Population Genetics, Natural Selection and Genetic Architecture

of the Selfing Syndrome in the Morning Glory *Ipomoea lacunosa*

by

Joanna Lucy Rifkin

University Program in Genetics and Genomics
Duke University

Date: June 2 2017
Approved:

_____
Mark Rausher, Supervisor

_____
Tom Mitchell-Olds

_____
John Willis

_____
Kathleen Donohue

Dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in the
University Program in Genetics and Genomics
in the Graduate School of Duke University

2017

ABSTRACT

Population Genetics, Natural Selection and Genetic Architecture

of the Selfing Syndrome in the Morning Glory *Ipomoea lacunosa*

by

Joanna Lucy Rifkin

University Program in Genetics and Genomics
Duke University

Date: June 2 2017
Approved:

_____
Mark Rausher, Supervisor

_____
Tom Mitchell-Olds

_____
John Willis

_____
Kathleen Donohue

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
University Program in Genetics and Genomics in the
Graduate School of Duke University

2017

Abstract

The evolutionary transition from outcrossing to self-pollination occurs frequently in flowering plants and has direct and indirect effects on genomics, life history and floral morphology. The life history and floral traits common to selfing plants are collectively called the "selfing syndrome." This dissertation uses the highly selfing morning glory *Ipomoea lacunosa* to address three major questions in the evolution of highly selfing plants: what are the genomic consequences of selfing, do the morphological changes associated with the transition to self-pollination result from natural selection or genetic drift, and how does the genetic architecture of those morphological traits affect their evolution? In the first chapter, we analyze genetic data from *I. lacunosa* and its outcrossing sister species *I. cordatotriloba* to compare the genomic consequences of selfing in *I. lacunosa* to theoretical predictions. We find that the reduction in genetic diversity is greater than that predicted by theory, suggesting a population bottleneck in *I. lacunosa*'s history. There is also evidence for the relaxation of natural selection. The second chapter combines these genetic data with phenotypic measurements in a Qst-Fst comparison to determine whether natural selection is responsible for life history and floral morphology differences between *I. lacunosa* and its outcrossing relative *I. cordatotriloba*. Our analyses reveal that several component traits in the selfing syndrome diverged in response to natural selection. Chapter Three uses a quantitative trait locus (QTL) mapping approach to characterize the genetic architecture of the selfing syndrome and investigate how genetic correlations between traits affected its evolution.

We find generally lower levels of genetic correlation between selfing syndrome traits than previous QTL studies of the selfing syndrome. The low level of genetic correlation indicates that independent selection on selfing syndrome traits is responsible for the evolution of the syndrome as a whole.

Table of Contents

List of Tables

List of Figures

Acknowledgements

There are so many people to thank that I'm sure I forgot someone. Sorry.

I thank my colleagues among the graduate students at Duke. It is an honor to have known all of you. In particular, my days would be drab indeed without labmates Rong-Chien Lin, Irene Liao, Allan Castillo, Kimmy Stanton and HanHan Xia. I thank my UPGG cohort, especially Megan Aldrup, Annie Jeong and Mike Landowski, for going through those first few years together. Colleagues in other labs, particularly Jenn Coughlan, for her sharpness and insight. My close friends Ashley Troth and Rose Keith, for always being there with the compassion to give me chocolate and the honesty to tell me I'm being a little dramatic.

I thank the senior students and postdocs who encouraged me, answered my questions, and provided advice and support from my arrival here: Tanya Duncan, who encouraged me to come to Duke, told me what a great adviser Mark was, handed on her project to me, and remains the staunchest supporter of its continuing evolution. Cathy Rushworth, Maggie Wagner and Liana Burghardt took time from their own studies to look out for me. And our lab's postdocs Talline Martins and Kate Ostevik, who came through for me with guidance and support at the very beginning and the very end.

I thank the greenhouse staff—Jenny Gordon, John Mays, Jorge Fidhel Gonzalez and Michael Barnes—and the several Duke University undergrads who have also helped with plant maintenance and data collection. I thank the faculty and staff of the

Duke Biology department, for their warmth and support throughout my time here, especially Jim Tunney and Anne Lacey, and the faculty and staff of the University Program in Genetics and Genomics. This work has been supported by funding from the NSF, the NIH, the Duke Biology Department, Duke University, and Sigma Xi.

I thank my dedicated, patient and insightful committee members, John Willis, Tom Mitchell-Olds, and Kathleen Donohue, for offering their enthusiasm and helping me to direct mine.

I thank my adviser Mark Rausher, who has shaped this dissertation and my development as a researcher in ways I suspect I am barely beginning to appreciate. His calm, rigor and commitment to scientific excellence inspire me as a scientist. His patience, compassion and integrity inspire me as a human.

I thank my parents Helen Palmer and Joshua Rifkin, for instilling in me unceasing curiosity, a love of learning, a fascination with the natural world, a strong work ethic, and a fondness for academia. Thank you for helping me achieve—to a first approximation—my original career ambition of "Bug Doctor." And the rest of my family and my in-laws. I love all of you.

Finally, I thank Matt Dougherty, a better partner than I could have imagined in love, life, and adventure. I'm glad you're here with me, Matt. Here, at the end of all things. *Contra mundum.*

Introduction

"It is hardly an exaggeration to say that Nature tells us, in the most emphatic manner, that she abhors perpetual self-fertilisation." (Darwin 1877)

The founder of evolutionary biology had inbreeding depression in mind when he wrote these words. Had he been reviewing a phylogeny of flowering plants, he might have thought differently, for the shift from outcrossing to selfing has occurred many times in angiosperms. Between 10% and 15% of flowering plants are highly selfing, and there have been many independent transitions, often coinciding with speciation events (Barrett, Arunkumar, and Wright 2014; Goldberg and Igić 2012; Barrett 2002). This common evolutionary transition has major genomic and morphological consequences for selfing species (Sicard and Lenhard 2011; Barrett, Arunkumar, and Wright 2014).

Transitions to selfing are expected to change the population genetic structure of a species in several ways. Theory predicts that the effective population size of a selfing species is reduced proportionally to the increase in selfing rate, and empirical studies have found results of more extreme reductions (Pollak 1987; B. Charlesworth, Morgan, and Charlesworth 1993; Nordborg and Donelly 1997; Foxe et al. 2010; Guo et al. 2009; Busch, Joly, and Schoen 2011; Pettengill and Moeller 2012). This reduction in effective population size also reduces the efficacy of natural selection (Glémin 2007; Slotte et al. 2013). High rates of selfing also increase homozygosity, reducing effective population

size still further (D. Charlesworth and Wright 2001; Pollak and Sabran 1992) and increasing the scale of the decay of linkage disequilibrium.

The effects of selfing are not only genetic, however. Like evolutionary changes between biotic pollinators (for example, from bees to hummingbirds), the transition to selfing is often accompanied by changes in many aspects of floral morphology (Sicard and Lenhard 2011; Ornduff 1969). Selfing plants produce smaller flowers with less nectar and pollen. They may also produce less color and scent, and develop faster across their life-history (Sicard and Lenhard 2011; Snell and Aarssen 2005). These changes, collectively described as the "selfing syndrome" by analogy with biotic pollination syndromes, are frequently explained as the result of natural selection. Multiple selective explanations have been proposed: selection favoring more efficient self-fertilization resulting in more general floral reductions; selection to re-allocate resources from costly floral display to growth, defense or reproduction; selection for faster development resulting in general size reductions; or selection to escape insect predation on flowers. However, it is also plausible that genetic drift could cause the reductions leading to selfing syndrome (Duncan and Rausher 2013a). The effects of genetic drift are exaggerated in smaller populations, and the efficacy of natural selection is reduced (D. Charlesworth and Wright 2001; Pollak and Sabran 1992; B. Charlesworth, Morgan, and Charlesworth 1993). Selection to maintain pollinator attraction may additionally be relaxed in selfing plants (Anderson and Busch 2006). Due to the combination of these

three factors, we must take seriously the possibility that that loss-of-function mutations in floral attractant characters could spread throughout selfing populations entirely via drift and without natural selection necessarily having to drive them.

Determining whether natural selection is responsible will answer only some questions about how the selfing syndrome evolved. If selection is responsible, understanding the genetic basis of the traits involved and how they relate to each other can allow inferences about which evolutionary trajectories are more and less plausible. How genes that control traits are distributed across a genome, the degree to which traits have a shared genetic basis, and how they interact epistatically—a group of concerns collectively described as "genetic architecture"—can affect how traits evolve (Hansen 2013). Floral traits specifically have been shown to have common genetic bases and exhibit strong trait correlations (Armbruster et al. 2004; Glover et al. 2015; Ashman and Majetic 2006; Smith 2015). If selection on one trait also affects another, then distinct components of the selfing syndrome may not require distinct selective events. If, in contrast, the traits in the selfing syndrome have separate genetic bases, they likely also required separate selective events to evolve.

Study System

In my dissertation, I address these questions in a highly selfing morning glory, *Ipomoea lacunosa*, and its mixed-mating sister species *I. cordatotriloba* (Duncan and

Rausher 2013b). Both species are broadly distributed in North America, with overlapping ranges in the southeastern United States (USDA and NRCS 2017). Compared to *I. cordatotriloba*, *I. lacunosa* exhibits many classic selfing syndrome traits, including smaller, more closed flowers, less nectar and pollen, less color, and faster early growth (Duncan and Rausher 2013a; McDonald et al. 2011; Sicard and Lenhard 2011; Snell and Aarssen 2005; Chapter 2). Some of these floral differences have been shown to result from natural selection (Duncan and Rausher 2013a).

I. lacunosa is an excellent system for addressing questions about the selfing syndrome. It has a small genome of about 470Mb (Duncan and Rausher 2013b) which has been sequenced (described briefly in Chapter 1; a fuller description will be published elsewhere) and benefits from genetic resources of its close relatives *I. trifida* and the sweet potato *I. batatas* ("Sweetpotato Genomics Resource" 2017; Hirakawa et al. 2015; Schafleitner et al. 2010). Although they are partially reproductively isolated, the two species can be crossed and produce viable hybrids, which facilitates genetic investigations such as quantitative trait locus (QTL) mapping (Duncan and Rausher 2013b). Like most Convolvulaceae, they can also be cloned easily with cuttings.

Outline of Chapters

Chapter One, "Evolution of Reduced Genetic Variation and Relaxed Selection in the highly selfing species *Ipomoea lacunosa*: a Genomic Perspective," tests the predictions

of population genetic theory about genetic variation, linkage disequilibrium, and the relaxation of natural selection. In it, we use a single nucleotide polymorphism (SNP) panel derived from transcriptome sequences from 60 individual accessions across 22 populations of the two species across a large portion of their range.

Chapter Two, "Selection is Responsible for the Selfing Syndrome of *Ipomoea lacunosa*," incorporates the genetic data in Chapter One and phenotypic data from the same populations in a Qst-Fst analysis. This allows us to determine whether the relationship between patterns of genetic and phenotypic differentiation is consistent with neutral patterns of divergence or with the action of natural selection altering traits.

Chapter Three, "The Genetic Architecture of the Selfing Syndrome in *Ipomoea lacunosa*," applies a QTL mapping approach to investigate the genetic basis of 19 selfing syndrome traits, spread across five trait-category "modules." Understanding whether traits that evolved in response to natural selection are genetically independent or tightly correlated allows inferences about possible evolutionary trajectories leading to *I. lacunosa*'s selfing syndrome. In this way, we can test the predictions of some of the selective explanations that have been proposed for the selfing syndrome's evolution.

1 Evolution of Reduced Genetic Variation and Relaxed Selection in the highly selfing species *Ipomoea lacunosa*: a Genomic Perspective

1.1 Introduction

The evolution of selfing from outcrossing is one of the most common evolutionary transitions in flowering plants (Barrett 2002). This transition is expected to have a number of population genetic consequences related to reduced effective population size and heterozygosity. First, the reduction in effective population size associated with increased selfing is expected to reduce standing levels of genetic variation because effective population size is reduced in proportion to the increase in selfing rate (Pollak 1987; B. Charlesworth, Morgan, and Charlesworth 1993; Nordborg and Donelly 1997). Additionally, reduction in effective population size is expected to cause relaxed purifying selection, leading to increases in the amount of non-synonymous variation relative to synonymous variation (Glémin 2007; Hazzouri et al. 2013) and perhaps decreases in coding usage bias in the selfing species (Hazzouri et al. 2013). Finally, the reduced heterozygosity associated with increased selfing reduces the effective recombination rate and as a consequence leads to increased linkage disequilibrium.

In addition to these direct effects of increased selfing, there may be additional effects on genomic diversity due to demographic factors that may be associated with the evolution of selfing. To the extent that selfing evolves in marginal populations with

frequent extinction and recolonization events, theoretical analyses suggest that effective population size may be reduced markedly compared to that of outcrossing populations, further eroding genetic variation (Pannell JR and Charlesworth 2000; Invarsson 2002). Consistent with these analyses, a number of recent studies suggest that population bottlenecks or long periods of reduced population size associated with the evolution of selfing have reduced genetic variation to a substantially greater extent than can be accounted for by simply the reduced population size that is a direct consequence of increased selfing (Foxe et al. 2010; Guo et al. 2009; Ness, Wright, and Barrett 2010; Busch, Joly, and Schoen 2011; Pettengill and Moeller 2012).

While such bottlenecks have clearly occurred in some species, they are not necessarily universally expected. It is often the case that the evolution of high selfing rates occurs in populations that ancestrally exhibit mixed mating. In such a situation, it is possible for alleles that increase selfing rate to spread by selection through large metapopulations, incorporating much of the ancestral genomic variation present in those metapopulations, ultimately giving rise to selfers with the same effective population size as the ancestral mixed mater (except for the reduction in effective population size due directly to increased selfing). In this case, one would not necessarily expect to see a reduction in genetic diversity of greater than ½ (Pollak 1987; B. Charlesworth, Morgan, and Charlesworth 1993; Nordborg and Donelly 1997).

Another expected consequence of increased selfing is a reduction in the efficacy

of purifying selection, which arises simply as a consequence of the reduced effective population size of selfers (B. Charlesworth, Morgan, and Charlesworth 1993; D. Charlesworth and Wright 2001). In addition, population bottlenecks and other demographic factors leading to prolonged small population sizes in selfers can further reduce the efficacy of selection. Previous investigations have produced mixed support for this expectation, with some studies finding evidence for relaxed selection (Cao et al. 2011; Qiu et al. 2011; Ness, Siol, and Barrett 2012; Brandvain et al. 2013; Slotte et al. 2013), while other studies have failed to do so (Wright, Lauga, and Charlesworth 2002; Sweigart and Willis 2003; Haudry et al. 2008; Escobar et al. 2010). Given the small number of taxa represented by these studies, our ability to form generalizations on this issue are limited. Moreover, few studies have attempted to determine the extent to which any relaxed selection observed is due to simply the transition to selfing *vs.* demographic reductions in population size.

The pair of morning glory sister species *Ipomoea lacunosa* and *I. cordatotriloba* constitute an excellent system for evaluating these expectations. *I. lacunosa* is highly selfing (selfing rate = 0.95, (Duncan and Rausher 2013a)), and exhibits many typical "selfing syndrome" traits, including reduced flower size and pollen production compared to *I. cordatotriloba* (McDonald et al. 2011; Duncan and Rausher 2013b). By contrast, *I. cordatotriloba* has a mixed mating system, with an average selfing rate of only about 0.5 (Duncan and Rausher 2013a). Because this species is self-compatible, the

increase in selfing in *I. lacunosa* may have been gradual, which in turn may mean that this species may not have experienced periods of reduced population size. Although the two species can be crossed, there is a strong crossing incompatibility (Diaz, Schmiediche, and Austin 1996; Duncan and Rausher 2013b), as well as some postzygotic incompatibility (unpublished data).

Using this pair of species, we take a population genomics approach to assessing the impact of increased selfing on genomic diversity and the efficacy of selection. Specifically, we use sequenced transcriptomes from approximately 30 individuals of each species to generate a genome-wide panel of SNPs, and show that: (1) genetic diversity in *I. lacunosa* is less than one quarter of that in *I. cordatotriloba*, (2) this reduction in diversity is greater than can be accounted for by the reduced effective population size due to a transition to selfing, implying that *I. lacunosa* experienced a substantial bottleneck during or since speciation; (3) the ratio of non-synonymous variation to synonymous variation is significantly higher in *I. lacunosa*, implying a relaxation of purifying selection; and (4) this relaxation of selection is primarily due to demographic effects rather than the change in mating system; and (5) codon usage bias is greater in *I. lacunosa*, suggesting that relaxation of selection was not of sufficient magnitude to affect codon-usage bias.

1.2 Methods

1.2.1 Study system

Ipomoea lacunosa and *I. cordatotriloba* are weeds in the Batatas section of the genus *Ipomoea* (Convolvulaceae). They are broadly distributed with overlapping ranges in the southern United States; *I. lacunosa* extends north into Canada and west to Texas, while *I. cordatotriloba* is found from North Carolina to Mexico (USDA and NRCS 2017); Figure 1).

The material included in this study was obtained from field collections made by the Rausher Lab, accessions from the USDA GRIN seed bank, and a collection from the Baskin Lab (Table S1.1, Appendix). Populations were coded as sympatric if both species were collected at the same site by the Rausher Lab, as allopatric either if the site was outside of the range of the other species or if only one species was identified at the Rausher Lab's collection site, and as unknown if the accession was within the range of both species but detailed information was not available. Under this coding scheme, our sample included 13 sympatric, 16 allopatric, and 2 unknown *I. lacunosa* accessions and 11 sympatric, 14 allopatric, and 4 unknown *I. cordatotriloba* accessions (Figure 1). At sympatric sites, individual plants of the two species were typically found growing within meters of each other, and were often intertwined. We have also observed pollinators (primarily bumblebees) moving between flowers of the two species.

