spanning the northern and southern limits of the species (approx. 34.8°N to 36.5°N). Similarly, *M. whitneyi* was collected from three populations spanning its rather limited geographic extent in the southern Sierra (approx. 36°N to 37°N) (Figure 1.4B). One population of *M. constrictus* (C3; Figure 4.1B) occurs in the transition between the Coastal and Transverse ranges in southwestern California in an area where *M. whitneyi* is not found. This population was classified as 'allopatric,' while the other two populations of *M. constrictus* occurring in mountain ranges with *M. whitneyi* were classified as 'sympatric' for purposes of analysis. As above, these classifications are more

related to geographic proximity than actual sympatry or allopatry because populations of these two species do not co-occur in nature either (J. Sobel, pers. obs.).

In efforts to maximize the effectiveness of parameter estimation in coalescent simulations, a minimum of six individuals per population were collected from each population (Felsenstein 2006). Leaf material from each plant was clipped and immediately placed in silica gel to desiccate the sample. Within three weeks of collection, DNA was isolated from the dried leaf material using MP Biomedicals FastDNA kits and FastPrep tissue homogenizer. DNA was suspended in water and stored at -20C.

Selection of loci for sequencing

Molecular markers were obtained through a collaboration with J. Willis and colleagues at Duke University. They have produced an EST (expressed sequence tag) library for one species of *Mimulus*, *M. guttatus*. EST sequences have been blasted against the *Arabidopsis* genome in search of homologues of the *Mimulus guttatus* ESTs, and EPIC (exon primed, intron crossing) primers are developed that amplify introns within genes that the two species have in common. Primers that successfully amplify a product in *M. guttatus* have been made available on the public website, www.mimulusevolution.org (Wu et al. 2008). Because these primers amplify putative introns, the resulting sequences are assumed to be non-coding, and are treated as neutral nuclear genetic markers.

Fifty primers were selected from among those that have successfully amplified products in *M. guttatus*, and have also been shown to amplify in *Mimulus cardinalis* and

M. lewisii (T. Bradshaw, pers. comm.). Marker development in *M. bicolor/filicaulis* and *M. constrictus/whitneyi* was carried out to identify PCR conditions that successfully amplify these four included species. Eleven markers were selected for *Mimulus bicolor/filicaulis* (Table 4.1A), and seven were selected for *M. constrictus/whitneyi* (Table 4.1B). Markers were selected that both amplified easily and spanned multiple linkage groups in a map of *M. cardinalis/lewisii* (Table 4.2) (T. Bradshaw, unpublished). While linkage maps are not available for either *M. bicolor/filicaulis* or *M. constrictus/whitneyi*, selecting markers that occur on multiple linkage groups in *M. cardinalis/lewisii* helps assure that the genomes of these species pairs will be effectively represented.

PCR and sequencing

DNA from field-collected tissue was amplified using standard PCR conditions with slight modifications to anneal temperatures as determined in optimization. Primer sequences used are provided in Table 4.2, and more details about each marker can be obtained through www.mimulusevolution.org. Amplified PCR products were run on ~2% agarose gels stained with SYBRSafe® DNA gel stain (www.invitrogen.com). PCR products were visualized by exciting gels with ~300nm blue light and observing through an orange emission filter. Agarose gels were poured with an extra row of wells, such that DNA was loaded into one set of wells and run toward the other set. The desired PCR product was monitored visually until it entered the second set of wells, upon which time the sample was recovered. Alternate lanes were used in loading to assure no cross contamination occurred between samples. DNA from recovered products was

precipitated by treatment with 3M sodium acetate pH 5.2, 100% ethanol, and incubation on ice. DNA was pelleted by centrifugation and resuspended in water after drying. 9uL of DNA sample was combined with 30pmol (in 3ul) of the appropriate forward primer, and sequencing was performed on an ABI 3730 Genetic Analyzer by the Michigan State University genomics center (www.rtsf.msu.edu).

Sequence analysis

Raw chromatograms were checked for quality manually using geneious 5.0 bioinformatics software (Drummond et al. 2010). Ambiguous base calls were corrected by hand, and alignments were performed in Clustal W version 1.82 (Thompson et al. 1994) using default parameters. Processed alignments were extracted into PHYLIP format (Felsenstein 2005) for further analysis.

Basic polymorphism analyses were conducted with the software SITES (Hey and Wakeley 1997; distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#SITES>). SITES provides data for each marker on the total number of polymorphisms, and further divides polymorphisms into exclusive, shared, and fixed differences between subgroups. Fixed polymorphisms are those that are polymorphic with respect to the entire group being analyzed, but are not polymorphic within groups and are fixed for different base pairs. Shared polymorphisms are those that are segregating within and between groups. Exclusive polymorphisms are those that are polymorphic in one subgroup, but fixed in another subgroup. For these analyses, each species pair contained one species that had both 'allopatric' and 'sympatric' sites with respect to the other species in the pair. Polymorphism analysis was conducted separately to compare allopatric and sympatric

populations to the alternate pair, such that the following comparisons were made (species A is the more widespread of the pair): 1) allopatric populations of species A vs. species B, 2) sympatric populations of species A vs. species B, and 3) all combined populations of species A vs. species B. Distributions of polymorphisms were used to estimate per locus mutation rates and recombination parameters, and pairwise F_{ST} values were calculated for allopatric, sympatric, and total populations.