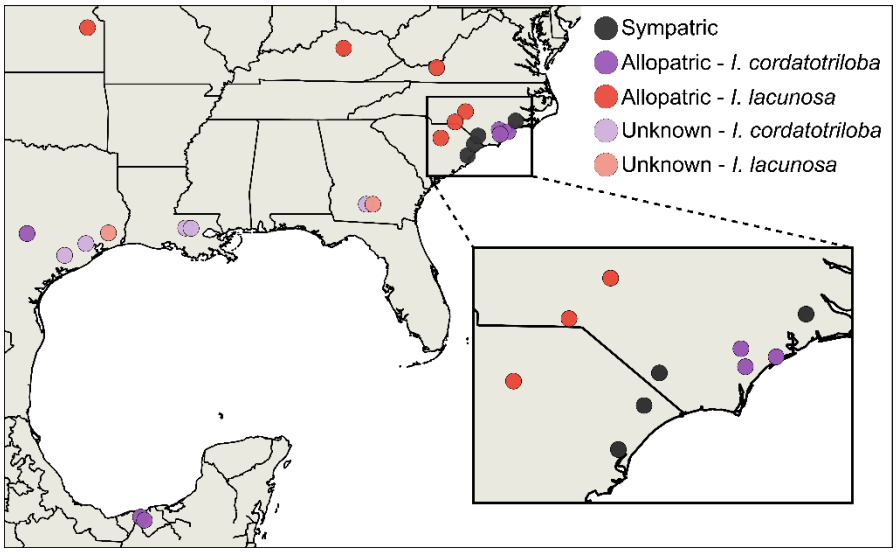


Figure 1-1 Map of the sites included in our study.
Dark grey circles indicate sympatric sites. Dark purple circles indicate allopatric I. cordatotriloba populations. Light purple circles indicated I. cordatotriloba populations for which it is not known whether I. lacunosa was growing sympatrically. Dark red circles represent allopatric I. lacunosa populations. Light red circles indicate I. lacunosa populations for which it is not known whether I. cordatotriloba was growing sympatrically.

1.2.2 Genome Sequencing and Assembly

For this investigation, we sequenced and assembled a preliminary draft of the *I. lacunosa* genome. Here we give a brief explanation of the methods used and a brief description of genome characteristics. A fuller description will be published elsewhere.

We sequenced two lanes of paired-end Illumina HiSeq 2500 100-bp reads, one HiSeq 2500 lane of a mate-pair library with a 5000-bp insert size, and 16 PacBio SmrtCells for the *I. lacunosa* draft assembly. The PacBio reads yielded approximately 50x coverage for the 500 Mb genome. The assembly was initially completed using CANU (Koren et al. 2017) using the PacBio reads, and subsequently improved with the Illumina

reads using the program Pilon (Walker et al. 2014). An initial annotation of the genome was performed using MAKER (Cantarel et al. 2008). The resulting genome had a total length of 431 Mb, with 2064 contigs, an N50 of approximately 550 Kb, and the length of the longest contig is approximately 4.7 Mb. This is consistent with a genome size of 497 Mb, as determined by flow cytometry (Duncan and Rausher 2013b). The annotation yielded 32,757 unigenes comprised of a total of 191,171 coding sequence features (exons).

1.2.3 Transcriptome Sequencing

One individual of each accession was grown in the Duke University Greenhouses from December 2014 – June 2016. For each accession, we extracted RNA from a single young leaf (0.5 – 2 cm) using a modified TRI Reagent (Sigma-Aldrich) protocol. The modifications include an additional TRI Reagent:chloroform cleanup step, adding the suggested amount of glycogen, and 3 ethanol washes before drying the RNA pellet. RNA was resuspended in 30µl of RNase-free water. RNA quality was assessed using the 2200 TapeStation system (Agilent Technologies); all samples had an RNA integrity score of 7 or higher. We generated RNA libraries starting with 4µg of total RNA using the KAPA Stranded mRNA-Seq Kit (KAPA Biosystems) and multiplexed using NEBNext Multiple Oligos for Illumina (New England BioLabs). The libraries were quality checked using the Bioanalyzer Agilent High Sensitivity DNA kit (Agilent Technologies) and Qubit Fluorometer (Thermo-Fisher Scientific). We pooled 20-21

libraries per sequencing lane; the samples were sequenced using three lanes on an Illumina HiSeq 4000 v4 platform running 150bp paired-end reads at the Duke Sequencing and Genomic Technologies Shared Resource.

1.2.4 SNP Identification

We identified SNPs using a modified version of the GATK best practices for RNASeq (Van der Auwera et al. 2013). Reads were aligned to the *I. lacunosa* draft assembly using STAR 2-pass (Dobin and Gingeras 2015). The resulting SAM files were processed using Picard tools (http://broadinstitute.github.io/picard). We called SNPs using the GATK Joint Genotyper in -erc GVCF mode. SNPs were hard-filtered according to the GATK recommendations for RNAseq data (Fisher Strand bias < 30, quality-by-depth < 2, SNP clustering) and with a minimum depth of 10 reads. We eliminated all non-variant sites and all SNPs not called in at least 60 individuals. The script used for these procedures is provided in the Supplementary Online Material. Although the minimum depth cutoff may eliminate some singletons, and thus bias our estimates of genetic diversity downwards, it should not bias estimates of differences in measures of genetic diversity between species or sets of samples.

1.2.5 Data Analysis

A set of scripts in APL (Iverson 1962) were written by one of the authors (MDR) to analyze SNP data. These scripts use as input genomic information from the *I. lacunosa* annotation GIFF3 file and transcriptome information and information about called SNPs

from GATK output files to calculate the various quantities reported. These scripts are available from the senior author upon request. Average heterozygosity for a species was the average individual heterozygosity, calculated as the proportion of individuals heterozygous at a SNP, averaged over all SNPs. This was calculated in two ways: (1) by averaging across all SNPs that were variable in at least one species, and (2) by averaging across all SNPs that were variable in the species of interest. Average π within a species was calculated by first calculating a difference value for each SNP for each pair of individuals. This value was 0 if the genotypes were homozygous and identical, 1 if the genotypes were homozygous for different alleles, and 0.5 otherwise. These values are the probabilities that an allele drawn randomly from each genotype will be different. These values were summed over SNPs, and the total divided by the total number of sites in the transcriptome (30,036,768) to yield a value of π for that pair of individuals. These π values were then averaged over all pairs of individuals within a species.

To identify whether SNPs resulted in synonymous or non-synonymous differences, we identified the coding-sequence feature in the annotated *I. lacunosa* genome that contained the SNP from information in the SNP and genome GFF3 files. We then identified the codon containing that SNP, inserted the alternate alleles into the appropriate position in the codon, and determined whether the two codons corresponded to the same amino acid or not. This process produced a dataset of 27,079 synonymous SNPs and 21,746 non-synonymous SNPs. Of the remaining 17,473 SNPs,

11,281 were classified as potentially regulatory because they occurred in non-coding regions of transcripts. The final 6,623 SNPs represent SNPs in transcripts that are not contained in the genome annotation.

To determine whether selection was relaxed in *I. lacunosa*, we asked whether π_N / π_S was greater for *I. lacunosa* than for *I. cordatotriloba.* We calculated π_S and π_N as described above, except that the average number of sites differing between individuals was divided by the number of synonymous and number of non-synonymous sites in the transcriptome (5,611,590 synonymous, 16,841,774 non-synonymous), respectively.

We also asked whether codon usage bias was more extreme in *I. lacunosa*. To do so, we first determined codon usage for each codon by counting the number of each of the 64 codons in the coding regions of the *I. lacunosa* genome (Table S1.2, supplementary online material). For any two codons encoding the same amino acid, we designated the codon with the greater number of genome-wide counts as the "more preferred" codon, and the one with the fewer number of genome-wide counts as the "less preferred" codon. Following Hazzouri *et al.* 2013, we then examined synonymous SNPs exhibiting fixed differences between the two species. For each such SNP, we determined which species used the more preferred codon, and totaled these counts for each species. The expectation is that if relaxed selection in *I. lacunosa* led to a decrease in codon usage bias, there should be more SNPs for which the more preferred codon is in *I. cordatotriloba* than in *I. lacunosa.* Because in some alternate codons for a particular amino acid have similar

genome-wide counts, it is not necessarily clear that one is more preferred than the other. To address this issue, we repeated this analysis, filtering out cases in which the two codons differ in number of genome-wide counts by less than a threshold. Specifically, if C_1 and C_2 are the genome-wide counts for two codons, with $C_1 > C_2$, we calculated the value T = C2 / C1. We then chose several different threshold values, α, (α = 1, 0.9, 0.8, 0,7, 0.6, 0.5, 0.4) and filtered out SNPs for which $T > \alpha$, corresponding to SNPs for which the difference between genome-wide counts was too small. In this process, decreasing α corresponds to increasing the stringency of this filtering.

We used bootstrapping for testing whether parameters differed between species. For each analysis we bootstrapped at two levels. First, within a species, we randomly resampled individuals with replacement, and second we randomly resampled SNPs with replacement.

All analyses were performed on two sets of samples. The first set included all individuals in the study. The second sample consisted of all samples collected from sites that were known to be allopatric, i.e. sites at which it is known that one species was present and the other absent. We examined this second set because there is the possibility of gene flow between the species at sympatric sites, which could complicate interpretation of our analyses. As will be seen, results are similar for both sets of samples.

1.2.6 Demographic Modeling

To determine whether there is evidence for *I. lacunosa* experiencing a recent period of substantially reduced population size, we used the software ∂a∂i (Gutenkunst et al. 2009) to model population size changes in the two species. In particular, we employed a model in which there was population growth (or decline) in a single ancestral population for a time *Ts*, at which point the population splits and there is a bottleneck in one of the populations (representing *I. lacunosa*). The model allows migration after the split; a more detailed analysis of the magnitude of migration, if any, will be described elsewhere. This population then recovers with exponential growth for a time *Tr*. In addition to estimating these two parameters, the model estimates *N1s*, the population size of ancestral species at the time of the split (relative to 1 at the beginning of the initial period of growth or decline), *N2s*, the population size of species 2 (*I. lacunosa*) immediately after the split, and *N2f*, the present population size of species 2, and *m*, the migration rate. We focus on this single model because we are primarily interested in whether there is evidence for a decline in the population size of *I. lacunosa*, and if so, estimating its magnitude.

Because *I. lacunosa* is highly selfing and *I. cordatotriloba* also selfs at a substantial rate, SNP alleles are not independent within an individual. Consequently, we used only one allele per individual, choosing the allele at random when an individual was heterozygous. There were thus a total of 30 and 31 alleles at each SNP for *I. cordatotriloba* and *I. lacunosa*, respectively, which we projected down to 28 individuals for each species,

yielding 23,992 synonymous SNPs for analysis. To find maximum likelihood solutions we used the BFGS optimizer Inference.optimize_log. We used a grid size of [10, 15, 20] for extrapolation. To identify the optimum parameter values, we performed searches starting at 8 widely separated parameter combinations. For each parameter combination we performed 10 perturbations to obtain starting parameter values and used a maximum of 10 iterations. Our final search used the best-inferred parameters as initial values and performed a search with a maximum of 20 additional iterations. To obtain uncertainties associated with the parameter estimates, we used ðaði's implementation of the Godambe Information Matrix approach (Coffman et al. 2015) based on 100 bootstrapped datasets. Approximate confidence intervals were estimated as the estimated value ± 1.96 x the estimated standard deviation.

1.3 Results

1.3.1 Genetic Diversity

By several measures, genetic diversity was substantially reduced in *I. lacunosa* compared to *I. cordatotriloba* (Table 1). The proportion of SNPs that were variable within *I. lacunosa* was approximately one fifth of the proportion in *I. cordatotriloba*, both when all samples were analyzed and when only known allopatric samples were analyzed. The proportion of variable SNPs with heterozygosity greater than 0 was 4.6 and 7 times greater for *I. cordatotriloba* for the two sample sets, although the average heterozygosity at variable sites was approximately the same for both species. Average within-species π

in *I. lacunosa* was less than one-sixth that of *I. cordatotriloba*. These results indicate that the reduction in genetic diversity in *I. lacunosa* compared to *I. cordatotriloba* are substantially greater than the value of ½ that would reflect the reduction in effective population size associated with a switch from complete outcrossing to complete selfing (Pollak 1987; B. Charlesworth, Morgan, and Charlesworth 1993; Nordborg and Donelly 1997). In our case, because *I. cordatotriloba* selfs about 50 percent of the time, the expected reduction is only about 1/3, which suggests that *I. lacunosa* underwent a population bottleneck at some point after its divergence from *I. cordatotriloba.*

Table 1-1 Comparison of genetic diversity parameters for the two species and the two sets of samples.
CORD = I. cordatotriloba. LAC = I. lacunosa. P is probability of no difference between species based on bootstrapping.

	All Samples			Known Allopatric Samples		
Parameter	CORD	LAC	P	CORD	LAC	P
Proportion SNPs variable within species	0.911	0.164	<0.001	0.610	0.126	<0.001
Proportion Variable SNPS with heterozygosity > 0	0.746	0.159	<0.001	0.804	0.115	<0.001
Heterozygosity (only variable sites within species)	0.00638	0.0804	>0.05	0.133	0.144	>0.05
Average π (x 106)	144	23.6	<0.001	130	22.9	<0.001

1.3.2 Linkage Disequilibrium

The reduced genetic diversity in *I. lacunosa*, coupled with increased selfing, is expected to reduce the effective recombination rate and lead to higher linkage disequilibrium in this species, compared to *I. cordatotriloba.* This expectation is realized. Maximal LD, as measured by the correlation between allele, is approximately 70 percent

greater in *I. lacunosa* (Figure 3). The distance required for LD to decay completely is about 40,000 bp for *I. lacunosa*, but only about 20,000 bp for *I. cordatotriloba.*

1.3.3 Relaxation of Selection

To determine whether there has been relaxed purifying selection in *I. lacunosa*, we calculated π_S and π_N and the π_N / π_S ratio for each species. As with the analysis of all SNPs combined, both π_S and π_N were significantly reduced in *I. lacunosa*, with this species exhibiting only 0.15 – 0.20 as much variation as *I. cordatotriloba* (Table 2A,B). By contrast, the π_N / π_S ratio was significantly higher in *I. lacunosa* for both sets of samples (Table 2A,B, Figure 2), consistent with relaxed selection in *I. lacunosa.* We also performed this analysis for variants unique to *I. lacunosa*, which we presume largely reflect variants that have arisen since the divergence of the two species. These variants exhibit a π_N / π_S ratio approximately 30 percent higher than for all variants within this species compared to variants unique to *I. cordatotriloba* (Table 2C,D), suggesting relaxed selection has been occurring for much of the time since divergence.

By contrast, a test for increased codon bias yields no evidence for relaxed selection. Regardless of the threshold value for the difference between genome-wide codon counts for two synonymous codons to be considered differentially preferred, *I. lacunosa* exhibits substantially more codon bias than *I. cordatotriloba* (Table 3). This pattern is the opposite that which would be expected if relaxed selection in *I. lacunosa* allowed more non-preferred codons to drift to fixation. One possible explanation for

this pattern is that our method of determining codon-usage bias was inappropriate because the two species differed substantially in which codons were preferred. However, this possibility seems unlikely because we found no variation, and thus no differences between species, at greater than 99 percent of all codons in the transcriptome. Differences in coding-use bias between the two species is thus minimal, justifying our determination of more preferred and less preferred codons based on the *I. lacunosa* genome. We conclude that some process other than relaxed selection is responsible for the apparently greater bias in *I. lacunosa.*

Table 1-2 Analysis of relaxed purifying selection in I. lacunosa.
CORD = I. cordatotriloba. LAC = I. lacunosa. Part A: values of πS , πN , and $\pi N / \pi S$ for the two species for the two sample sets. Part B: Significance value of bootstrap test for the indicated contrasts. Part C. values for only those SNPs unique to a given species. Part D. Significance of comparisons in part C.

	All samples			Known allopatric samples		
A	π_S	π_N	π_N / π_S	π_S	π_N	π_N / π_S
Within *CORD*	0.00110	0.000241	0.219	0.00111	0.000241	0.218
Within *LAC*	0.00017	0.0000488	0.286	0.000165	0.0000488	0.296
LAC / CORD	0.155	0.202	1.31	0.149	0.202	1.36

B	All samples	Known allopatric samples
CORD π_S *vs. LAC* π_S	P = 0.007	P = 0.003
CORD π_N *vs. LAC* π_N	P < 0.001	P < 0.001
CORD π_N / π_S *vs. LAC* π_N / π_S	P < 0.001	P < 0.001

	All samples			Known allopatric samples		
C	π_S	π_N	π_N / π_S	π_S	π_N	π_N / π_S
Within *CORD*	0.000987	0.000244	0.247	0.000907	0.000226	0.249
Within *LAC*	0.000061	0.0000234	0.383	0.000061	0.000023	0.381
LAC / CORD	0.062	0.094	1.55	0.067	0.102	1.530

D	All samples	Known allopatric samples
CORD π_S *vs. LAC* π_S	P = 0.002	P = 0.002
CORD π_N *vs. LAC* π_N	P < 0.001	P < 0.001
CORD π_N / π_S *vs. LAC* π_N / π_S	P < 0.001	P < 0.001

Table 1-3 Analysis of biased codon usage.
Entries are number of SNPs at which a species has the more preferred codon for SNPs that show fixed differences between species. Threshold is the maximum ratio of (genome-wide codon counts for less preferred codon)/(genome-wide codon counts for more preferred codon) for a SNP to be included in the counts. $\chi 2$ indicates value of $\chi 2$ for test of equality of counts for the two species.