Tests of neutrality were conducted using HKA software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#HKA>), which uses coalescent simulations to implement the Hudson, Kreitman, and Aguidae (HKA) statistical test for selection on loci (Hudson et al. 1987). Actual values of Tajima's D (1989) obtained from the polymorphism analysis were compared to simulated distributions in HKA from 10,000 coalescent simulations to test for significant departure from neutrality. Significant positive values of Tajima's D are consistent with balancing selection, estimates that do not differ from zero are assumed to be neutral, and significant negative values are consistent with excess rare variants. Because the markers employed in this study are introns, significant departures from expectations under neutrality are likely to arise by introgression, and can be taken as evidence for gene flow between species.

Values estimated from observed polymorphisms in SITES analysis were used to test for significant departure from an isolation model of speciation implemented in WH software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#WH>) (Kliman et al. 2000; Wakeley and Hey 1997; Wang et al. 1997). Mutation and recombination rates generated in polymorphism analysis were used to parameterize 1000 coalescent simulations, and both χ^2 and WH test statistics were examined to test for significant

departures from randomness among shared polymorphisms. The simple speciation model fit is one where zero migration occurs, and significantly higher amounts of shared polymorphisms is consistent with introgression after speciation.

Results

Within the *Mimulus bicolor/filicaulis* species pair, patterns of shared and exclusive polymorphisms do not indicate a qualitative difference between ‘allopatric’ and ‘sympatric’ comparisons of *M. bicolor* with *M. filicaulis* (Table 4.3A). If gene flow between the two species was higher in ‘sympatric’ populations, the expectation is that shared polymorphisms would be higher and exclusive polymorphisms lower in sympatric populations. However, there is no evidence of differences among the distribution of these polymorphism types (Table 4.3A). There are no fixed differences between *M. bicolor* and *M. filicaulis* in any subgrouping suggesting either very recent divergence or high levels of gene flow. F_{ST} values confirm the lack of differentiation between subgroups of *M. bicolor* and *M. filicaulis* (Table 4.4A), with no significant differences among groups (one-way ANOVA; $p=0.756$). Included in this comparison is *M. bicolor* in allopatry to other *M. bicolor* in the region of sympatry, suggesting that *M. filicaulis* is no more differentiated at the sequences analyzed than *M. bicolor* is from other populations.

Several loci show evidence of departure from neutrality as revealed by HKA analysis (Hudson et al. 1987). Actual values of Tajima’s D calculated from polymorphism analysis show significant departure from null distributions provided by coalescent simulations. In *Mimulus bicolor* 7 of the 11 loci depart from a random expectation and 3 of the 11 depart in *M. filicaulis* ($p<0.05$; Table 4.5A). The extremely

low levels of differentiation made fitting an WH isolation model difficult because the model does not perform well when expectations of fixed differences approach zero. Of 1000 coalescent simulations attempted with mutation and recombination parameters from polymorphism analysis only 46 yielded estimates of expected distributions of polymorphisms across all types (shared, exclusive, and fixed differences). The other 954 simulations ended with no estimates due to null values in the expected fixed differences category. While comparing χ^2 and WH test statistics generated from the actual data to only 46 simulated values is not an appropriate statistical test, both actual values were higher than any of the simulated values, suggesting that the isolation model can be rejected in this species pair.

Patterns within *Mimulus constrictus* and *M. whitneyi* differ markedly from *M. bicolor/filicaulis*. While polymorphism analysis reveals little in the way of fixed differences (only 1 fixed difference in total analysis; Table 4.3B), exclusive polymorphisms outnumber shared polymorphisms by more than 2.5 to 1, providing much higher levels of differentiation between subgroups within this species pair. While exclusive polymorphisms and shared polymorphisms do not differ significantly between allopatric and sympatric pairings of *M. constrictus* with *M. whitneyi* ($p=0.482$ and $p=0.996$ respectively), introgression is suggested by the fact that there are 7 fixed differences when comparing allopatric *M. constrictus* to *M. whitneyi*, while there are only 2 fixed differences between sympatric *M. constrictus* and *M. whitneyi*. F_{ST} values indicate that there is no difference in average F_{ST} between allopatric *M. constrictus* and *M. whitneyi* (mean $F_{ST}=0.512$ (0.21)) and sympatric *M. constrictus* and *M. whitneyi*

(mean $F_{ST}=0.466$ (0.056)) (Table 4.4B; $p=0.59$). However, individual loci show interesting patterns of differentiation. For example, *MgSTS 558* has an F_{ST} value of -0.008 when comparing allopatric populations of *M. constrictus* to sympatric populations of the same species, but when comparing to *M. whitneyi*, it has much higher levels of F_{ST} (0.255 in allopatric and 0.377 in sympatric comparisons). *MgSTS79* shows a similar pattern (Table 4.4.B).