	All samples				Known allopatric samples			
Threshold	CORD	LAC	χ^2	P	CORD	LAC	χ^2	P
1	80	112	5.33	0.025	342	494	27.63	0.001
0.9	80	112	5.33	0.025	342	494	27.63	0.001
0.8	80	112	5.33	0.025	342	494	27.63	0.001
0.7	80	112	5.33	0.025	342	494	27.63	0.001
0.6	74	106	5.69	0.025	309	466	31.85	0.001
0.5	60	89	5.64	0.025	258	367	19.01	0.001
0.4	53	81	5.85	0.025	236	343	19.77	0.001

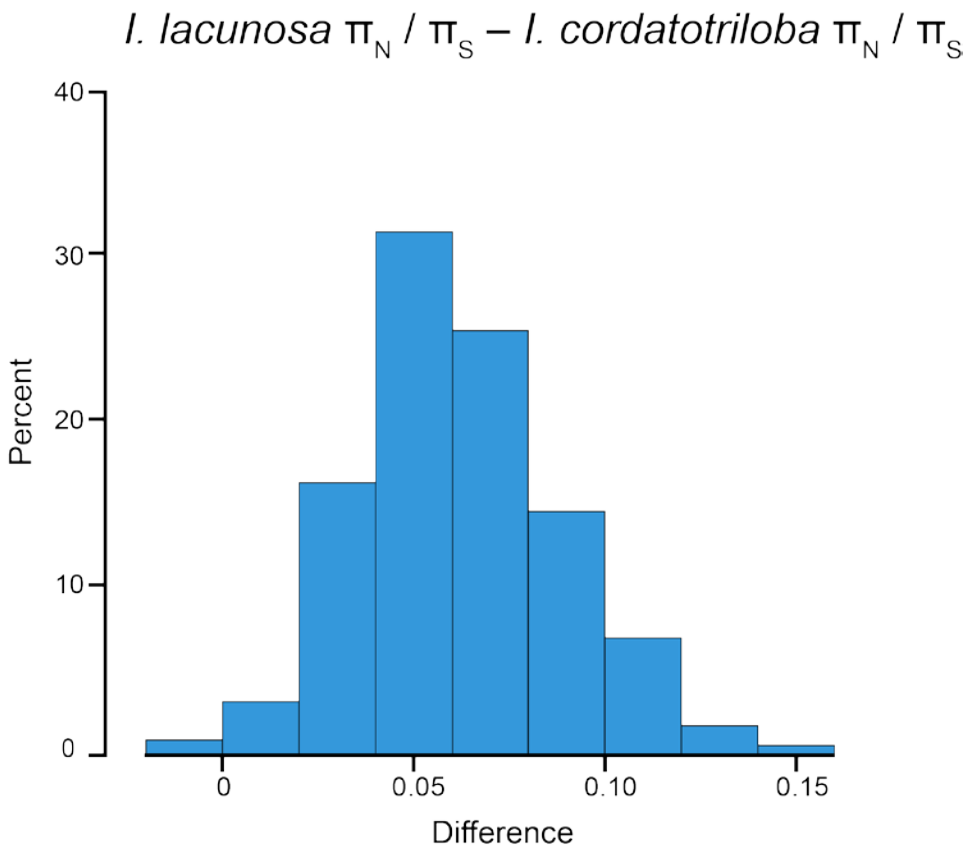


Figure 1-2 Frequency distribution of 1000 bootstrapped differences in π_N / π_S ratio between the two species.
Values greater than 0 indicate π_N / π_S is greater for *I. lacunosa.*

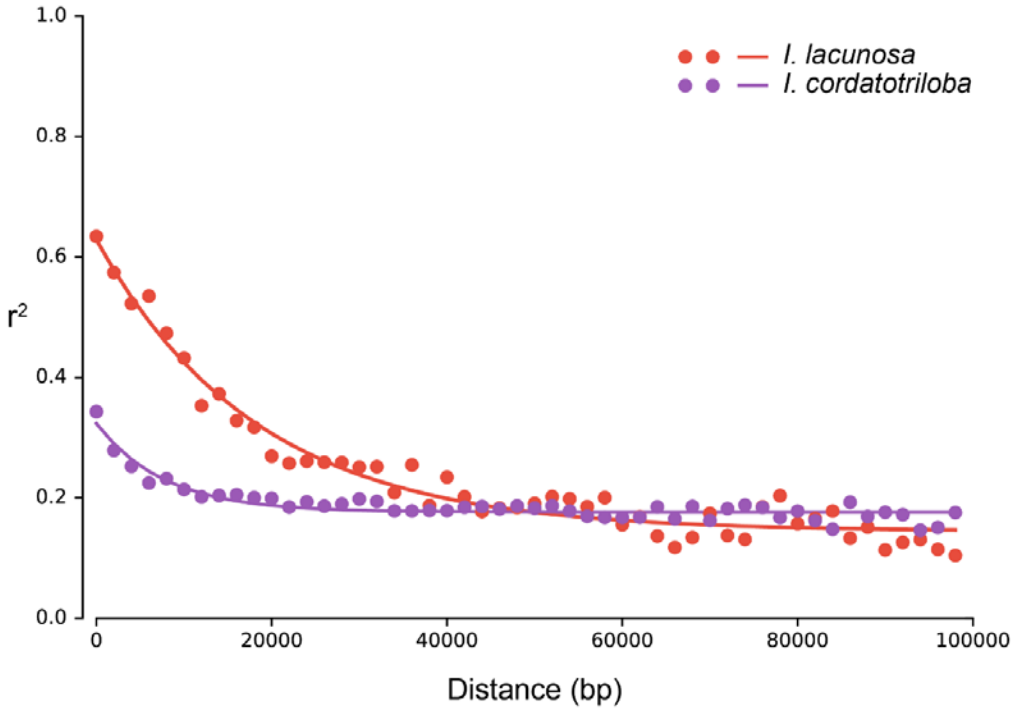


Figure 1-3 Linkage disequilibrium decay in *I. cordatotriloba* and *I. lacunosa.*
Points are averages for a given distance bin.

1.3.4 Demographic Modeling

Demographic modeling revealed a period of likely population decline in the ancestral population leading up to the split between the two species (Figure 4, Table 4), although the confidence interval for population size overlaps 1.0 and thus is consistent with either no change over time or possibly a slight increase. At the time of the split,

there is a sharp reduction in population size of *I. lacunosa*, with the estimate population size dropping to only 0.026 that of the estimate for *I. cordatotriloba* at that time. Finally, there is a slight recovery in I. lacunosa, with population size increasing approximately 8-fold to the present. By extrapolating the trend in *I. cordatotriloba* size from *Ts* to the present, the estimated present population size of *I. cordatotriloba* is $N1f = 0.318$. Relative to this population size, that of *I. lacunosa* is substantially lower ($N2f / N1f = 0.261$), although the large uncertainties means that equal final population sizes cannot be ruled out.

Table 1-4 Estimates and uncertainties of population demographic parameters.

	Parameter					
	$N1_s$	$N2_s$	$N2_f$	m	T_s	T_r
Estimate	0.386	0.01	0.083	0.808	0.337	0.069
Standard Deviation	0.387	0.076	0.283	0.918	0.418	0.331
Confidence Interval	(0, 1.44)	(0, 0.159)	(0, 0.638)	(0, 2.61)	(0, 1.156)	(0, 0.718)

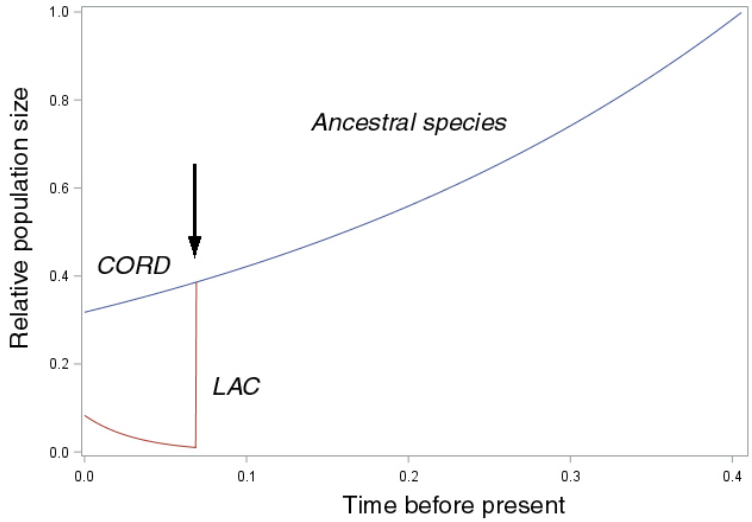


Figure 1-4 Change in population size of the two species as estimated using ∂a∂i. Population sizes are expressed as fractions of the ancestral species at the time of exponential decline. Time is expressed in units of $4 N_0 \mu$, where N_0 is the population size at time $T = T_R + T_S$ before present. The arrow indicates the time of speciation. *CORD = I. cordatotriloba. LAC = I. lacunosa.*

1.4 Discussion

1.4.1 Patterns of genetic diversity

Highly selfing species are expected to exhibit a reduction in variation compared to their outcrossing ancestors. When nearly complete selfing evolves from an obligately

outcrossing ancestor , with no demographic reduction in population size, genetic variation is expected to be reduced by ½ because the effective number of chromosomes is reduced by ½ (Pollak 1987; B. Charlesworth, Morgan, and Charlesworth 1993; Nordborg and Donelly 1997). Consequently, when sister species are compared, one of which is highly selfing, and the other highly outcrossing, the amount of genetic variation in the selfer should be approximately ½ that of the outcrosser. In the case of the highly selfing *I. lacunosa*, however, its sister species *I. cordatotriloba* is not an obligate outcrosser, but selfs approximately 50 percent of the time (Duncan and Rausher 2013a). The expected reduction in variability due to the transition to selfing in *I. lacunosa* should thus be only approximately 33 percent. Stated in another way, the amount of variation in *I. lacunosa* should be about 67 percent that of *I. cordatotriloba.*

In contrast to this expectation, *I. lacunosa* exhibits a much greater reduction in genetic variation, such that, depending on the measure of variation used (Tables 1 and 2), variation in *I. lacunosa* is only 15 – 20 percent of that in *I. cordatotriloba.* This additional reduction in variation is consistent with a substantial reduction in population size during the divergence of *I. lacunosa* from *I. cordatotriloba.* This inference is consistent with our demographic modeling, which indicates an approximately 97 percent reduction in *I. lacunosa* population size in the time right after speciation.

This pattern of reduced variation is similar to that reported for several other outcrossing-selfing pairs of species, including pairs in the genera *Mimulus*, *Capsella*,

Collinsia, and *Clarkia* (Guo et al. 2009; Brandvain et al. 2013; Brandvain et al. 2014; Hazzouri et al. 2013; Pettengill and Moeller 2012), and in several of these species the inference that the selfing species underwent a period of reduced population size has been supported by coalescent modeling. Our investigation adds to the growing list of species that have undergone population size reductions during the evolution of high levels of selfing. We note, however, that several other highly selfing species do not show similar signs historical population size reductions (e.g. *Eichhornia paniculata* (Ness, Wright, and Barrett 2010); *Leavenworthia alabamica* (Busch, Joly, and Schoen 2011); and *Arabidopsis lyrata* (Foxe et al. 2010), indicating that the evolution of high selfing rates is not invariably accompanied by population reduction. Identifying factors that determine whether or not population reduction accompanies the evolution of extreme selfing thus remains a challenging issue.

In addition to reduced genetic diversity in *I. lacunosa*, we found that this species exhibits substantially greater linkage disequilibrium than *I. cordatotriloba.* This difference is expected because the reduced diversity is expected to be associated with a lower effective recombination rate (Nordborg 2000). Similar results have been obtained in other comparisons of selfing and outcrossing sister species (Pettengill and Moeller 2012; Brandvain et al. 2014), suggesting this is a general pattern in highly selfing species. What is unclear is the extent to which this reduction is due simply to increased selfing. Because population size reduction also leads to reduced genetic variation, and thus

reduced opportunity for recombination, it may also contribute to a reduction in LD in highly selfing species. Moreover, given that in our system, as well as in others cited above, population size reduction appear to have a greater effect on genetic diversity than simply switching to extreme selfing, population size reduction has the potential to produce a much larger effect on LD. An interesting direction of future research will be to devise and apply methods for evaluating the relative magnitude of these two effects on the increase in LD seen in selfers.

1.4.2 Relaxed Selection

Our finding of an elevated π_N / π_S ratio in the highly selfing *I. lacunosa* is consistent with findings of other studies that have attempted to detect relaxed selection in selfing species (Hazzouri et al. 2013; Brandvain et al. 2013; Brandvain et al. 2014; Arunkumar et al. 2015). While relaxed selection is an expected consequence of the reduction in effective population size associated with a transition from outcrossing to selfing, it is also an expected consequence of the demographic reduction in plant numbers that we have inferred from both the magnitude of genetic variation reduction in *I. lacunosa* and from our demographic analyses. A crude estimate of the relative contribution of these two processes to the overall relaxation of selection can be obtained by evaluating the efficacy of selection. A natural definition of the efficacy of selection is the probability of fixation of a deleterious mutation relative to the probability of a neutral mutation. This probability is given by Charlesworth (2009) as $Q = Ne\,s / (1 - e^{-2}$

$^{Ne\,s}$), where *Ne* is effective population size and *s* is the magnitude of selection against the mutation (Brian Charlesworth 2009). The difference $Q_{LAC} - Q_{CORD}$ reflects the difference in relative probability of fixation, and thus the difference in efficacy of selection for the two species. This difference can be partitioned into two parts: $Q_{LAC} - Q^*_{CORD}$ and $Q^*_{CORD} - Q_{CORD}$, where Q^*_{CORD} is the value corresponding to $Ne = 2/3\ Ne_{CORD}$, reflecting the reduction in effective population size that would be expected because of the reduction in selfing rate from the transition from *I. cordatotriloba* (selfing rate = 0.5) to *I. lacunosa* (selfing rate > 0.95). From these differences in *Q* values, the proportion of the overall increase in the probability of fixation that is due solely to the increase in selfing rate is then $q = (Q^*_{CORD} - Q_{CORD}) / (Q_{LAC} - Q^*_{CORD})$, while the proportion due to population reduction is $1 - q$.

We estimated *Ne* for each species as $Ne = \pi_S / 4\mu$ (Charlesworth 2009), for various values of μ bracketing values estimated for various dicots (1.6×10^{-9} to 5.2×10^{-9} for *Solanum* (Roselius, Stephan, and Städler 2005); 7.7×10^{-9} for *Arabidopsis* (Yang et al. 2015); and 1.5×10^{-8} for the Brassicaceae (Koch, Haubold, and Mitchell-Olds 2000)). With these estimates, the change from mixed mating to selfing accounts for at most 40 percent of the change in efficacy of selection, and usually much less (Table S1.3, Appendix A) for selection coefficients of at least 10^{-6}. Although these crude estimates obviously do not take into account demographic fluctuations in the populations of either species, they suggest that reduction in absolute population size is responsible for the

majority of selection relaxation in *I. lacunosa.*

Relaxed selection in *I. lacunosa* might also be expected to lead to reduced codon usage bias in this species (Marais, Charlesworth, and Wright 2004; Qiu et al. 2011). However, we did not find reduced coding usage bias, despite the evidence for relaxed selection provided by comparison of π_N / π_S ratios. Hazzouri *et al.* (2013) likewise found evidence for relaxed purifying selection in a highly selfing species of *Collinsia* yet no evidence of decreased codon usage bias. We suggest that this pattern may be a consequence of the extremely weak selection acting to promote codon usage bias (Zeng and Charlesworth 2009). In particular, if the effective population size of *I. cordatotriloba* is already small enough to generate essentially no selection for codon bias, a further reduction of *Ne* in *I. lacunosa* is expected to have little effect (McVean and Charlesworth 2000; Qiu et al. 2011). One caveat to this conclusion, however, is that our power to detect relaxation of codon usage bias may have been low because translational selection tends to occur most prominently in highly expressed genes, and we did not examine gene expression levels. Future analyses taking expression levels into account may lead to revised conclusions.

2 Selection is Responsible for the Selfing Syndrome of *Ipomoea lacunosa*

2.1 Introduction

The evolutionary transition from outcrossing to self-pollination has occurred repeatedly in flowering plants (Barrett 2002). Typically, this transition in mating system is accompanied by reductions in floral display: highly selfing species produce smaller flowers with less nectar, less scent, shorter pedicels or cymes, and lower pollen-to-ovule ratios than their outcrossing relatives (Goodwillie et al. 2010; Sicard and Lenhard 2011; Ornduff 1969). These features of highly selfing species are termed, by analogy with biotic pollination syndromes, the "selfing syndrome." In addition, there are several traits that are not traditionally considered part of the selfing syndrome, but which often occur in highly selfing species. For example, because highly selfing plants are often weedy and occur in marginal habitats, rapid growth is often at a premium and may be favored by selection (Snell and Aarssen 2005). Additionally, although floral display size includes both the number of flowers and the size of those flowers, flower number and inflorescence size in selfing species has been less studied than flower size.

Although the selfing syndrome has evolved repeatedly in many species, the reasons for these reductions remain unclear. In particular, there has been little effort to determine whether trait reductions result from natural selection or the accumulation of mutations through genetic drift. Both alternatives are plausible. On the one hand, the repeated evolution of the selfing syndrome represents a convergent evolutionary

response to a change in mating system, and convergent evolution often is indicative of changes being driven by similar selective pressures (Stern 2013). Moreover, several types of selection have been proposed to drive selfing-syndrome evolution. For example, it has been argued that because costly floral displays are no longer needed to attract pollinators in highly selfing species, selection favors reallocation of resources away from these displays to other fitness-enhancing functions, producing fewer, smaller flowers with less nectar and scent, as well as inflorescences with fewer flowers (Goodwillie et al. 2010). Similarly, because less pollen is needed for self-fertilization, it has been hypothesized that selection reallocates resources used to produce pollen to other functions (Goodwillie et al. 2010). Alternatively, it has been suggested that reduced floral size and display size have been selected to reduce attractiveness to herbivores (McCall and Irwin 2006; Sicard and Lenhard 2011). And, as described above, it has been argued that selection often favors faster development in highly selfing species because they tend to grow in marginal habitats.

By contrast, genetic drift is also a plausible explanation for the evolution of selfing-syndrome traits. Without the necessity to attract insect pollinators, selfing plants no longer experience purifying selection to maintain display traits, and may accumulate mutations that reduce floral size, pollen production, nectar production and display size through genetic drift (Duncan and Rausher 2013a). Moreover, self-pollination increases homozygosity, which reduces effective population size, thereby increasing the potential

effects of genetic drift (D. Charlesworth and Wright 2001; Pollak and Sabran 1992).