Departure from neutrality as revealed by HKA tests shows that a handful of loci have more negative values of Tajima's D than expected from null simulations (Table 4.5B). These include *M. constrictus* copies of *MgSTS 46*, *MgSTS 79*, and *MgSTS 595*. While significant negative D values indicate an overabundance of rare alleles, the fact that these three loci do not exhibit relatively low F_{ST} values indicates that gene flow is most likely not responsible for the pattern. Due to the higher levels of differentiation, simulations performed under WH isolation model performed much better in this species pair. In the total analysis (both allopatric and sympatric *M. constrictus* compared to *M. whitneyi*), of the 1000 simulations, 322 provided estimates to compare actual values of χ^2 and WH test statistics. In both cases, calculated test statistics for the actual data did not differ significantly from the simulated data, indicating that an isolation can not be rejected ($\chi^2=39.718$, $p=0.506$; WH=12.00, $p=0.640$). The WH isolation model was also tested for only sympatric *M. constrictus* vs. *M. whitneyi*. In this simulation, 382 informative estimates were made from the 1000 runs, and similarly, an isolation could not be rejected ($\chi^2=42.146$, $p=0.644$; WH=10.00, $p=0.471$).

Discussion

Estimates of reproductive isolation suggest that *Mimulus bicolor* and *M. filicaulis* are incompletely isolated (Sobel Chapter 3). When two species show incomplete isolation in the lab, two alternative exist: 1) forms of isolation that were not measured are maintaining species and/or 2) gene exchange is occurring at a rate related to estimates of isolation. In the recently diverged species pair *M. bicolor* and *M. filicaulis* it appears that the second alternative is at play. Polymorphism analysis and coalescent simulations provided evidence that introgression is occurring between these two species, which corroborates the relatively low value of reproductive isolation obtained for this species pair. Indeed, along with evidence from isolation estimates, additional phylogenetic work on this species pair may reveal that there is insufficient genetic differentiation to categorize these two taxa as distinct biological species. Given the rarity of *M. filicaulis*, these results, along with evidence for a near total lack of crossing barriers (Sobel, Chapter 3), suggests that conservation strategies are necessary to insure that this taxonomic species is not lost due to hybridization with *M. bicolor*.

While the total analysis suggested rampant gene exchange in *Mimulus bicolor* and *M. filicaulis*, individual sequences showed distinct patterns. For example, locus *MgSTS 133*, showed significantly higher values of F_{ST} for between species comparisons than within *M. bicolor*, suggesting that this locus may be resisting introgression due to selection. In addition, locus *MgSTS 46* showed extremely low levels of polymorphism relative to other sequences in this study, suggesting a history of purifying selection at a nearby locus. However, polymorphism was low in both species of this pair, so selection near this locus preceded divergence.

Estimates of reproductive isolation between *Mimulus constrictus* and *M. whitneyi* suggest that this pair of species is closer to complete isolation than *M. bicolor* and *M. filicaulis*. However, the estimated strength of 0.97 leaves the possibility that the remaining isolation could either be accounted for by unmeasured barriers, or could allow some amount of introgression to occur between the species. The overall analysis provided evidence that all loci included in the study are substantially differentiated between these species. Coalescent simulations failed to find evidence for a pattern of polymorphism that departs from neutrality, but the pattern of fixed differences in allopatry and sympatry suggests that these two species may sometimes hybridize. The allopatric population of *M. constrictus* had seven fixed differences relative to *M. whitneyi*, and these differences were reduced to just a single fixed difference at *MgSTS 130* when comparing all *M. constrictus* sequences to *M. whitneyi*. This means that the fixed differences in allopatry must be exclusive polymorphisms in sympatric populations of *M. constrictus*, and in sympatry the exclusive polymorphism consists of the two base pairs that are fixed in the alternate allopatric and *M. whitneyi* samples. This suggests that sample size and power may be responsible for failing to find significant departures from neutrality for this species pair, and additional sequencing may be necessary to fully investigate introgression in this species pair. However, the overall result is consistent with the measured isolation.

In previous studies, it has been common to find discordance between measures of isolation and estimates of gene flow. For example, in a study of hydrenid beetles, Urbanelli (2002) found that isolated populations showed significantly higher gene flow than expected based on the isolation estimated by geography, and that adjacent populations showed lower levels than expected. This example highlights the importance

of using geographic isolation carefully as a measure of isolation. While complete allopatric separation may prevent the flow of genes, only geographic differences based on genetic differences are bound to maintain any degree of permanence over time (Sobel et al. 2010). Therefore, these ecologically similar but geographically distinct groups are bound to exchange more genes than ecologically differentiated populations that are closer in proximity.