Experimental evolution and field experiments have implicated selection in reducing the distance between reproductive organs under pollinator limitation (Moeller and Geber 2005; Roels and Kelly 2011). However, to our knowledge, the role of selection in shaping selfing-syndrome traits that do not directly affect selfing rate has been shown only once, for corolla size traits in the highly selfing morning glory *Ipomoea lacunosa* (Duncan and Rausher 2013a). This study used a Qst-Fst approach to demonstrate that divergence in corolla size and anther-stigma separation between *I. lacunosa* and its mixed-mating sister species *I. cordatotriloba* was driven by natural selection.

Qst-Fst comparisons between quantitative traits and neutral markers offer a method for differentiating between selection and drift as possible causes for population divergence (Leinonen et al. 2013). When the genes responsible for individual traits are not known, either because the traits are polygenic, because the system is a non-model species, or both, molecular signatures of selection (Nielsen 2005) cannot be used to detect selection on divergent traits. In these situations, a Qst-Fst approach, which uses phenotypic measurements to estimate the differentiation in loci underlying quantitative traits of interest and comparing that differentiation to neutral markers, can differentiate between selection and drift as explanations for phenotypic divergence.

In this investigation, we have adopted this approach to determining the extent to which natural selection contributed to evolution of the selfing syndrome in the

divergence between *I. lacunosa* and *I. cordatotriloba.* This study greatly extends the initial investigation by Duncan and Rausher (2013) in several ways. First, it includes traits besides measurements of floral size, including nectar and pollen production and display size (flower number). It thus allows us to determine whether selection may have caused the evolution of some syndrome traits (e.g. flower size), but not others. Second, it examines traits that are not traditionally considered part of the selfing syndrome, but are thought be correlates of high selfing rates (inflorescence size and growth rate). Finally, this study includes a broader array of populations and loci sampled.

2.2 Methods

2.2.1 Study system

Ipomoea lacunosa and *I. cordatotriloba* are weeds in the series Batatas of the Convolvulaceae (USDA and NRCS 2017). The two species have overlapping distributions in North America; I. lacunosa is found in the Eastern United States from Florida to Canada and west to Texas, and I. cordatotriloba occurs from Mexico to North Carolina. We have observed both species growing intertwined in the same habitat.

Both species are self-compatible. However, *I. lacunosa* is highly selfing (selfing rate > 0.95), while *I. cordatotriloba*'s selfing rate varies among populations and averages around 0.5 (Duncan and Rausher 2013a). The highly selfing *I. lacunosa* produces small, white flowers with less pollen than *I. cordatotriloba*, which produces medium purple flowers. Crossing barriers separate the two species but artificial hybrids can be produced

(Diaz, Schmiediche, and Austin 1996; Duncan and Rausher 2013b).

2.2.2 Samples and plant culture

For this study, we used plants covering a wide geographic range obtained from the Rausher Lab's field collections, an accession from the Baskin Lab, and accessions from the USDA's GRIN seed bank (see Table S1, Chapter 1). The majority of plants used were selfed in the greenhouse for at least one generation, however four were derived from field-collected seeds or seeds produced in the greenhouse from field-collected plants and were thus potentially open-pollinated. Individuals from the same female parents from each site, but not the same individuals, were used for genetic and phenotypic measurements.

Seeds were scarified and planted in four-inch pots in Fafard 4P soil and maintained in a growth room under 16-hour days at 78 degree Fahrenheit. After four weeks, conditions were changed to 13-hour days at 65 degrees Fahrenheit to trigger flowering. When flower buds appeared, plants were moved to the greenhouse. Because we have observed that some floral traits in these species are plastic between growth room and greenhouse conditions, plants were maintained under greenhouse conditions for two weeks before the onset of measurements.

2.2.3 Phenotypic measurements

We included measurements of 9 selfing-syndrome characters in our analysis: early growth (first three internodes), total flowers per day, corolla length, corolla width,

nectar volume and nectar sugar concentration, and total nectar sugar. As controls, we also measured the vegetative traits of leaf length:width ratio, leaf dissection and sepal length. Previous research in these species has found that floral traits (corolla length) diverged in a manner consistent with selection but vegetative traits (leaf length:width ratio) did not. This study includes a wider range of floral attraction and display traits as well as a flower-adjacent vegetative trait (sepal length).

Early growth measurements of the first three internodes (cotyledons to first true leaf, first true leaf to second true leaf, and second true leaf to third true leaf) were taken with a rule on day 21. After plants were moved to the greenhouse, floral measurements were performed between 8:00 and 12:00. Because flowers remain open for less than a day, it is not necessary to standardize flower age. Three flowers were measured on different days from each individual. Corolla length and width, cyme length and style length were measured using a digital caliper (Mitutoyo Digimatic CD6" CS).

Nectar volume was quantified as follows: the day before the flower opened, the bud was capped with a plastic straw covered with parafilm. The following morning, all nectar was extracted from the base of the flower with a 2µl microcapillary tube (Drummond Scientific). The height of the nectar in the tube was measured with the digital caliper. Because each tube is 32 mm long and holds 2 µl in total, this measurement was converted to volume with the following equation: V = (2µl * height of nectar in tube)/32 mm. Nectar sugar concentration was quantified by expelling all of the

nectar from the microcapillary tube onto a Master-53M ATAGO refractometer. The refractometer was standardized with water at the start of the day's measurements. Because refractometer readings are often imprecise with low volumes, the nectar was diluted with 2.5µl water before measurement. Refractometer readings were converted into concentrations according to the methods of Bolten (Bolten et al. 1979). Pollen production was quantified by removing anthers the day before anthesis, allowing them to dry overnight, and resuspending them in 500uL 70% ethanol. Both the size and number of pollen grains were measured using a Coulter counter (). Leaf length, width and degree of dissection (on a five-point scale from all leaves entire to all leaves lobed) were scored for three mature leaves on each individual.

2.2.4 Phenotypic comparisons

To determine which traits differed between the two species, we performed a nested ANOVA using "Fit Model" in JMP Pro 13 (SAS Instititute 2017) using restricted estimation maximum likelihood. For each trait, we analyzed a model in which species was included as a fixed effect, site nested within species and female parent nested within site nested within species as random effects. Measurements on two selfed offspring per female parent provided the error effect. We applied a sequential Bonferroni correction for multiple comparisons (Rice 1989).

2.2.5 SNP calling

To derive distributions of Fst, we called SNPs from leaf transcriptomes

(described fully in Rifkin et al. 2017 / Chapter 1). Briefly, whole RNA was extracted from leaf transcriptomes using a modified TRI-Reagent (Sigma-Aldrich) protocol. Libraries were prepared using NEBNext Illumina barcodes and sequenced on three lanes of Illumina Hi-Seq 4000. SNPs were called using a modified version of the GATK best practices for RNASeq (Van der Auwera et al. 2013) and hard-filtered for coverage and quality. Synonymous SNPs were identified from our draft annotation (Rifkin et al. 2017 / chapter 1).

To control for non-independence of linked SNPs, we performed all analyses on a subset of SNPs separated by at least 40kb. Analysis of linkage disequilibrium indicated that LD decays to background levels within 40kb for both species (Chapter 1). This filtering produced a final set of 3,893 SNPs for estimating Fst.

2.2.6 Calculation of Fst and Qst

Fst was calculated for each SNP using the method described by Weir and Cockerham (1984). Overall Fst was then calculated by applying the weighted average across loci that they describe.

Because our analysis of characters revealed little evidence of population structure within each species, we ignored populations when calculating Qst. In this situation, Qst is defined by the formula

$$Qst = \sigma^2_{BetwSp} / (\sigma^2_{BetwSp} + \sigma^2_{A}) \qquad \text{Eq. (1)}$$

where σ^2_{BetwSp} is the between-species component of variance and σ^2_{A} is the additive

genetic variance for the trait (Whitlock 2008). For each character, we estimated σ^2_A from the results of a nested ANOVA with Species as the top-level effect, Female Parent nested within Species, and offspring nested within Female Parent. Because the offspring were produced by selfing, the Female Parent variance component, σ^2_{BetwFP}, represents approximately $(1 + F1)\sigma^2_A$, where $F1$ is the inbreeding coefficient of the female parent (Holland, Nyquist, and Cervantes-Martinez 2003); Eq. 5b). Although for both species $F1$ is likely high, it is unknown. Consequently, as a conservative approach, we assume $F1 = 0$. This is conservative because it maximizes the estimate of σ^2_A and thus minimizes Qst. Our estimate of Qst is thus

$$Qst = \sigma^2_{BetwSp} / (\sigma^2_{BetwSp} + \sigma^2_{BetwFP}) \qquad \text{Eq. (1)}$$

We used bootstrapping to compare Fst and Qst (Duncan and Rausher 2013a), employing an APL program written by MDR. For each bootstrap sample, we sampled both loci and individuals within species randomly with resampling. To test for a significant difference between Fst and Qst, we generated 500 bootstrap estimates of Fst and Qst. Representing these estimates as vectors **Fst** and **Qst**, we calculated the difference vector, **D** = **Qst** —**Fst** and determined the proportion of **D** elements that were ≤ 0. We performed this analysis only on the 10 characters that showed a significant difference between species. We applied a sequential Bonferroni correction for multiple comparisons.

2.3 Results

2.3.1 *Differences between species*

Using a nested ANOVA, we found little evidence for population differentiation within species (Table S2.1). After a sequential Bonferroni correction, the population effect was not significant for any of the traits. This does not necessarily mean there is no real population differentiation, only that we were not able to detect it, probably because we scored a maximum of 3 accessions per population.

Generally, traits typically associated with the selfing syndrome (corolla length and width, nectar volume and nectar sugar concentration) were significantly smaller or lower in *I. lacunosa*, even after a sequential Bonferroni correction (Table 1). In early growth, internodes on day 21 after germination were significantly longer in *I. lacunosa*. In addition, number of flowers produced per day was significantly greater and cyme length significantly shorter in *I. lacunosa*. All of these differences are in the expected direction for selfing-syndrome traits. By contrast, vegetative traits (leaf length/width ratio, leaf dissection, and sepal length) did not differ significantly between the species.

Table 2-1 Trait means, standard deviations and nested ANOVA results.
Asterisks (*) indicate traits where species difference was significant at the 0.05 level after a sequential Bonferroni correction. Relative divergence was calculated as the absolute value of the difference between the two species means divided by whichever species mean was larger for the trait in question.

Trait	*I. lacunosa* mean (st. dev.)	*I. cordatotriloba* mean (st. dev.)	Nested ANOVA, species effect (F ratio, P value)	Relative divergence
Internode 1 (mm)*	13.37 (3.71)	5.62 (2.12)	87.8342, P<0.0001	0.58
Internode 2 (mm)*	13.35 (6.78)	6.44 (2.96)	21.6556, P=0.0001	0.52
Internode 3 (mm)*	52.20 (20.50)	21.15 (14.86)	47.4394, P<0.0001	0.59
Flowers per day*	4.08 (1.79)	2.04 (1.33)	17.33, P=0.0011	0.5
Flowers on inflorescence	2.08 (0.80)	2.37 (1.24)	1.27, P=0.273	0.13
Cyme length (mm)*	9.46 (3.39)	16.67 (10.27)	8.6962, P=0.0082	0.43
Corolla length (mm)*	20.01 (1.41)	31.10 (2.95)	202.3758, P<0.0001	0.36
Corolla width*	15.09 (1.35)	32.34 (3.89)	480.7105, P<0.0001	0.53
Nectar volume (μL)*	0.792 (0.37)	3.08 (0.89)	111.5224, P<0.0001	0.74
Nectar sugar concentration (mg/mL)*	21.90 (70.82)	31.12 (42.39))	50.8611, P<0.0001	0.31
Leaf length/width	1.50 (3.04)	1.10 (0.11)	0.7128, P=0.407	0.27
Leaf dissection	0.12 (0.21)	0.25 (0.23)	3.783, P=0.067	0.54
Sepal length (mm)	11.00 (0.85)	10.82 (1.00)	0.44, P=0.514	0.02

2.3.2 *Fst-Qst analysis*

By comparing the genetic differentiation at neutral markers, Fst, to the morphological differentiation, Qst, we can assess whether trait differentiation is consistent with selection. If Qst > Fst, that suggests that trait divergence is greater than can be accounted for by drift alone. If Qst = Fst, however, trait divergence is consistent

with neutral processes.

In performing this analysis, we analyzed only the nine traits that exhibited significant differences between the two species. For all of the floral traits, except number of flowers produced per day and cyme length, the bootstrap analysis indicated that Qst > Fst, with all comparisons significant at the $P < 0.05$ level after a sequential Bonferroni correction (Fig. 1). In addition, Qst was significantly greater than Fst for all three internode lengths. Thus, divergence in some floral traits appears to have been driven by natural selection. However, flowers produced per day and cyme length do not appear to have diverged as a result of selection. As none of the vegetative traits had diverged significantly between species, we did not perform Fst-Qst comparisons for them: traits that are not diverged cannot have diverged because of natural selection.

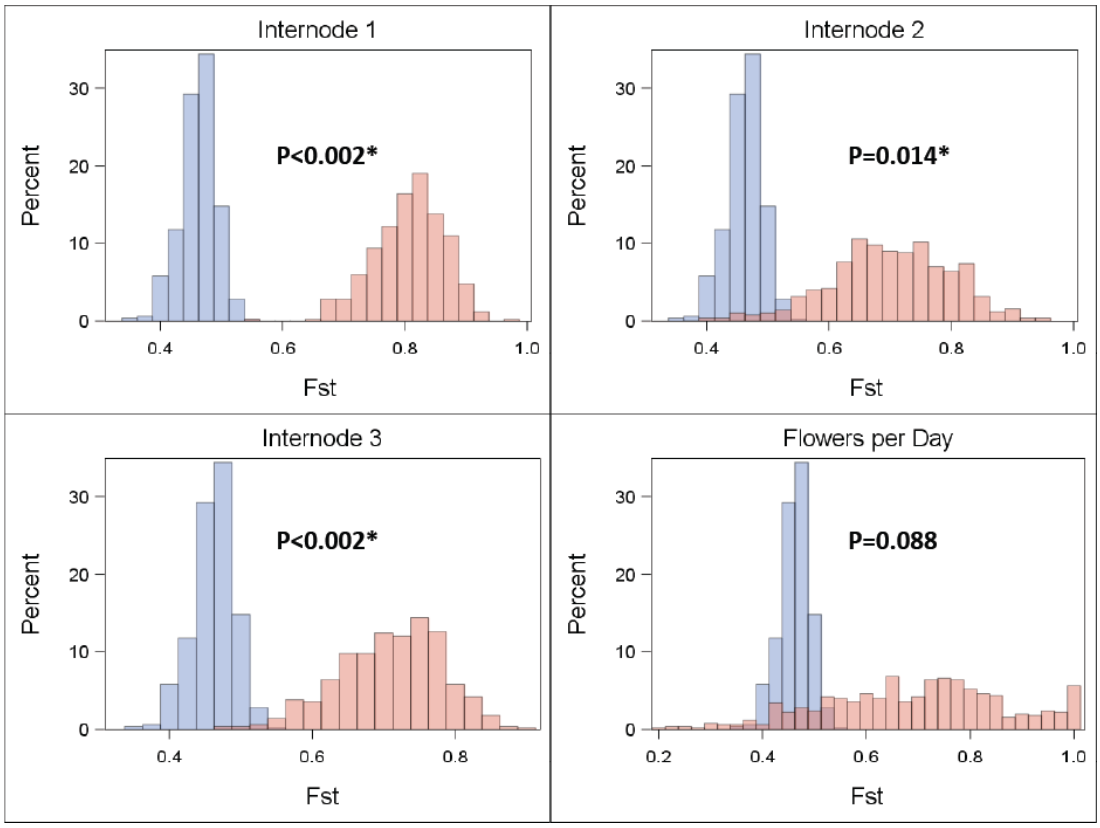
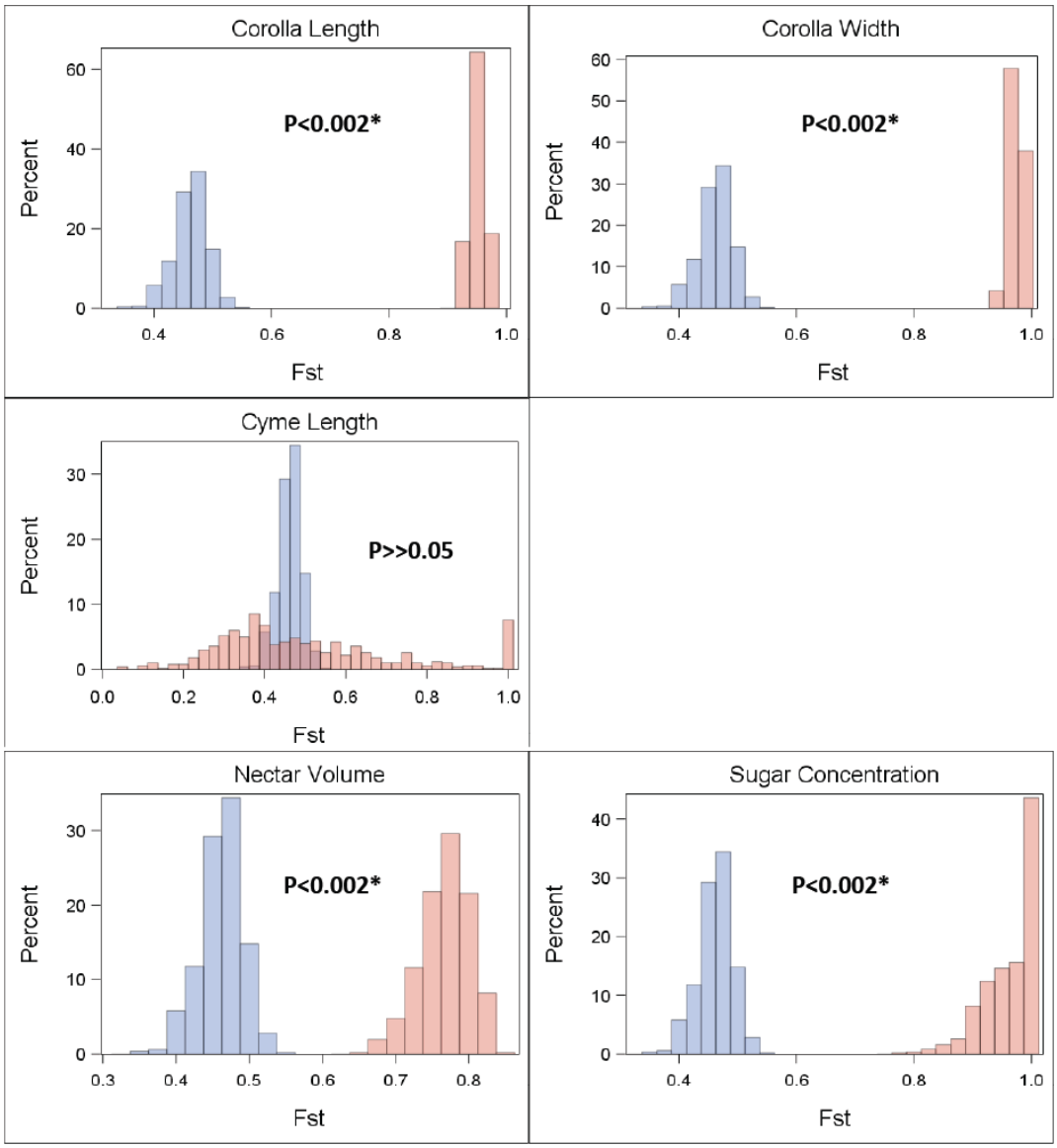


Figure 2-1 Fst vs Qst analyses.
Traits: early growth (Internode 1, Internode 2, Internode 3), life history (Flowers per Day), floral morpohology (Corolla length, Corolla Width), Inflorescence (Cyme length) and Nectar (Nectar volume, Sugar Concentration) traits. Fst values in blue, Qst values in red. Asterisk (*) indicates significant after a sequential Bonferroni correction for multiple comparisons.