More commonly, studies of gene flow find less evidence of introgression than expected based on laboratory measurements of isolation (e.g. Counterman and Noor 2006; Llopart et al. 2005). In these circumstances, laboratory measurements suggest that mating preferences are incomplete and intrinsic postzygotic isolation is missing or weak, leading to the assumption that gene flow will be quite high in nature. The finding of lower introgression rates than expected may again be related to estimates of isolation based upon habitat preferences. Very few estimates of isolation have been made for either ecogeographic isolation (but see Ramsey et al. 20003) or microhabitat isolation (but see Kay 2006). These forms of ecological isolation precede mating isolation and intrinsic postzygotic isolation in the sequence of barriers experienced by organisms in nature. Therefore, they have the potential to greatly impact gene flow before later acting barriers have a chance to do so. The close correspondence between measures of isolation and gene flow in the presented study suggests that including measures of ecogeographic isolation in estimates of total isolation will aid in harmonizing measures of isolation and introgression.

Summary

Estimates of introgression can help corroborate measures of reproductive isolation obtained in the laboratory. In this study, introgression was investigated for two species pairs of *Mimulus* that differ in their estimated degree of reproductive isolation. *Mimulus bicolor/filicaulis* had been previously estimated to have medium strength isolation, far from an amount necessary to complete speciation. Estimates of introgression based on coalescent simulations of polymorphism data corroborate this value, suggesting that these two taxonomic species share genes frequently (or have done so in the very recent past). Through two lines of evidence, this taxonomic species pair have been shown to not be biological species. *Mimulus constrictus/whitneyi* exhibit higher values of reproductive isolation in the laboratory by ecogeographic isolation, relative seed set differences, and hybrid inviability. The high level of total reproductive isolation inferred for this species pair was corroborated by gene flow analysis. No evidence of excess shared polymorphism was detected, though comparisons of allopatric and sympatric groups suggest small amount of gene flow may occur. Taken as a whole, contrasting patterns of introgression between these two species pairs suggests that measurements of reproductive isolation presented in chapter 3 are reasonably accurate. This accuracy was achieved in large part through incorporation of ecogeographic isolation into estimates of total reproductive isolation.

Table 4.1A. Molecular markers sequenced for analysis of introgression within *Mimulus bicolor* / *M. filicaulis* species pair. Populations of *M. bicolor* that flank the range of *M. filicaulis* were considered sympatric (sym), while others were considered allopatric (allo) (see Figure 4.1A).

Marker	Aligned length (bp)	Number of individuals sequenced			<i>M. filicaulis</i>
		<i>M. bicolor</i> (allo)	<i>M. bicolor</i> (sym)	<i>M. bicolor</i> (total)	
<i>MgSTS 28</i>	220	31	17	48	18
<i>MgSTS 46</i>	270	34	13	47	20
<i>MgSTS 50</i>	283	35	18	53	21
<i>MgSTS 51</i>	68	33	24	57	24
<i>MgSTS 55</i>	123	31	20	51	17
<i>MgSTS 122</i>	170	24	17	41	19
<i>MgSTS 133</i>	165	34	19	53	18
<i>MgSTS 219</i>	122	32	19	51	27
<i>MgSTS 273</i>	75	36	26	62	26
<i>MgSTS 283</i>	153	34	22	56	15
<i>MgSTS 345</i>	148	27	25	52	21

Table 4.1B. Molecular markers sequenced for analysis of introgression within *Mimulus constrictus* / *M. whitneyi* species pair. Populations of *M. constrictus* that occur within same mountain range as populations of *M. whitneyi* were classified as sympatric (sym), while others were considered allopatric (allo) (see Figure 4.1B).

Marker	Aligned length (bp)	Number of individuals sequenced			
		<i>M. constrictus</i> (allo)	<i>M. constrictus</i> (sym)	<i>M. constrictus</i> (total)	<i>M. whitneyi</i>
<i>MgSTS 46</i>	189	7	10	17	22
<i>MgSTS 79</i>	151	10	11	21	19
<i>MgSTS 104</i>	284	5	3	8	14
<i>MgSTS 130</i>	362	7	5	12	12
<i>MgSTS 464</i>	85	10	11	21	21
<i>MgSTS 558</i>	83	9	6	15	14
<i>MgSTS 595</i>	145	10	11	21	19

Table 4.2. Primer sequences used to amplify *MgSTS* markers in this study. Linkage group refers to linkage map locations of each marker in *M. cardinalis/lewisii* (T. Bradshaw, unpublished).

Marker name	Linkage Group	Forward (5'-3')	Reverse (5'-3')
<i>MgSTS</i> 28	4	CGATTGTTGATGCAATCAGG	GCACGGCATACTGAAGACAA
<i>MgSTS</i> 46	3	AGTGCTGCAGCTCAATGCTA	GCTGGATCGATTCTGAGGAC
<i>MgSTS</i> 50	1	CCTACGTGACAGAACGTCCA	TCCTGACCACCTCCAAGAAC
<i>MgSTS</i> 51	4	CCTGATGTTGACGGTTTCCT	TCAGCAGCCTTGATGATGTC
<i>MgSTS</i> 55	8	CCTCACTTCCAAAGCTCCTCA	TGTGGACACAGTTTCAGCGTA
<i>MgSTS</i> 79	3	GATCTTTTGAGCGACAGGAAG	TGTTGTCTACCATCTTCTCTCATACA
<i>MgSTS</i> 104	4	CCGAACCCCTTACACAAGGAC	GGCACAAACACAACCAAGATG
<i>MgSTS</i> 122	1	GGACACAGCCCTTGAAAGCAT	ATCACTGCGTCTGCGTATTG
<i>MgSTS</i> 130	6	ACTCGCCCTTCGATTCTAT	TCCGGGATGACTTCCTTCTT
<i>MgSTS</i> 133	6	GGACAAATCCTTGCTGCAT	CGCTTCTCCTCCTTTGTGCG
<i>MgSTS</i> 219	2	CAACACCCCATGCAAAAGAAC	CTTGCGAAGAAATACCATGC
<i>MgSTS</i> 273	3	CCCTGTTACTTGTGCCATCC	GGAAAGCTGGAGACTTGATGC
<i>MgSTS</i> 283	5	CGGGTAAATAGACCCCAATAGTCC	ATCACATCCGGGTGATAACC
<i>MgSTS</i> 345	6	GGCTCTTGATGAAGAGTGG	GATCTCGGACTTCCTCAACG
<i>MgSTS</i> 464	7	GAAGGTCTTGCCCTACACTTATCG	AGCAAAGTTGCACCCCATACC
<i>MgSTS</i> 558	2	GTCGAATATGCCAATTGAAGC	TCCACGGCAAGAACTACTCC
<i>MgSTS</i> 595	6	TTAACCAAATTCGGATGAACC	AGGATGGATCGAAGAGAACG

Table 4.3A. Results of polymorphism analysis using SITES software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#SITES>) (Hey and Wakeley 1997) for *Mimulus bicolor* and *M. filicaulis*. ‘Shared’ and ‘exclusive’ and ‘total’ refer to the number of nucleotide polymorphisms of each of these types. ‘Fixed’ refers to the number of fixed differences between each category and *M. filicaulis*. All comparisons from *M. bicolor* are made in reference to *M. filicaulis*. Therefore, the number of exclusive polymorphisms in *M. filicaulis* can be calculated by subtracting the ‘shared’ quantity from any of the three comparisons.

Marker	Allopatric				<i>M. bicolor</i>				<i>M. filicaulis</i>			
	Shared	Exclusive	Fixed	Shared	Shared	Exclusive	Fixed	Shared	Exclusive	Fixed	Total	Total
MgSTS 28	67	62	0	52	54	0	0	80	76	0	115	115
MgSTS 46	0	2	0	0	3	0	0	0	5	0	1	1
MgSTS 50	47	45	0	39	63	0	0	57	74	0	113	113
MgSTS 51	15	23	0	7	23	0	0	17	27	0	23	23
MgSTS 55	40	35	0	34	36	0	0	45	48	0	59	59
MgSTS 122	17	58	0	11	18	0	0	18	67	0	31	31
MgSTS 133	6	14	0	3	13	0	0	7	24	0	17	17
MgSTS 219	27	29	0	26	17	0	0	36	32	0	44	44
MgSTS 273	3	8	0	3	5	0	0	4	10	0	4	4
MgSTS 283	22	48	0	27	31	0	0	29	55	0	33	33
MgSTS 345	7	18	0	7	10	0	0	7	22	0	14	14
Mean (st. dev.) per bp	0.146 (0.11)	0.207 (0.12)	0	0.119 (0.095)	0.160 (0.11)	0	0	0.174 (0.13)	0.260 (0.13)	0	0.250 (0.18)	0.250 (0.18)

Table 4.3B. Results of polymorphism analysis using SITES software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#SITES>) (Hey and Wakeley 1997) for *Mimulus constrictus* and *M. whitneyi*. ‘Shared’ and ‘exclusive’ and ‘total’ refer to the number of nucleotide polymorphisms of each of these types. ‘Fixed’ refers to the number of fixed differences between each category and *M. whitneyi*. All comparisons from *M. constrictus* are made in reference to *M. whitneyi*. Therefore, the number of exclusive polymorphisms in *M. whitneyi* can be calculated by subtracting the ‘shared’ quantity from any of the three comparisons.