2.4 Discussion

The floral and life-history changes that constitute the selfing syndrome are distinct from morphological and physiological changes that directly affect selfing, such as reduced herkogamy and loss of self-compatibility. Although these associated floral and life-history changes have been observed in a wide range of taxa, and various

selective explanations have been proposed for them, to our knowledge there has been only one other demonstration that selection is indeed responsible for these changes (Duncan and Rausher 2013). Confirming that selection is responsible for these changes is a key prerequisite for understanding how and why they evolve. Genetic drift is also a plausible explanation for these floral reductions, given that selfing plants experience relaxed selection to attract pollinators and their reduced effective population size makes them more susceptible to the effects of drift (D. Charlesworth and Wright 2001; Pollak and Sabran 1992).

In this study, we have extended Duncan and Rausher's (2013) original analysis of corolla length and width in *I. lacunosa* and *I. cordatotriloba* to include early growth and nectar traits. In addition, we have sampled a much larger number of presumably neutral loci (their study used only 4 microsatellite loci), and have sampled individuals from a geographically wider range. We have demonstrated that *I. lacunosa* has evolved classic selfing syndrome traits: reduced flower size and reduced nectar production. Moreover, we have shown that divergence between species in these traits was likely driven by natural selection, based on our Fst-Qst analysis. Determining that traits did indeed diverge under selection is a necessary prerequisite to identifying what form that selection took.

Several selective explanations have been proposed for the evolution of the selfing syndrome. Selection to increase selfing rates by reducing anther-stigma separation could

cause general reductions in flower size (Sicard & Lenhard 2011). Highly selfing species may redirect resources away from floral display and from pollen production to other functions (Goodwillie et al. 2010). Or, selection to reduce flower size may be mediated by flower-eating herbivores (McCall & Irwin, 2006; Sicard & Lenhard, 2011). Finally, selection for generally faster development in marginal habitat might lead to general size reductions. Differentiating among these explanations and identifying which is acting in any given example of the selfing syndrome should be a major focus of future research into the selfing syndrome.

In addition to these traditional selfing-syndrome characters, we also examined a set of early growth characteristics. Researchers have long observed an association between selfing, weediness and an annual life history, and selfing plants have been found to develop faster at several life stages (Snell and Aarssen 2005; Fishman et al. 2015). We found that stem elongation during the first three weeks after germination was significantly faster in *I. lacunosa*, consistent with evolution of more rapid growth. And as with floral characters, our Fst-Qst analysis indicates that this divergence was most likely caused by natural selection.

By contrast, although *I. lacunosa* produces more flowers per day than *I. cordatotriloba*, a pattern again consistent with rapid growth and early flowering, our analysis provided no evidence that this difference was caused by natural selection. The relationship between flower size, flower number and mating system is complex. On the

one hand, both the flower size and flower number components of display size increase outcrossing (Goodwillie et al. 2010). On the other hand, there are costs and tradeoffs associated with producing many flowers. Some studies have found evidence for a tradeoff between flower size and flower number (Ashman and Majetic 2006; Sargent et al. 2007). In self-compatible species, a larger number of flowers per plant also increases the likelihood of geitonogamous selfing, which may be selected against both because of inbreeding depression and because of a cost to outcross male fitness (Barrett and Harder 1996). The pattern we observe between *I. lacunosa* and *I. cordatotriloba* - in which the selfing species produces an increased number of flowers that is not consistent with selection - may thus also reflect the breakdown of regulatory mechanisms that limit the number of flowers in *I. cordatotriloba* in response to the relaxation of selection to limit flower number.

We also found no evidence that selection was responsible for the divergence in cyme length. As cyme length is correlated with corolla length and corolla width (Chapter 3), it is possible that selection on these traits was sufficient to cause observable phenotypic divergence but not to lead to a Qst distribution that did not overlap the Fst distribution. Both phenotypic patterns in the populations measured and QTL results (Chapter 3) are consistent with this. All three traits are significantly diverged, and the mean relative divergence between species is similar (cyme length 0.43, corolla length 0.36, corolla width 0.53; Table 1). However, unlike corolla length and corolla width, the

two species' phenotypic distributions for cyme length overlap extensively (Figure S2.1). Cyme length varies widely across *I. cordatotriloba* populations, but that range of variation is dramatically reduced in *I. lacunosa*, covering a small subset of the *I. cordatotriloba* variation. Although the traits are only moderately correlated (corolla length and cyme length 0.350, corolla width and cyme length 0.338; Chapter 3), the single cyme length QTL identified overlaps QTLs for both corolla length and corolla width and accounts for 28% of the difference between grandparents in the mapping population (Chapter 3). The simple genetic architecture of cyme length and its correlation with traits that were under selection could thus account for the dramatic reduction in cyme length in *I. lacunosa* through selection for reduced corolla length and width leading to the fixation of a low corolla length allele in *I. lacunosa*, while the trait's broader phenotypic range in *I. cordatotriloba* leads to a wide distribution of Qst that overlaps Fst.

In contrast to early growth, we found no evidence that vegetative traits had diverged at all. In general, floral and vegetative traits are not tightly correlated (Ashman and Majetic 2006), so it is not surprising that selection on floral traits did not result in change in vegetative traits. However, the early growth differences we observed might lead to the expectation of more general vegetative differences in response to a more marginal habitat, as leaf shape is known to correlate with ecological context (Nicotra et al. 2011). As the ranges of the two species overlap, it seems likely that both are subject to similar pressures on vegetative growth and therefore have not diverged. It is also

possible that measurements more sensitive to physiological differences (e.g. specific leaf area or leaf dry weight) might reveal significant differences in vegetative characteristics, especially since *I. lacunosa*'s range extends much further north than *I. cordatotriloba*'s and therefore it may have experienced selection to tolerate a broader range of temperatures (Wilson, Thompson, and Hodgson 1999; USDA and NRCS 2017). However, a study in *Clarkia* found no relationship between photosynthetic rate (a physiological trait associated with drought tolerance) and selfing rate (Ivey et al. 2016). In general, the morphological divergence between the two species is far greater for floral and early growth than for vegetative traits.

Our Fst-Qst comparison clearly indicates that multiple traits involved in the selfing syndrome diverged in response to selection. Determining the pattern of selective pressures led to that divergence will depend, in part, on an understanding of the relationship between the traits: if all are tightly correlated genetically, then direct selection on a single trait (corolla length, for example) could lead to divergence in all traits. If, on the other hand, the traits are genetically distinct, each would have to change separately in response to direct selection on that character. In the next chapter, a quantitative trait locus (QTL) mapping approach will reveal that correlations between floral and early growth traits, and even among floral traits, are generally not strong, and QTLs underlying these traits show generally low levels of overlap (Chapter 3). Taken together, these results this strongly suggest that the selfing syndrome in *I. lacunosa* is the

result of multiple changes across the genome in response to strong direct selective pressures on the different traits.

3 The Genetic Architecture of the Selfing Syndrome in Ipomoea lacunosa

3.1 Introduction

Transitions from outcrossing to self-pollination have occurred repeatedly across the angiosperm phylogeny (Barrett 2002; Barrett, Arunkumar, and Wright 2014). These transitions in mating system are frequently accompanied by a set of stereotyped changes in floral morphology. Compared to their outcrossing and mixed-mating relatives, selfing species typically exhibit reductions in flower size, color, scent, nectar production and pollen-to-ovule ratio (Sicard and Lenhard 2011). This suite of characters is termed the "selfing syndrome" (Sicard and Lenhard 2011) by analogy with biotic pollination syndromes (Fenster and Armbruster 2004).

In addition to these floral changes, several recent studies have found that some life-history traits are also correlated with the transition to high selfing rates. For example, a phylogenetic study of selfing and outcrossing species found that selfing species exhibit faster development at many life-history stages (Snell and Aarssen 2005). Fishman et al. (2015) demonstrated that self-pollinating *Mimulus* flower earlier than outcrossing relatives (Fishman et al. 2015). There is some theoretical justification for a relationship between these phenological transitions and high selfing rates: high selfing rates tend to be associated with species inhabiting marginal habitats, where there is a selective premium on rapid development (Snell and Aarssen 2005). These considerations suggest that in addition to the traditional suite of floral traits, rapid

growth and earlier flowering should be considered for inclusion in the selfing syndrome.

The repeated, correlated evolutionary change in a suite of characters associated with the transition to high selfing rates raises three fundamental questions. First, does the selfing syndrome evolve as a result of genetic drift, or in response to natural selection? Second, if selection is responsible, what selective agents are responsible? Finally, do these traits evolve via *independent evolution* consisting of mutations in different genes and direct selection on each trait, or through *correlated evolution*, as would occur if the traits are genetically correlated and some trait changes result from an indirect response to selection on others?

These questions can be addressed in part by identifying the genetic architecture of the traits involved using a quantitative trait locus (QTL) approach. QTL analyses can provide information about whether syndrome traits evolve independently. If QTLs for different traits do not co-localize, then one may conclude that those traits have evolved independently. By contrast, co-localization of QTLs may suggest that traits are genetically correlated, and thus do not evolve independently, although this conclusion must be further substantiated by identifying the causal loci underlying the QTLs. Gene identification also allows analysis of whether molecular signatures of natural selection are present, which can eliminate the hypothesis that syndrome traits evolved primarily by genetic drift (Nielsen 2005). The relationship between QTL allelic effects and species

differences can also support or contradict a role for selection: if allelic effects are generally in the same direction as species differences, a history of directional selection is suggested (Orr 1998).

In this investigation, we focus on assessing the degree to which selfing-syndrome traits evolve independently by determining the degree to which QTLs for syndrome traits co-localize. In particular, we attempt to determine where the species we study lie along a continuum of genetic architecture. At one extreme of this continuum, genetic substitutions affecting syndrome traits are highly pleiotropic, resulting in changes in all syndrome traits being mediated by changes in the same genes. In this situation, we expect QTLs to co-localize for all traits. At the other extreme, substitutions uniquely affect individual traits, and no QTLs co-localize.

Understanding where a species lies along this continuum has the potential to provide information about the nature of the selection pressures that have molded the selfing syndrome. For example, in many species, the evolution of high selfing rates is mediated by the evolution of style and anther lengths to bring these organs into contact (Chang and Rausher 1998; Motten and Stone 2000; Schueller 2004; Takebayashi, Wolf, and Delph 2006). To the extent that other syndrome traits are pleiotropically affected by these changes, which is likely if they share similar developmental pathways, selection *for increased selfing* is responsible for the evolution of these other traits, which evolve as correlated responses. If, however, genetic changes affecting anther and stigma position

do not also affect other floral traits, the evolution of other floral traits cannot be due to a correlated response to selection for increased selfing. Instead, there must be another cause: either genetic drift, or selection on floral traits distinct from selfing (e.g., selection to redirect resources away from unnecessary pollinator attractants). As another example, if genetic changes causing floral size reduction pleiotropically affect pollen and/or nectar production, selection to reduce flower size (to save resources, reduce florivory, or improve selfing efficiency) would indirectly cause reduction in pollen or nectar production. By contrast, if those genetic changes do not also affect nectar and/ or pollen production, reduction in those traits must have been independent of other traits.

Several studies in four other taxa have used a QTL mapping approach to determine the degree to which traits in the selfing syndrome have evolved independently: *Mimulus* (Lamiales:Phrymaceae; Lin and Ritland 1997; Fishman, Kelly, and Willis 2002; Fishman et al. 2015), *Solanum* (Solanales:Solanaceae, Bernacchi and Tanksley 1997; Georgiady, Whitkus, and Lord 2002), *Leptosiphon* (Ericales:Polemoniaceae, Goodwillie, Ritland, and Ritland 2006) and *Capsella* (Brassicales:Brassicaceae, Slotte, Hazzouri, and Stern 2012). All these studies included measurements of flower size and shape and the structure of reproductive organs, and all except one included a measure of either self-incompatibility or anther-stigma separation. Developmental timing traits, inflorescence size, pollen number and ovule number were each included in two studies. In general, these investigations found high and positive

correlations among syndrome traits across F2 or backcross individuals. This result is consistent with many of the same genetic changes influencing different syndrome traits; however, it does not definitively demonstrate this type of genetic architecture because such correlations may arise even if different mutations cause change in different traits. In particular, if two different mutations, one affecting each trait, are fixed independently but are genetically linked, even relatively loosely, they will tend to be inherited together in a mapping population and thus contribute to a positive correlation.

Better evidence for particular genetic changes affecting multiple syndrome traits is provided by demonstrations that QTLs for different syndrome traits co-localize. Most studies of the selfing syndrome have included multiple floral morphology traits, but few have examined pollen number, phenology or inflorescence size. This limits our capacity to make inferences about the selfing syndrome as a whole. In the studies that have investigated them, QTLs for syndrome traits tended to exhibit a moderate degree of QTL co-localization. To quantify and compare QTL co-localization across studies, we averaged across a matrix of pairwise overlaps for all traits included in the study. This provided a measure which ranged between 0 (no QTL overlapped any other QTL) to 1 (every QTL overlapped every other QTL). Across all traits, the average pairwise overlap between QTLs for selfing syndrome traits varies widely between studies (Bernacchi and Tanksley 1997, 0.31; Lin and Ritland 1997, 0.23; Georgiady 2002, 0.35; Fishman 2002, 0.66; Goodwillie 2006, 0.54; Slotte 2012, 0.36; Fishman 2015, 0.47). These results are consistent

with a moderately correlated genetic architecture of syndrome change (i.e. genetic changes exhibiting pleiotropic effects on several syndrome traits), but also suggest that many genetic changes were independent in different characters. However, the small number of studies, especially those including inflorescence size, phenology and pollen production, precludes conclusive inferences about how general these results are, and thus suggests the need for further investigations on different species. Moreover, there is a need to incorporate additional syndrome characters (e.g. flower color and nectar production) that have not previously been examined.

We present here the results of a QTL mapping study designed to identify genomic regions that contribute to evolution of selfing-syndrome traits in the wild morning glory *Ipomoea lacunosa*. Previous studies in this species, which has diverged from the more outcrossing *I. cordatotriloba*, have demonstrated that the reduction in floral size in *I. lacunosa* was driven by natural selection (Duncan and Rausher 2013a; Chapter 2). In addition, we have shown that selection was also responsible for faster early growth and decreased investment in nectar (Chapter 2). Our primary goal is to determine the extent to which QTLs for different syndrome traits co-localize, and thus evaluate the extent to which the evolution of different syndrome traits may have been independent or correlated. In particular, we ascertain the likelihood that natural selection on floral size documented previously caused correlated responses in other selfing-syndrome traits. In doing so we include important syndrome traits that have

seldom been examined previously (e.g. reduction in pollen and nectar production, loss of floral pigmentation, and phenological characters). A secondary goal is to identify genomic regions that can be further targeted for identification of genes associated with the evolution of selfing syndrome traits.

3.2 Methods

3.2.1 Study system

Ipomoea lacunosa and *I. cordatotriloba* are weeds in section Batatas of the genus *Ipomoea* (Convolvulaceae; (Diaz, Schmiediche, and Austin 1996). The species have overlapping distributions in the southeastern United States (USDA and NRCS 2017). Previous research has demonstrated that *I. lacunosa* is highly selfing (selfing rate 0.95), that *I. cordatotriloba* populations vary in selfing rate from (0 to 0.95, with an average of 0.5), and that the morphological divergence in corolla size between the two species is due to selection (Duncan and Rausher 2013a).

3.2.2 Plant material

Seeds were collected from a wild *I. lacunosa* population in Kinston, North Carolina (35.23971, -77.57392) and a wild *I. cordatotriloba* population in Conway, South Carolina (33.94713, -79.01940) (Duncan and Rausher 2013b). These were inbred for one generation to generate the lab strains "Lacunosa ProTruck" (LPR) and "Cordatotriloba Double A" (CAA). Individuals from these strains were crossed, using LPR as male and CAA as female, to generate F1 hybrids. A single F1 hybrid was selfed to generate an F2

mapping population.

The F2 hybrid mapping population was planted in spring of 2014 and flowered in fall and winter of 2014. Seeds were scarified with sandpaper and germinated in the dark in petri dishes. Any seeds that were not visibly swollen within two days were re-scarified with a razor blade. When taproots were visible in most individuals, the seeds were transferred to 4" plastic pots filled with Fafard 4P soil and maintained under 16-hour days a growth room at approximately 80°F for six weeks. The light cycle was changed to 13-hour days for two weeks at 65°F to induce flowering, after which plants were transferred to the Duke Research Greenhouses and maintained under greenhouse conditions with 13-hour days.