Marker	<i>M. constrictus</i>						<i>M. whitneyi</i>		
	<u>Allopatric</u>			<u>Sympatric</u>			<u>Total</u>		
	Shared	Exclusive	Fixed	Shared	Exclusive	Fixed	Shared	Exclusive	Fixed
MgSTS 46	0	6	1	4	9	0	4	15	0
MgSTS 79	1	2	0	1	3	0	1	5	0
MgSTS 104	9	21	0	2	6	0	11	22	0
MgSTS 130	5	15	4	8	15	2	10	26	1
MgSTS 464	0	1	2	0	5	0	0	6	0
MgSTS 558	5	8	0	4	1	0	5	9	0
MgSTS 595	0	2	0	1	1	0	1	3	0
Mean (st. dev.) per bp	0.0161 (0.022)	0.0403 (0.033)	0.0057 (0.008)	0.0160 (0.016)	0.0300 (0.020)	0.0008 (0.002)	0.0230 (0.021)	0.1044 (0.261)	0.0004 (0.001)
									0.0693 (0.082)

Table 4.4A. Pairwise F_{ST} values from molecular polymorphism analysis performed in SITES (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#SITES>) (Hey and Wakeley 1997) for *Mimulus bicolor* and *M. filicaulis*. ‘Sym’ refers to populations of *M. bicolor* that flank *M. filicaulis*, while ‘allo’ refers to all other populations (See Figure 4.1A).

Marker	<i>M. bicolor</i> (allo v. sym)	<i>M. bicolor</i> (allo) v. <i>M. filicaulis</i>	<i>M. bicolor</i> (sym) v. <i>M. filicaulis</i>	<i>M. bicolor</i> v. <i>M. filicaulis</i> (overall)
<i>MgSTS</i> 28	0.041	0.015	-0.02	-0.006
<i>MgSTS</i> 46	0.039	-0.174	0.065	-0.061
<i>MgSTS</i> 50	0.032	0.049	0.083	0.056
<i>MgSTS</i> 51	0.004	0.123	0.121	0.121
<i>MgSTS</i> 55	0.01	-0.052	-0.013	-0.039
<i>MgSTS</i> 122	0.023	0.111	0.07	0.091
<i>MgSTS</i> 133	0.126	0.313	0.372	0.315
<i>MgSTS</i> 219	0.022	0.056	0.038	0.043
<i>MgSTS</i> 273	0.166	0.375	0.178	0.245
<i>MgSTS</i> 283	0.028	0.085	0.096	0.082
<i>MgSTS</i> 345	0.07	0.181	0.037	0.103
Mean (st. dev.)	0.051 (0.05)	0.098 (0.15)	0.093 (0.11)	0.086 (0.11)

Table 4.4B. Pairwise F_{ST} values from molecular polymorphism analysis performed in SITES (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#SITES>) (Hey and Wakeley 1997) for *Mimulus constrictus* and *M. whitneyi*. ‘Sym’ refers to populations of *M. constrictus* that occur in mountain ranges also containing *M. whitneyi*, while ‘allo’ refers to all other populations (See Figure 4.1B).

Marker	<i>M. constrictus</i> (allo v. sym)	<i>M. constrictus</i> (allo) v. <i>M. whitneyi</i>	<i>M. constrictus</i> (sym) v. <i>M. whitneyi</i>	<i>M. constrictus</i> v. <i>M. whitneyi</i> (overall)
<i>MgSTS 46</i>	0.318	0.581	0.556	0.52
<i>MgSTS 79</i>	0.091	0.465	0.452	0.456
<i>MgSTS 104</i>	0.438	0.319	0.459	0.281
<i>MgSTS 130</i>	0.172	0.606	0.508	0.553
<i>MgSTS 464</i>	0.325	0.887	0.441	0.647
<i>MgSTS 558</i>	-0.008	0.255	0.377	0.305
<i>MgSTS 595</i>	0.138	0.471	0.468	0.467

Table 4.5A. Estimates of Tajima's D (1989) and the mutation rate parameter θ ($4N\mu$) per base pair in *Mimulus bicolor* and *M. filicaulis* species. Significant departure from neutrality for distributions of polymorphisms (i.e. fixed differences, shared, and exclusive polymorphisms) was conducted using coalescent simulations using HKa software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software/HKA>) (Hey and Wakeley 1997). Actual values of Tajima's D were compared to the distribution of randomly generated values from 10,000 simulation runs, implementing the Hudson, Kreitman, and Aguiade (HKA) test for significance ((Hudson et al. 1987)). Significant p-values are in bold.

Marker	<i>M. bicolor</i>			<i>M. filicaulis</i>		
	θ	Tajima's D	p-value	θ	Tajima's D	p-value
MgSTS 28	0.231	-1.600	0.031	0.203	-1.472	0.057
MgSTS 46	0.005	-2.010	0.001	0.001	-1.198	0.086
MgSTS 50	0.126	-1.804	0.012	0.137	-1.755	0.021
MgSTS 51	0.173	-1.718	0.018	0.111	-0.997	0.158
MgSTS 55	0.273	-1.176	0.100	0.259	-1.596	0.039
MgSTS 122	0.169	-2.198	0.002	0.078	-2.022	0.080
MgSTS 133	0.485	-2.111	0.002	0.034	-1.439	0.061
MgSTS 219	0.152	-1.322	0.075	0.115	-1.955	0.010
MgSTS 273	0.048	-1.421	0.064	0.016	-1.417	0.074
MgSTS 283	0.151	-0.583	0.002	0.084	-0.066	0.217
MgSTS 345	0.046	-1.169	0.119	0.028	-0.793	0.234
Mean (st dev)	0.169 (0.13)	-1.556 (0.48)		0.097 (0.08)	-1.337 (0.56)	

Table 4.5B. Estimates of Tajima's D (1989) and the mutation rate parameter θ ($4N\mu$) per base pair in *Mimulus constrictus* and *M. whitneyi* species. Significant departure from neutrality for distributions of polymorphisms (i.e. fixed differences, shared, and exclusive polymorphisms) was conducted using coalescent simulations using HKA software (distributed by J. Hey; <http://genfaculty.rutgers.edu/hey/software#HKA>) (Hey and Wakeley 1997). Actual values of Tajima's D were compared to the distribution of randomly generated values from 10,000 simulation runs, implementing the Hudson, Kreitman, and Aguiade (HKA) test for significance (Hudson et al. 1987). Significant p-values are in bold.

Marker	θ	<i>M. constrictus</i>		<i>M. whitneyi</i>	
		Tajima's D	p-value	Tajima's D	p-value
MgSTS 46	0.0313	-1.651	0.038	-1.215	0.105
MgSTS 79	0.0117	-1.975	0.013	0.332	0.224
MgSTS 104	0.0564	-0.394	0.351	-0.452	0.376
MgSTS 130	0.0443	-1.261	0.174	-1.261	0.093
MgSTS 464	0.0201	-1.446	0.078	-0.163	0.453
MgSTS 558	0.0563	-1.088	0.141	-1.338	0.090
MgSTS 595	0.0081	-1.861	0.012	0.668	0.134
Mean (st dev)	0.0326 (0.02)	-1.3822 (0.54)		-0.490 (0.81)	
			0.0470 (0.030)		



Figure 4.1A. Distribution of populations sampled for molecular analysis in *Mimulus bicolor* and *M. filicaulis*. A projected ecological niche map is used as background (Chapter 2, Figure 2.5C) with black areas corresponding to unsuitable habitat, dark gray suitable to *M. bicolor* alone, white suitable to *M. filicaulis* alone, and light gray suitable to both species. Populations of *M. bicolor* that flank the distribution of *M. filicaulis* were classified as sympatric (B4-B5), and all other populations were classified as allopatric (B1-B3; B6-B7). Five populations of *M. filicaulis* reside within a very limited geographic extent in the cluster marked F1-F5.