3.2.3 Phenotypic measurements

3.2.3.1 Life history traits

We measured seven early growth traits (days to opening of first three leaves, height on day 21 from germination, length of first three internodes on day 21 from germination) and two floral phenology traits (average measurement date and whether any flowers were produced) in 424 F2 individuals. Early growth traits were also measured in selfed offspring of the grandparent individuals (N=46 *I. lacunosa,* 47 *I. cordatotriloba*). Total height and internode length were measured with a rule. Leaf opening was scored as the point when focal leaf unfolded to an angle of greater than 90 degrees. Early growth traits in F1 individuals were scored in a separate growout in

summer of 2013 (N=13 *I. lacunosa*, 14 *I. cordatotriloba* and 7 F1). Because morning glories are fast-growing vines and required frequent pruning, the first flower measured was not necessarily the first flower produced. However, measurements were generally taken soon after the onset of flowering. Therefore average measurement date is an approximation of onset of flowering. Not all F2s produced flowers (this was confirmed by collecting seed capsules at the end of the experiment: plants not recorded as having flowered also had no selfed seeds, indicating that we did not simply miss the opportunity to measure them), so whether a plant ever produced flowers was also included as a binary trait.

3.2.3.2 Floral traits

Flowers were measured daily between 9AM and 12PM. Phenotypic measurements were taken opportunistically from all flowering individuals on different days until at least 5 flowers had been measured (mean 3.76). Additional measurements were taken from individuals that produced many flowers to confirm consistency of flower morphology over time. Floral traits were measured at the same time from cloned individuals of the *I. cordatotriloba* and *I. lacunosa* grandparents and the F1 parent of the mapping population. Measurements from the grandparents were compared using a two-tailed t-test with unequal variances.

We measured nine floral traits (color, flowers open per day, cyme length, flowers or buds on cyme, corolla length, corolla tissue length, corolla width, anther-stigma

position, nectar volume and style-plus-ovary length) in the 357 F2 individuals that produced flowers and in clones of the grandparents and the F1. These can be roughly grouped into four floral trait categories, affecting different tissues: flower size and shape traits ("floral morphology"), characters of the inflorescence and floral vegetative tissue ("inflorescence"), nectar, and pollen.

Corolla measurements are described in Figure 1. All floral traits were measured externally on intact flowers from the base of the ovary except for style-plus-ovary length, which was measured on dissected flowers. Because flower shape (measured as length / width) has also been shown to differ between these species (Duncan and Rausher 2013a), two measures of flower shape (corolla length / corolla width and corolla tissue length / corolla length) were included in analyses.

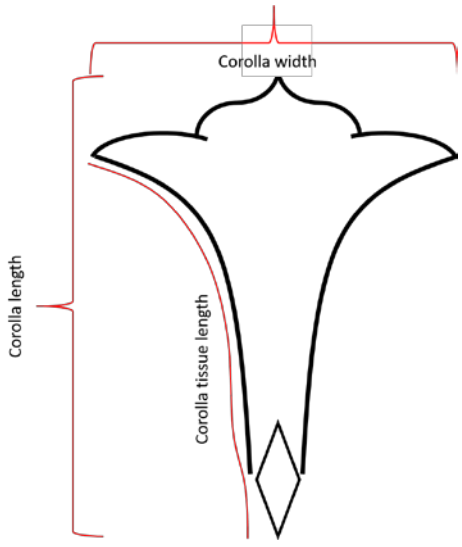


Figure 3-1 Floral morphological measurements of corolla length, corolla width and corolla tissue length.

Floral measurements were taken with a digital caliper (Mitutoyo Digimatic CD6"

CS). Anther-stigma separation was scored using a method adapted from (Duncan and Rausher 2013a): if no anthers touched the stigma, it was scored as either -1 (stigma below all anthers) or 1 (stigma above all anthers). If one anther touched the stigma and the rest did not, it was scored as either -0.5 (stigma below all but one anther) or 0.5 (stigma above all but one anther). If the stigma touched 2-5 anthers, anther-stigma separation was scored as 0. Nectar was measured by applying a capillary tube (Drummond, 0.4mm diameter) to the nectary of a flower and measuring the height of nectar removed with digital calipers; volume was calculated according to the formula for volume of a cylinder ($V=\pi r^2 h$).

As both species produce the same number of ovules (McDonald et al. 2011), pollen:ovule ratio can be measured using pollen number. We measured pollen number from 1-2 flowers each of 112 F2 individuals and the grandparents and F1 using aliquots of lactophenol-stained pollen and ImageJ software (Schindelin et al. 2012) according to the methods of Costa and Yang (2009). In addition, number and diameter of pollen from an overlapping set of 1-4 flowers each from 68 F2 individuals and the grandparents and F1 were measured on a Coulter Z2 counter by a collaborator.

Mean measurements were calculated for every individual in R using the function **aggregate** (Team 2016). Pearson correlation coefficients and significance were calculated with the function **rcorr** in the R package **Hmisc** (Harrell Jr, with contributions from Charles Dupont, and many others. 2016), and we applied a Bonferroni correction for

multiple comparisons. Because plants varied inconsistently and multiple measures were obtained from only a subset of individuals, we compared analyses of the phenotypic data across three different cutoffs for number of flowers sample per individual (1, 2 and 5). Sample size cutoffs for these comparisons were made based on the number of flowers sampled, ignoring missing variables. QTL analyses were performed using all individuals, including those with only one flower measured.

3.2.4 Marker identification

To develop markers for constructing a linkage map and identifying QTLs, we used double-digest restriction-site associated DNA sequencing (ddRAD), a reduced-representation genotyping-by-sequencing approach (Peterson et al. 2012). Under this approach, samples are uniquely identified with a two-part combinatorial barcoding index: after restriction enzyme digest, up to 96 in-line forward barcodes are ligated to the sticky end left by one restriction enzyme, and up to 12 unique Illumina sequencing primers are added to the other end of the DNA fragment using a PCR primer adapter ligated to the sticky end left by the second restriction enzyme. Restriction enzymes were selected based on an *in silico* digest of the *I. lacunosa* draft assembly, using the R package **simrad** (Lepais and Weir 2016).

We isolated DNA using ThermoFisher GeneJet Plant DNA kits from young leaf tissue frozen at -80°C, and homogenized using a Geno Grinder 2000 (SPEX Sample Prep). DNA samples were digested using EcoR1 and MSP1 (New England Biolabs)

according to the manufacturer's instructions and cleaned with AMPure XP PCR purification kits (Agencourt). To ensure optimal conditions for adapter ligation, concentrations of the cleaned, digested samples were quantified using a Qubit dsDNA BR assay kit (Thermo Fisher). Ligation molarities were calculated using both the NEBiolabs calculator (New England Biolabs) and the calculator provided by Peterson et al. 2012, with similar results. We ligated 96 eight-base in-line forward barcodes (Parchman, Gompert, and Buerkle 2011) to 100ng per sample of DNA using T4 DNA ligase (New England Biolabs) according to the manufacturer's instructions. At this point, samples were pooled by plate and a gel size selection was performed by the Duke Sequencing and Genomic Technologies Core Resource (DSGTCR). These five pools (four of 96 samples, one of 37 samples) were then amplified using PCR primers containing 5 separate Illumina Sequencing Primers (Peterson et al. 2012), pooled, and sequenced by the DSGTCR.

We sequenced five lanes on the Illumina HiSeq (XXX) platform. Differences in the size ranges selected led to uneven coverage. We initially resequenced the original library, which did not improve the coverage of low-coverage individuals. We therefore compensated by resequencing the 192 lowest-coverage individuals in two successive rounds of one and two lanes.

3.2.5 Bioinformatics pipeline

We used the Genome Analysis Tool Kit (GATK) pipeline to identify SNP markers

following the suggested best practices for RADSeq in a nonmodel organism (Van der Auwera et al. 2013). Raw reads were demultiplexed into a file for each sample using Sabre (https://github.com/najoshi/sabre). Reads from each individual were aligned to the *I. lacunosa* draft assembly reference genome using BWA-MEM (Li 2013). SAM files were converted to BAM files and prepared for analysis with the Picard tools **CleanSam**, **AddOrReplaceReadGroups**, **FixMateInformation**, and **SortSam** (http://broadinstitute.github.io/picard.). Samples from multiple lanes representing the same individual were combined with the Picard tool **MergeSamFiles** . Genotypes were called in individuals using the GATK **HaplotypeCaller** in –ERC mode. All resulting g.vcf files, as well as 215 F3 individuals from a separate experiment not described here, were genotyped together using the GATK **JointGenotyper**.

Genotype calls were hard-filtered as follows: SNPs were separated from indels, all SNPs were filtered by quality-by-depth, Fisher strand bias and mapping quality using GATK's **VariantFiltration** tool. Individual genotypes were filtered by minimum genotype quality (GQ>15). The resulting genotype calls were exported to a SNP table using the GATK tool **VariantsToTable**, and then were filtered using a custom Python script as follows: multiallelic markers, markers genotyped in fewer than 100 individuals, and singleton calls were removed. Using a second custom Python script, markers were compared against a variant database assembled from whole-genome sequencing of the two grandparents of the mapping population. Only markers in which both grandparents

were homozygous for different alleles were retained. For all alleles and genotypes described, "A" refers to the *I. lacunosa* reference and "B" to the *I. cordatotriloba* non-reference allele.

Because selfed F3 offspring from the F2 mapping population had been genotyped for another experiment, additional genotypes were imputed as follows using a custom Python script: any marker that was heterozygous in an F3 but uncalled in its F2 parent was imputed as a heterozygote. This supplied 146887 missing calls in 210 F2 individuals.

3.2.6 Linkage map construction

Map construction started with a raw dataset of 16968 sites genotyped in 422 individuals (including the grandparents and F1 parent of the mapping population). Genotype coverage was 50% and average genotype frequencies over all loci were AA:AB:BB = 24:48.8:27.3.

The data were filtered in R/QTL (Broman, Wu, and Churchill 2003) according to the recommendations in (Broman 2010): individuals with fewer than 1000 markers genotyped and markers genotyped in fewer than 375 individuals were removed. 151 markers with significant segregation distortion were removed for initial map construction. The filtered dataset included 1652 markers in 421 individuals with 90% coverage and average genotype frequencies of AA:AB:BB 22.7:53.8:23.5.

We used R/QTL and MSTMap to assemble the markers into linkage groups,

which produced 16 linkage groups (N=15; **XXX**). The sixteenth linkage group consisted of eight markers on contig 1898 of the draft assembly spanning 159 kilobases.

We used MSTMap (Wu et al. 2008) to order the markers (see Appendix XX for settings). This produced the same linkage groups as R/QTL with the exception of a few singletons that did not fit into any linkage group. We imported the marker order from MSTMap into R/QTL and removed the markers that MSTMap did not find positions for.

After marker ordering, the map was adjusted using the R/QTL functions **ripple** and **droponemarker**. We tested for segregation distortion using the R/QTL function **geno.table** and applied a Bonferroni correction to the results of the chi-square test for deviations from 1:2:1 segregation. Suspected genotyping errors were converted to NA calls according to the methods described in (Broman 2010).

3.2.7 QTL mapping

For all phenotypes except color, nectar volume and whether a plant ever flowered, we scanned for QTLs using the **scanone** function with Extended Haley-Knott regression in R/QTL (Broman and Sen 2009; Broman 2010). For color and whether a plant ever flowered, we used a binary model, and for nectar volume, we used both Extended Haley-Knott regression and a non-parametric analysis because of the skewed distribution. Significance thresholds were generated by Extended Haley-Knott permutation analysis (Broman and Sen 2009). We compared our QTL to both genome-wide and chromosome-specific significance thresholds (Garner et al. 2015).

The phenotypic effect of the genotypes at the marker nearest to each QTL was calculated in R using the **effectplot** function in R/QTL (Broman, Wu, and Churchill 2003). From this, additive effect was estimated as the difference between the trait value of the two homozygotes divided by two. Dominance was calculated as ABS(trait value of heterozygote – average of homozygotes)/additive genetic effect.

We used two methods to quantify overlap among QTLs. First, we created a pairwise matrix of traits and averaged across this matrix. For every pair of traits, we counted in both directions the percentage of QTLs that overlapped: if two traits had two and five QTLs respectively and there was a single overlap, then 0.5 and 0.2 would be entered. We also counted the number of other traits with overlapping QTLs for every trait.

3.2.8 QTL sign test

The distribution of QTL effect sizes and directions can provide evidence for the action of past natural selection. In particular, if the direction of QTL effects coincides with the direction of differences between putatively selected groups more often than expected by chance, that suggests that they fixed as the result of selection (Orr 1998). For traits affected by at least 4 significant QTLs, we used the R function **fitdistrplus** (Delignette-Muller and Dutang 2015) to assess the distributions of our data and performed a QTL sign test according to the methods of Muir et al. (2014). As measures of effect size, we used both the relative homozygous effects of QTLs and the proportion of

variance explained. We estimated the differentiation between the grandparents in terms of the standard deviation of each grandparent separately. Results were very similar for all measures of effect size and standard deviation.

3.3 Results

3.3.1 Grandparental and F1 hybrid differences

There were consistent and significant morphological, inflorescence, pollen, nectar and life history differences between the grandparents of the mapping population. They differed both in the morphological and reproductive traits typically included in the selfing syndrome (corolla length and width, style-plus-ovary length, pollen production) and in floral and life-history traits that are seldom included in studies of the selfing syndrome (color, inflorescence size, nectar production, early growth; Table 1).

The differences between the flowers of the grandparents of the mapping population are consistent with a pattern of reduced investment in floral display in the highly selfing species, and broadly consistent with the differences identified between the two species more generally in Chapter 2. *I. lacunosa* produces smaller, narrower flowers (i.e., with a greater length-to-width ratio) on shorter inflorescences, containing less nectar and less, smaller pollen. In contrast to the flowers, the self-pollinator's early growth is increased rather than reduced. Both grandparents produce their first leaves at the same rate, but *I. lacunosa*'s stem elongation is dramatically accelerated such that seedlings are 58% taller at day 21. Although they are differentiated between the species

(Tanya M Duncan and Rausher 2013b; Chapter 2), neither anther-stigma separation nor flowers produced per day differed between the populations used as parents of our mapping population.

In general, floral morphological traits in the F2s exhibited partial dominance of the *I. cordatotriloba* grandparent: cyme length, corolla length, corolla width and style-plus-ovary length in the F2s were all significantly higher than the midparent value. In contrast, the F2s produced significantly less nectar than the midparent value (Table 1; 1-sample t-test against midparent average with a Bonferroni correction for multiple comparisons, $p << 0.0001$). By one measure of pollen count, F2s produced less pollen than either parent (1-sample t-test against midparent average with a Bonferroni correction for multiple comparisons, $p << 0.0001$) and by the other slightly more than *I. lacunosa* but dramatically less than *I. cordatotriloba*. Onset of flowering and flower production were highly variable; 64 individuals (15.2%) never produced flowers under greenhouse conditions, and only 250 (59.4%) produced three or more measured flowers. Plants that produced three or more flowers began flowering an average of 35 days earlier, grew more buds per inflorescence on larger inflorescences, produced wider corollas and produced more nectar (Table S1). Thus, the earlier-flowering plants more closely resembled their *I. cordatotriloba* grandparent. This correlation between phenological and morphological traits suggests overlap of the genetic bases of these traits, and may reflect differences that have evolved in day-length or temperature cues

for flowering between the two species.

Early growth measurements were collected from all 421 F2 individuals. For height and internode elongation, the faster growth of *I. lacunosa* appeared dominant ($p<0.001$; Table 1). In contrast to the F2s, the F1 individual generally did not deviate from the midparent value in floral traits after correcting for multiple comparisons. However, in a separate growout, F1s demonstrated faster earlier growth that could reflect dominance of the *I. lacunosa* early growth traits or heterosis (Table S2).

Floral traits of F2s showed a greater variety of distributions than early growth traits. Cyme length, corolla measurements and style-plus-ovary length were normal or normal-like. However, flowers per day, flowers per cyme and nectar production were all dramatically skewed. Color was distributed bimodally, and the frequency of purple flowers did not differ from the 3:1 ratio expected for a trait controlled by a single locus (X-squared = 3.47, df = 1, P = 0.062).

Table 3-1 Grandparent, F2 and F1 trait means for life history and floral traits phenotyped.
Traits marked with an asterisk (*) differed significantly between the parents of our mapping population after correcting for multiple comparisons (T-test, except for pollen traits marked with a ^, which were compared with a Welch's T-test because of small sample sizes).
Abbreviations: , L1, first leaf open, L2, second leaf open, L3, third leaf open, H, height day 21, I1, internode 1 day 21, I2, internode 2 day 21, I3, internode 3 day 21, NL, number of leaves day 21, D, flower date, CyL, cyme length,FI, flowers on inflorescence, CL, corolla length, CtL, corolla tissue length, CW, corolla width, CL-W, corolla length / width, CTL-CL, corolla tissue length / corolla length, SL, style length, AS, anther-stigma position, NV, nectar volume, PCI, pollen count ImageJ, PCC, pollen count Coulter, PD, pollen diameter
IL M, *Ipomoea lacunosa* mean, IL SD, *Ipomoea lacunosa* standard deviant, IC M, *Ipomoea cordatotriloba* mean, IC SD, *Ipomoea cordatotriloba* standard deviation, p IL-IC, p value of t-test comparing *I. lacunosa* and *I. cordatotriloba*, F2M, F2 hybrid mean, t F2-M, t value of F2 vs. midparent t-test, p F2-M, p-value of F2 vs. midparent t-test, F1M, F1 hybrid mean, t F1-M, t value of F1 vs. midparent t-test, p F1-M, p-value of F1 vs. midparent t-test.