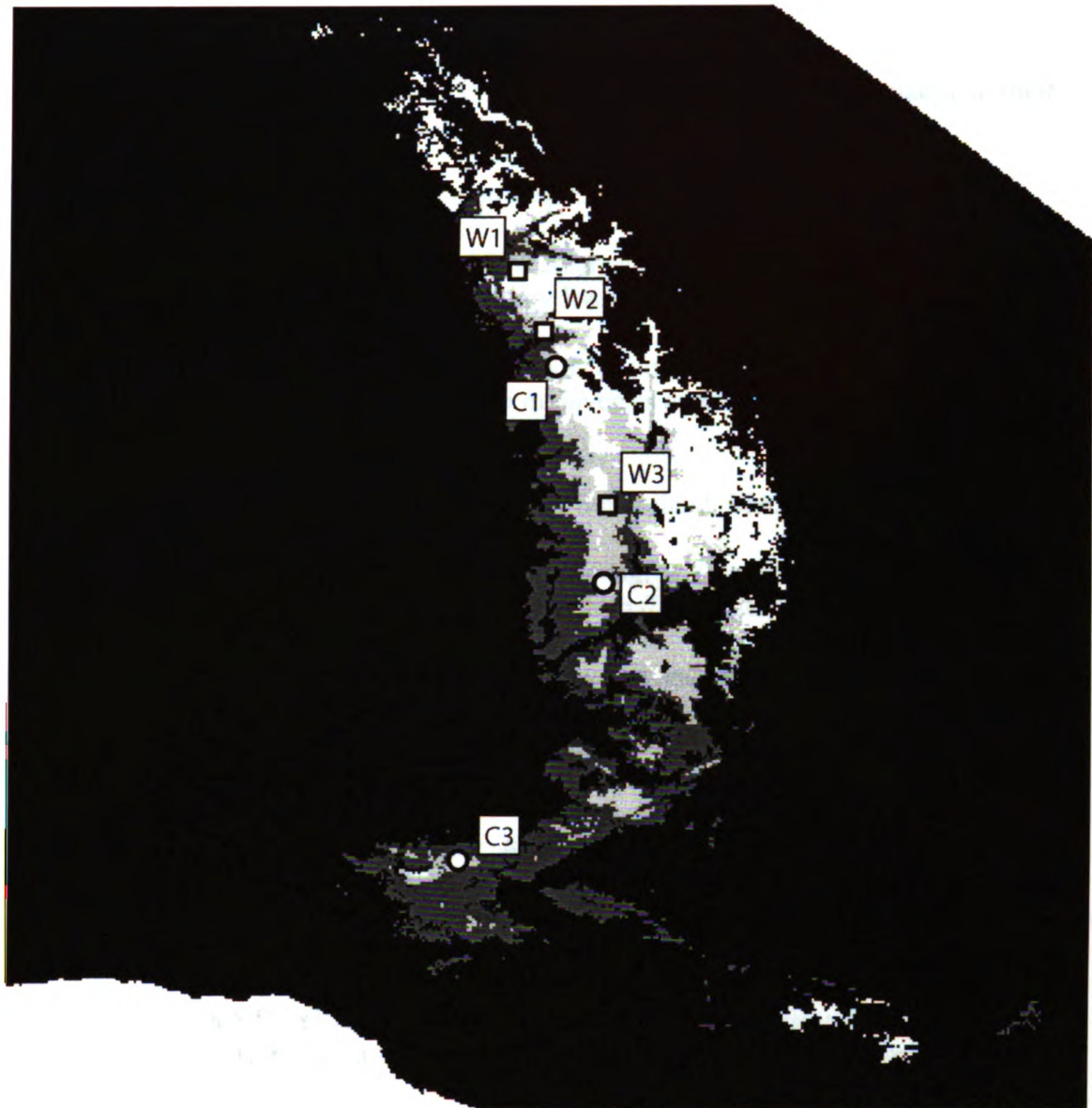


Figure 4.1B. Distribution of populations sampled for molecular analysis in *Mimulus constrictus* and *M. whitneyi*. A projected ecological niche map is used as background (Chapter 2, Figure 2.5G) with black areas corresponding to unsuitable habitat, dark gray suitable to *M. constrictus* alone, white suitable to *M. whitneyi* alone, and light gray suitable to both species. Two populations of *M. constrictus* occur in middle altitudes of the Sierra, where *M. whitneyi* occurs at high elevation. These two populations were classified as sympatric with *M. whitneyi* for purposes of analysis (C1, C2). The other *M. constrictus* populations (C3) occurs in transition area between the Coastal and Transverse ranges in an area without *M. whitneyi*. This population was considered allopatric for purposes of analysis.

Literature Cited

- Arnold, M. L., and S. A. Hodges. 1995. Are natural hybrids fit or unfit relative to their parents. *Trends in Ecology & Evolution* 10:67-71.
- Bolnick, D. I. 2004. Waiting for sympatric speciation. *Evolution* 58:895-899.
- Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology Evolution and Systematics* 38:459-487.
- Counterman, B. A., and M. A. F. Noor. 2006. Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. *American Naturalist* 168:682-696.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sunderland, MA, Sinauer Associates.
- Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse et al. 2010. Geneious v5.0, Available from <http://www.geneious.com>.
- Egan, S. P., P. Nosil, and D. J. Funk. 2008. Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* 62:1162-1181.
- Feder, J. L., X. F. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski et al. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States of America* 102:6573-6580.
- Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- . 2006. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* 23:691-700.
- Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145:833-846.
- 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.
- Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153-159.

- Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated Neotropical gingers. *Evolution* 60:538-552.
- Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter et al. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156:1913-1931.
- Lexer, C., and A. Widmer. 2008. The genic view of plant speciation: recent progress and emerging questions. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363:3023-3036.
- Llopart, A., D. Lachaise, and J. A. Coyne. 2005. An anomalous hybrid zone in *Drosophila*. *Evolution* 59:2602-2607.
- Magurran, A. E. 1998. Population differentiation without speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 353:275-286.
- Mallet, J. 2008. Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363:2971-2986.
- Mayr, E. 1942, *Systematics and the origin of species*. New York, Columbia University Press.
- Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology* 17:2103-2106.
- Nosil, P., S. R. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* 62:316-336.
- Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57:1520-1534.
- Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy et al. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211-1216.
- Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152:713-727.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295-315.

- Strasburg, J. L., and L. H. Rieseberg. 2008. Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris* - large effective population sizes and rates of long-term gene flow. *Evolution* 62:1936-1950.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal-W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673-4680.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* 3:1572-1578.
- Urbanelli, S. 2002. Genetic divergence and reproductive isolation in the *Ochthebius* (*Calobius*) complex (Coleoptera : Hydraenidae). *Heredity* 88:333-341.
- Via, S. 2001. Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology & Evolution* 16:381-390.
- Wakeley, J., and J. Hey. 1997. Estimating ancestral population parameters. *Genetics* 145:847-855.
- Wang, R. L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* 147:1091-1106.
- Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220-230.
- Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 175:1883-1893.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03220 8930