	IL M	*IL SD*	*IC M*	*IC SD*	p IL-IC	F2M	t F2-M	p F2-M	F1M	t F1-M	p F1-M
L1	11.53	3.13	11.22	2.78	0.62	11.6	1.3	0.19	NA	NA	NA
L2	13.78	3.35	13.07	2.95	0.29	13.59	1.02	0.31	NA	NA	NA
L3	16.04	3.39	15.5	3.15	0.43	15.92	0.83	0.41	NA	NA	NA
H*	470.61	246.26	297.28	179.97	<.0001	442.08	4.94	<.0001	NA	NA	NA
I1*	15.2	3.56	6.47	1.75	<.0001	11.56	3.91	<.0001	NA	NA	NA
I2*	17.02	6.44	7.72	3.2	<.0001	16.38	6.87	<.0001	NA	NA	NA
I3*	67.41	28.45	22.72	12.92	<.0001	63.27	11.49	<.0001	NA	NA	NA
NL	5.74	2.79	6.19	3.03	0.44	5.96	0.29	0.78	NA	NA	NA
D	NA	NA	NA	NA	NA	158.65	NA	NA	NA	NA	NA
CyL	15.88	5.62	20.92	6.05	0.03	23	12.24	<.0001	16.05	-2.96	0.03
FI	1.15	0.36	1.41	0.49	0.12	1.42	5.53	<.0001	NA	NA	NA
CL*	2.15	1.5	30.74	2.23	<.0001	26.65	22.16	<.0001	24.55	-1.22	0.28

CtL*	3.15	1.82	35.53	2.5	<.0001	30.73	24.85	<.0001	28.22	-0.34	0.75
CW*	4.15	1.82	35.53	2.5	<.0001	21.33	13.84	<.0001	20.47	1.04	0.34
CL-W*	1.35	0.14	1.19	0.1	0	1.26	-1.42	0.16	1.2	-2.38	0.06
CTL-CL	1.13	0.04	1.16	0.04	0.04	1.15	8.19	<.0001	1.15	0.78	0.47
SL*	11.16	0.69	17.52	1.26	<.0001	15.91	24.35	<.0001	15.58	8.57	0
AS	-0.08	0.18	-0.03	0.12	0.44	0.1	NA	NA	NA	NA	NA
NV*	0	0	1.31	0.83	<.0001	0.21	-37.44	<.0001	0.4	-2.47	0.06
PCI*^	450	NA	946.25	NA	NA	372.53	-12.42	<.0001	251.25	NA	NA
PCC*^	140	16	706	162.6	0.0396	204.4	-15.10	<.0001	323.5	NA	NA
PD*^	69.54	0.136	81.29	1.07	0.0043	69.32	-14.70	<.0001	71.85	NA	NA

3.3.2 Correlations among traits

Examination of character correlations across F2 individuals can identify which traits are genetically correlated and suggest which traits may have diverged between the two species because of linkage or pleiotropic effects in underlying genes. Trait correlations are illustrated in Figure 2, Table 3, and Table S3. In general, life history traits were strongly correlated among themselves. There were strong positive correlations between leaf number, height and internode elongation, all of which negatively correlated with leaf opening (i.e., plants that opened leaves earlier were taller and had more leaves at day 21; Figure 2). Days to leaf opening and leaf number at day 21 were moderately correlated with flowering date and flowers produced per day: plants that leafed out later and had fewer leaves at day 21 and flowered earlier. Since leaf traits did not differ between species, this may reflect variation in nutrient uptake or microclimate rather than another aspect of the inter-specific growth rate differences.

Floral traits were more strongly correlated within than between subgroups. The floral morphological traits of corolla length, corolla width, and style-plus-ovary length were all significantly, positively correlated. These traits were also significantly, positively correlated with nectar volume and cyme length, although the magnitude of correlation was smaller. Pollen production, however, was unrelated to all other floral traits. Pollen number and pollen diameter were significantly, positively correlated, (not shown in Figure 2) although the correlation remained significant after correcting for

multiple comparisons only for one measure of pollen number (Coulter counter). This pattern suggests that rather than experiencing a trade-off between number of pollen grains and investment in each pollen grain, *I. lacunosa* has evolved to jointly reduce investment in both number and size. There were moderately strong correlations between some life history traits and pollen traits (e.g. days to leaf 3 and pollen diameter, 0.46) but none that were significantly different from 0 after applying a Bonferroni correction to correct for multiple comparisons. These correlations are also on average stronger in the smaller subset of pollen assessed with the Coulter counter than in the larger sample of pollen measured in ImageJ. As pollen was collected and quantified from the plants that produced the most flowers, these correlations therefore probably reflect the trends described above in which plants flowered abundantly.

Significant correlations among traits were moderately strong, with correlation coefficients ranging from 0.23-0.91 (mean 0.48) for floral traits and 0.22-0.96 (mean 0.66) for life history traits. Correlations across these two sets of traits were generally not significant, suggesting that genes involved in divergence of floral morphological traits had few pleiotropic effects on timing traits and *vice versa*. Flowering date, which correlated with both early growth (leaf opening, number of leaves at day 21) and floral (cyme length, flowers per inflorescence, style-plus-ovary length, nectar volume, corolla tissue length) was the only trait that appeared in both groups, which may reflect the action of general regulators of growth and development.

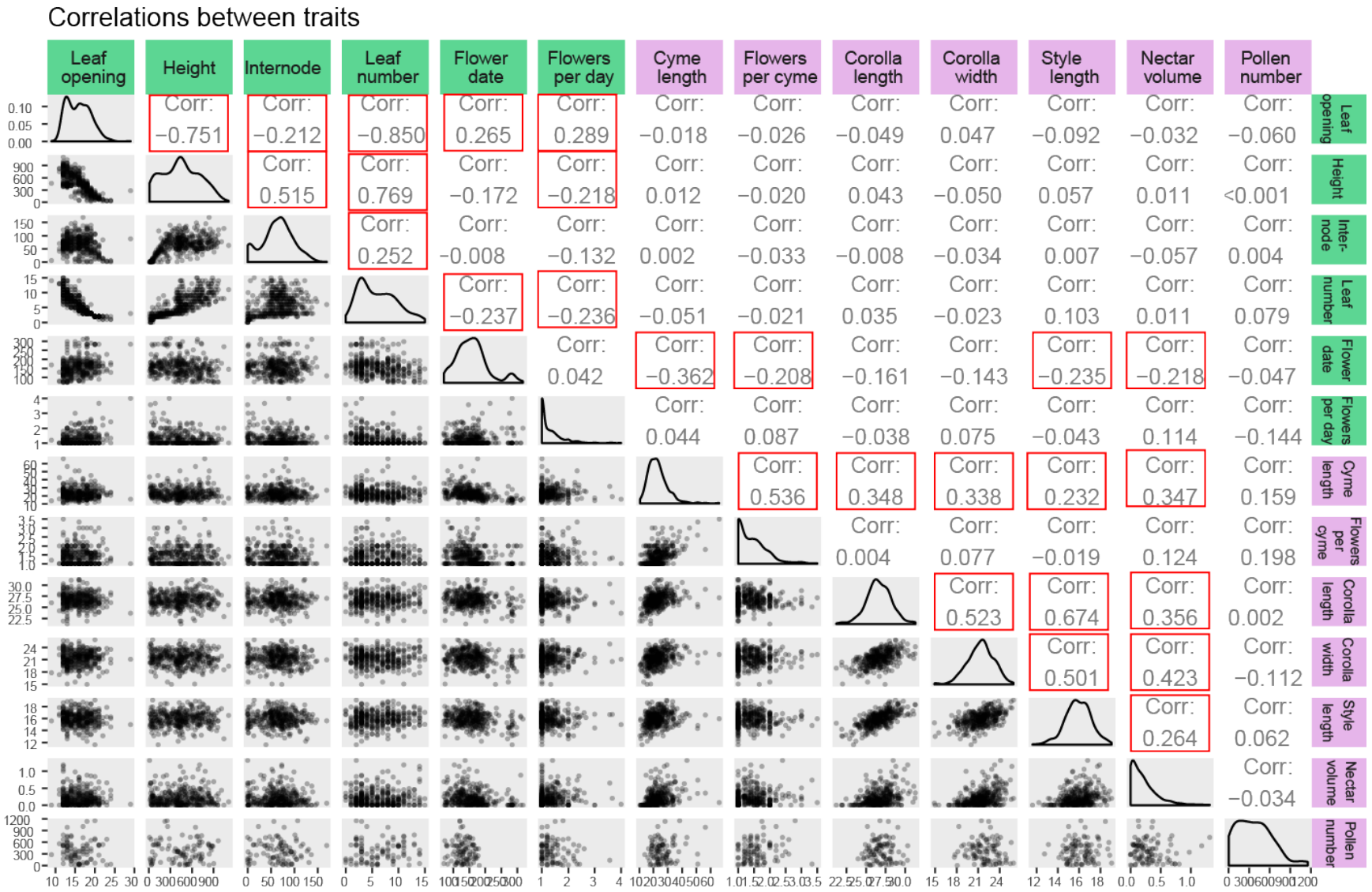

Correlations between traits
Leaf opening
Height
Internode
Leaf number
Flower date
Flowers per day
Cyme length
Flowers per cyme
Corolla length
Corolla width
Style length
Nectar volume
Pollen number
Corr: −0.751
Corr: −0.212
Corr: −0.850
Corr: 0.265
Corr: 0.289
Corr: −0.018
Corr: −0.026
Corr: −0.049
Corr: 0.047
Corr: −0.092
Corr: −0.032
Corr: −0.060
Corr: 0.515
Corr: 0.769
Corr: −0.172
Corr: −0.218
Corr: 0.012
Corr: −0.020
Corr: 0.043
Corr: −0.050
Corr: 0.057
Corr: 0.011
Corr: <0.001
Corr: 0.252
Corr: −0.008
Corr: −0.132
Corr: 0.002
Corr: −0.033
Corr: −0.008
Corr: −0.034
Corr: 0.007
Corr: −0.057
Corr: 0.004
Corr: −0.237
Corr: −0.236
Corr: −0.051
Corr: −0.021
Corr: 0.035
Corr: −0.023
Corr: 0.103
Corr: 0.011
Corr: 0.079
Corr: 0.042
Corr: −0.362
Corr: −0.208
Corr: −0.161
Corr: −0.143
Corr: −0.235
Corr: −0.218
Corr: −0.047
Corr: 0.044
Corr: 0.087
Corr: −0.038
Corr: 0.075
Corr: −0.043
Corr: 0.114
Corr: −0.144
Corr: 0.536
Corr: 0.348
Corr: 0.338
Corr: 0.232
Corr: 0.347
Corr: 0.159
Corr: 0.004
Corr: 0.077
Corr: −0.019
Corr: 0.124
Corr: 0.198
Corr: 0.523
Corr: 0.674
Corr: 0.356
Corr: 0.002
Corr: 0.501
Corr: 0.423
Corr: −0.112
Corr: 0.264
Corr: 0.062
Corr: −0.034

Figure 3-2 Pearson correlations among traits (off-diagonals) and trait distributions (diagonal).
Correlations in red boxes are significant after correcting for multiple comparisons. Life history traits are highlighted in green and floral traits in purple. Only Internode 3 is shown to conserve space. Pollen count represents the ImageJ pollen count. Pollen diameter is not shown.

3.3.3 Linkage map construction and transmission ratio distortion

The final linkage map included 1647 markers in 16 linkage groups (N=15, (Austin 1978)) from 418 individuals. The mean number of markers per linkage group was 103 (range 7-187) and the total map length was 2,843 centimorgans. The mean distance between markers was 1.7 centimorgans (range 1.2-38). Markers were not evenly distributed on the linkage groups. In particular, linkage groups 5, 7, 9, 11, 12, 14 and 15 had large gaps (>15cm). We believe that both this and the large overall map and contig lengths reflect genotyping errors that were not removed by our filtration process. In addition, our SNP markers represented only approximately 400 of the 2,065 scaffolds in the draft genome assembly, so some of the large gaps may reflect missing data. In the future, a binning-based imputation approach (such as FSFHAP, implemented in TASSEL; (Bradbury et al. 2007)) could produce a map with improved coverage that is less affected by genotyping errors. Despite these shortcomings, this map is consistent with our expectations: markers on the same contigs in the draft assembly are placed together in the correct order and the number of linkage groups is approximately correct. It is thus adequate for preliminary QTL mapping.

The supernumerary sixteenth linkage group contained markers from contig 1898 in the draft assembly. To determine the nature of this additional linkage group we performed an NCBI BLAST search for the entire contig to determine whether it contained many regions with homology to the mitochondrial or chloroplast genomes

(Johnson et al. 2008). BLAST hits included a small amount of chloroplastic-like predicted genes but no known mitochondrial or chloroplastic genes. Therefore to determine its approximate position, we reran MSTMap with a less stringently filtered marker set consisting of 3443 markers. Because of missing data, these grouped into few large linkage groups that were sufficient to give a rough idea of which contigs were in linkage disequilibrium with each other. In the resulting map, markers on contig 1898 grouped with markers from two contigs with several markers near the beginning of LG7 (175, 28623) and several contigs that were not represented in the final, most stringently filtered marker set. We suspect therefore that because of the incompleteness of the data, the gaps between LG16 and LG7 were too big to be bridged.

In the final map, 125 markers (7.59%) exhibited significant transmission ratio distortion. These included both isolated markers (on LG2, 5, 6, 8 and 9) and large blocks of distorted markers (on LG4, LG11, LG13, LG15 and LG16). 48 markers, on LG9 and LG11, showed an excess of the *I. lacunosa* allele (genotype ratios AA:AB:BB 34.5:48:17.5). The block on LG9 was extremely small (3 SNPs spanning base pairs in contig 27114) but appears to be correctly assembled, as the three distorted markers are nested among other markers from the same contig that do not show evidence of distortion. In contrast, a large block of 45 markers on LG 11 showed an excess of the *I. lacunosa* allele and a deficit of the *I. cordatotriloba* allele (average genotype ratios 33.4:53.0:13.7). This distorted region included markers from 12 contigs and spanned 37% of the linkage group. 76

markers, on LG2, LG4, LG5, LG6, LG13, LG15 and LG16, showed an excess of the *I. cordatotriloba* allele. These were divided between single markers and small groups on LG2, LG5, LG6, LG8 and LG9, large fractions of LG4 (35%), LG13 (41%), LG15 (32%) and the whole of LG16. Only the distorted region on LG4 had an excess of both *I. cordatotriloba* genotype homozygotes and heterozygotes (13.1:56.4:30.5).

3.3.4 Quantitative trait locus analysis: strength and direction of QTL effects

We identified 21 QTLs affecting 9 of 19 phenotypes measured that were significant at the whole-genome level (Table 2, Supplemental Figures S3.1-S3.19). An additional 30 QTLs, which affected 7 of those phenotypes and 10 additional phenotypes, were significant at the chromosome level. Most traits (6/9 genome-level, 13/17 chromosome-level) were affected by more than one QTL, with style-plus-ovary length and corolla length being the most highly polygenic (4 and 7 or 7 and 8 loci respectively). 95% Bayesian confidence intervals and 1.5-LOD intervals for QTL position both averaged around 60cM (62.65, 58.23). We identified a single locus for flower color on LG6, which explained 100% of the variation.

QTL studies of floral morphology have generally identified multiple QTLs of small-to-moderate effect per trait, and our results were consistent with this pattern. Across all traits, individual QTL relative homozygous effects (the allelic effect relative to the difference between the grandparents) averaged 28.7%. Effect sizes were small (relative homozygous effect <20%) for most floral trait QTLs identified, and floral traits

were also the most likely to be polygenic. Although the effect sizes were individually small, the sum of QTL effects explained the majority of the variation in corolla length (69%), style-plus-ovary length (102%) and shape. Apart from the single-locus trait affecting color, QTL affecting flower shape, early growth and pollen production had the strongest effect sizes relative to the difference between the parents.

Almost all (19/23, 48/52) of the QTL effects we identified were in the same direction as the difference between the species. However, we identified seven antagonistic QTLs with effects in the opposite direction. Three, affecting nectar volume and a measure of shape (corolla length / corolla width) and style-plus-ovary length, were significant at the genome wide level and had a relative homozygous effect of 6%, 39% and 44% respectively. An additional two antagonistic QTLs that were significant at the chromosome level affected the same measure of shape and style-plus-ovary length with relative homozygous effects of 39% and 6%.

Most QTLs (39/52) we identified were partially dominant (dominance deviations between 10% and 90% of the strength of the additive effect). Only QTLs affecting color, pollen, flowering time and whether plants flowered exhibited complete dominance. Across all QTLs exhibiting either complete or partial dominance, the *I. lacunosa* allele was dominant in 25/39 and the *I. cordatotriloba* allele was dominant in 14/39. This pattern of mostly *I. lacunosa* dominance was repeated within floral traits, while dominance was evenly split between *I. cordatotriloba* and *I. lacunosa* alleles for life history QTLs.

Phenotypic means suggested a general pattern of dominance of *I. cordatotriloba* for most floral traits and of *I. lacunosa* for most life history traits. In contrast to this, though, the pattern of dominance among the dominant and partially dominant QTL we identified did not match the phenotypic patterns. The F1 mean also differed significantly from the midparent value in fewer traits than did the F2 mean. This unexpected pattern has been observed in QTL studies of the selfing syndrome in both *Mimulus* and *Leptosiphon* (Fishman et al. 2015; Goodwillie, Ritland, and Ritland 2006). In both cases, there is a pattern of phenotypic dominance favoring the larger-flowered species that is not recapitulated in the QTL effects.

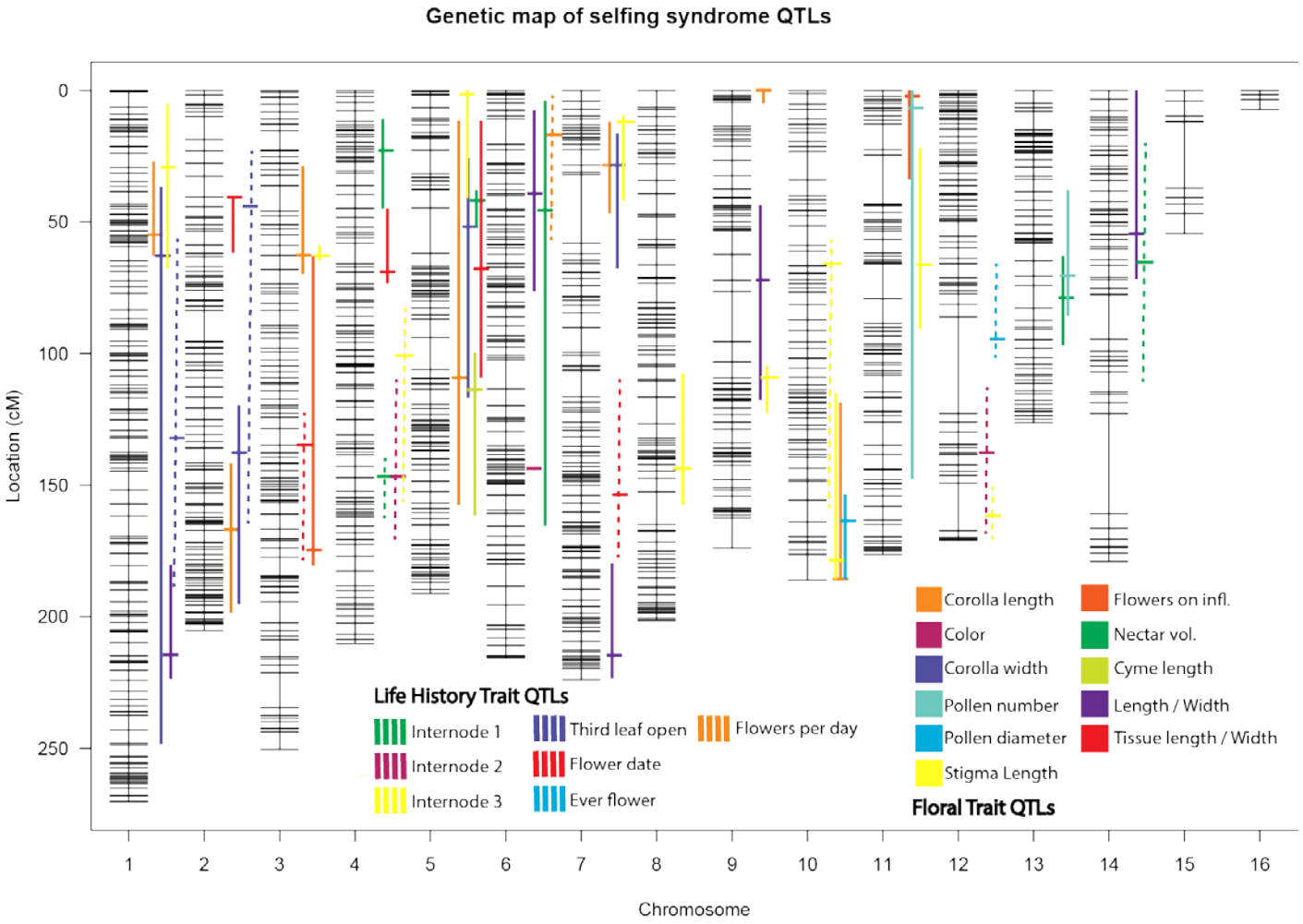


Figure 3-3 Genetic map of selfing syndrome QTLs. Floral QTLs are solid, life history QTLs are striped.

Table 3-2 Quantitative trait loci (QTLs) identified in our analysis.
Columns list trait, linkage group, position on linkage group, log odds ratio of the QTL, Bayesian 95% confidence interval for the location of the QTL, whether it passed a cutoff of genome-wide significance, type (additive, A, dominant, D, partially dominant, PD, over-dominant, OD, or under-dominant, UD), the direction (1 = allelic effects in the same direction as the interspecific difference, 0 = in the opposite direction), the additive effect, the relative homozygous effect (2x the additive effect relative to the difference between the grandparents), the sum of all relative homozygous effects for multi-QTL traits (all, genome-wide significant only), the percent of variation explained, the dominance deviation, the dominant allele if applicable, and the dominance relative to the relative homozygous effect.

Trait	LG	Position	LOD	Bayesian 95% CI	Sig. gen. wide	Dir	Add. Eff.	RHE	Sum RHE	%Var	Type	Dom.	Dom. Sp	Dom. / RHE
Internode 1 day 21	4	147	3.85	140, 163	N	1	1.191	0.273	0.55, 0	3.5	PD	0.858	C	0.720
	14	50	3.38	20, 111	N	1	1.198	0.274		3.0	PD	0.177	C	0.148
Internode 2 day 21	4	147	3.34	110, 171	N	1	3.349	0.720	1.47, 0	4.0	PD	1.778	C	0.531
	12	138	3.77	113, 169	N	1	3.463	0.745		4.3	PD	1.264	C	0.365
Internode 3 day 21	12	162	4.4	151, 171	Y	1	10.278	0.460	0.83, 0.46	4.4	PD	1.648	L	0.160
	11	66	3.33	56.75, 159	N	1	1.555	0.070		3.0	OD	13.291	NA	8.545
	4	101	3.3	83, 157	N	1	6.727	0.301		2.9	UD	10.034	NA	1.491
Third leaf open	1	132.4	3.07	56.52, 189	N	NA	0.292	1.073	0.76, 0	2.9	OD	1.059	NA	3.626
	2	44.3	3.2	23, 165	N	NA	0.469	1.724		3.0	OD	1.021	NA	2.175
Day flower measured	3	135	3.572	122.68, 179	N	NA	15.652	NA		3.9	D	3.913	L	0.250
	7	154	3.553	110, 178	N	NA	13.680	NA		3.8	PD	13.377	L	0.978
Ever flowered	13	94.73	4.19	66, 102	Y	NA	0.085	NA		3.2	D	0.084	C	0.998
Flowers per day	6	17	3.14	1.89, 57	N	NA	0.127			3.8	PD	0.047	L	0.368
Cyme Length	5	114	3.62	100, 162	N	1	2.165	0.283	0.28, 0	4.7	PD	0.595	L	0.275

Flowers on infl.	11	2.25	3.19	0, 34	N	1	0.148	0.920	1.78, 0	5.8	PD	0.060	L	0.406
Color	6	144	69.78	144.0, 144.16	Y	1	0.500	1.000	1	98.6	D	0.479	C	0.958
Cor Length	3	62.7	4.52	29, 70	Y	1	0.578	0.097	0.69, 0.45	4.5	A	0.007	NA	0.013
	1	55	5.05	27,63	Y	1	0.586	0.098		2.8	A	0.041	NA	0.071
	7	28.5	7.24	12, 47	Y	1	0.641	0.107		4.7	PD	0.448	L	0.699
	9	0	12.35	0,5	Y	1	0.872	0.146		9.7	PD	0.298	C	0.342
	5	109.4	3.54	11.57, 158	N	1	0.431	0.072		1.6	PD	0.270	L	0.626
	2	167.2	3.69	142, 199	N	1	0.511	0.085		2.1	PD	0.095	L	0.185
	10	186.1	3.67	119, 186.1	N	1	0.525	0.088		3.0	PD	0.065	L	0.125
Cor. Width	2	138	4.11	120, 195.46	Y	1	0.601	0.100	0.39, 0.3	4.7	PD	0.099	L	0.165
	7	28.5	4.62	16.54, 68	Y	1	0.608	0.101		4.8	PD	0.347	L	0.571
	1	63	4.22	36.61, 249	Y	1	0.621	0.103		4.5	PD	0.102	L	0.164
	5	52	3.87	26, 117	N	1	0.507	0.084		6.2	PD	0.329	L	0.649
	3	175	3.44	63, 181	N	1	0.139	0.862		4.8	PD	0.060	L	0.432
Shape (L/W)	9	72.2	4.85	43.76, 118.08	Y	0	0.032	0.396	1.86, 0.84	4.4	PD	0.019	L	0.581
	1	214.9	5.09	180.57, 224.06	Y	1	0.036	0.445		5.3	PD	0.011	C	0.301
	14	54.6	3.11	7.6, 76.62	N	1	0.022	0.264		2.9	OD	0.031	NA	1.419
	6	39.3	3.4	0, 71.91	N	1	0.030	0.365		3.0	PD	0.002	A	0.079
	7	215.1	3.62	180.11, 223.96	N	0	0.031	0.387		5.1	PD	0.011	L	0.350
Shape (Lvar/L)	5	68	4.56	11.6, 109.54	Y	1	0.010	0.632	0.76, 0.63	5.7	PD	0.006	C	0.651
	4	69.1	3.32		N	NA	0.000	0.013		4.4	UD	0.012	NA	61.866
	2	40.6	3.7	40.6, 62	N	1	0.009	0.110		4.2	PD	0.004	L	0.411
Style-plus-ovary Length	8	144	3.98	108, 158	Y	0	0.381	0.120	1.02, 0.83	4.2	A	0.026	NA	0.068
	1	29.43	5.8	5, 68	Y	1	0.461	0.145		4.8	A	0.024	NA	0.052
	7	12	6.86	9.712, 42	Y	1	0.490	0.154		6.5	PD	0.270	L	0.551
	9	109.29	9.78	105, 123	Y	1	0.602	0.189		8.9	A	0.092	NA	0.153

	3	62.7	13.29	58.99, 65	Y	1	0.701	0.220		13.6	A	0.072	L	0.103
	10	179	3.28	115.36, 186.15	N	1	0.014	0.005		14.6	UD	0.316	NA	21.901
	11	72	3.08	22, 91	N	0	0.204	0.064		15.6	UD	0.283	NA	1.392
	5	1.58	3.61	0, 41	N	1	0.392	0.123		16.6	PD	0.046	L	0.117
Nectar volume	4	23	4.16	11, 45	Y	0	0.042	0.064	0.37, 0.32	0.6	D	0.036	C	0.857
	13	79	3.99	63, 97	Y	1	0.051	0.078		2.9	PD	0.020	C	0.392
	5	42	9.04	38, 52	Y	1	0.098	0.149		9.0	PD	0.055	L	0.563
	6	45.65	3.49	4, 166	N	1	0.053	0.081		2.4	UD	0.070	NA	1.310
PollenIJ	11	6.67	3.08	0, 148	N	1	112.473	0.453	0.45, 0	13.6	UD	168.669	NA	1.500
PollenCC	13	70.6	3.68	38,86	N	1	75.722	0.268	0.27, 0	22.1	A	15.072	NA	0.199
PollenDia	10	164	3.04	154, 186.1489	N	1	1.482	0.252	0.25, 0	18.5	D	2.138	L	1.443

3.3.5 QTL overlap

Although QTLs except one overlapped at least one other QTL, both of our methods of QTL overlap quantification revealed extremely low average overlap across QTLs (Table 3, Table 4) and the results are similar across pairwise overlap and number of traits overlapped. Both phenotypic and genetic data suggest moderate independence within and considerable independence between the selfing syndrome modules (Table 3). Life history traits are the most genetically independent, with average F2 phenotypic correlations below 0.1 for all traits except pollen (which was measured from a small subsample of plants) and average QTL overlap of below 0.15 with all traits. Inflorescence traits are moderately correlated with each other (0.54) and the few inflorescence QTL identified do not overlap. They also exhibit substantial genetic independence from other floral traits, with mean between-module correlations of less than 0.24 and mean QTL overlaps of less than 0.29. Nectar production is more closely related to floral morphology than any other trait, but still exhibits considerable independence (correlation, 0.37; QTL overlap, 0.25). Pollen count generally exhibits high independence from all floral modules, with average between-module correlations of below 0.18 and average QTL overlap of under 0.19.

Between-module trait correlations and QTL overlaps averaged 0.14 and 0.12 respectively. Within-module correlations and QTL overlaps are higher than between-module correlations for all traits where they can be quantified (i.e., where more than one

trait in that module was measured), ranging from 0.35 for life history traits to 0.67 for floral morphology traits (0.08 and 0.28 QTL overlap respectively). Although these correlations are generally higher, there is substantial genetic independence: no within-module value of phenotypic correlation or QTL overlap approaches the value of 1.0 that would imply complete lack of genetic independence.

Table 3-3 Table of phenotypic correlations (top right) and QTL overlaps (bottom left) within and between trait groupings.
Parentheses enclose minimum and maximum values. NA indicates that only a single trait was measured and correlation / overlap could not be calculated. Traits included in phenotypic correlations Floral morphology: corolla length, corolla tissue length, corolla width, style plus ovary length. Nectar: nectar volume. Pollen: pollen count Coulter, pollen count ImageJ, pollen diameter. Inflorescence: cyme length, flowers on inflorescence. Life history: leaf 1 opening, leaf 2 opening, leaf 3 opening, number of leaves day 21, height day 21, length of internodes 1, 2 and 3 on day 21, day flower measured, flowers per day. Traits included in QTL overlap Floral morphology: corolla length, corolla width, shape (length divided by width), shape (length divided by tissue length), style plus ovary length. Nectar: nectar volume, cyme length, flowers on inflorescence. Pollen: pollen count (combined), pollen diameter. Life history: leaf 3 opening, length of internodes 1, 2 and 3 on day 21, day flower measured, flowers per day, ever flowered.

(N traits) (N QTLs)	Floral morphology (4)	Nectar (1)	Inflorescence (2)	Pollen (3)	Life history (10)
Floral morphology (27)	0.665 (0.5, 0.91) 0.284 (0, 0.857)	0.373 (0.26, 0.45)	0.179 (0, 0.39)	0.07 (0.01, 0.16)	0.061 (0.01, 0.24)
Nectar (4)	 0.251 (0, 0.667)	NA NA	0.235 (0.12, 0.35)	0.043 (0.01, 0.1)	0.068 (0.01, 0.22)
Inflorescence (3)	 0.281 (0, 1)	 0 (0, 0)	0.54 (0.54) 0 (0, 0)	0.085 (0.01, 0.21)	0.0545 (0, 0.36)
Pollen (3)	 0.145 (0, 1)	 0.1875 (0, 0.5)	 0.125 (0, 0.5)	0.467 (0.36, 0.72) 0 (0, 0)	0.204 (0.02, 0.46)
Life history (13)	 0.134 (0, 1)	 0.089 (0, 1)	 0.038 (0, 0.5)	 0 (0, 0)	0.350 (0.01, 0.96) 0.083 (0, 1)

Table 3-4 QTL overlap between traits.

Traits: C, Color, CL, Corolla length, CW, Corolla width, D, Flower date, EF, Ever flowered, FI, Flowers on inflorescence, FD, Flowers per day, I1, Internode 1, I2, Internode 2, I3, Internode 3, L, Leaf opening, NV, Nectar volume, CyL, Cyme length, PC, Pollen count, PD, Pollen diameter, LW, Shape (length/width), LV, Shape (tissue length over length), SL, Style plus ovary length. Overlap is scored relative to traits at column heads.

			C	**CL**	**CW**	**D**	**EF**	**FI**	**FD**	**I1**	**I2**	**I3**	**L**	**NV**	**CyL**	**PC**	**PD**	**LW**	**LV**	**SL**
	N traits overlapped	N QTLs	1	7	4	2	1	2	1	2	2	3	2	4	1	1	1	5	3	8
C	1	1	x	0	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	0
CL	8	7	0	x	2/4	0	0	1/2	0	0	0	0	2/2	1/4	1/1	0	1/1	1/5	0	6/8
CW	7	4	0	2/7	x	0	0	0	0	0	0	0	2/2	1/4	1/1	0	0	1/5	1/3	3/8
DG	2	2	0	0	0	x	0	1/2	0	0	0	0	0	0	0	0	0	1/5	0	0
EF	0	1	0	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0
FI	3	2	0	1/7	0	1/2	0	x	0	0	0	0	0	0	0	1/2	0	0	0	2/8
FD	2	1	0	0	0	0	0	0	x	0	0	0	0	1/4	0	0	0	1/5	0	0
I1	4	2	0	0	0	0	0	0	0	x	1/2	1/3	0	0	0	0	0	1/5	1/3	0
I2	2	2	0	0	0	0	0	0	0	1/2	x	2/3	0	0	0	0	0	0	0	0
I3	5	3	0	0	0	0	0	0	0	1/2	2/2	x	0	0	0	1/2	0	1/5	0	1/8
L	5	2	0	2/7	2/4	0	0	0	0	0	0	0	x	0	0	0	0	1/5	1/3	1/8
NV	6	4	1/1	1/7	1/4	0	0	0	1/1	0	0	0	0	x	0	1/2	0	1/5	2/3	0
CyL	3	1	0	1/7	1/4	0	0	0	0	0	0	0	0	0	x	0	0	0	1/3	0
PC	2	1	0	0	0	0	0	1/2	0	0	0	1/3	0	1/4	0	x	0	0	0	1/8
PD	2	1	0	1/7	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0	1/8
LW	9	5	0	1/7	1/4	1/2	0	0	1/1	1/2	0	1/3	1/2	1/4	0	0	0	x	0	1/8
LV	6	3	0	0	1/4	0	0	0	0	1/2	0	0	1/2	2/4	1/1	0	0	0	x	1/8
SL	9	8	0	6/7	3/4	0	0	2/2	0	0	0	1/3	1/2	0	0	1/2	1/1	1/5	1/3	x

3.3.6 Evidence for directional selection in QTL effects

To determine which traits exhibited patterns consistent with a history of selection, we performed a QTL sign test on traits for which we found at least 4 QTLs (corolla length, corolla shape, corolla width, nectar volume, style-plus-ovary length) (Orr 1998; Muir, Pease, and Moyle 2014). Our QTL effect size distributions included both exponential and log-normal distributions (Table S4). However, the QTL sign test implemented by Muir et al. (2014) is not highly sensitive to distributions of effect sizes, and we therefore used the method for exponential distributions. Although the majority of our QTL effects overall are in the same direction as the difference between the species, we could reject the null hypothesis for only one trait (corolla length, P=0.02). The distribution of QTL effect sizes and directions thus provides evidence that corolla size diverged under the influence of selection, but we cannot reject the hypothesis that divergence in other traits resulted from genetic drift.

3.4 Discussion

I. lacunosa represents a clear example of the evolution of the selfing syndrome (Duncan and Rausher 2013a; Sicard and Lenhard 2011). Its high selfing rate is accompanied by the canonical selfing-syndrome traits of smaller, narrower, unpigmented flowers, reduced nectar and reduced pollen:ovule ratio (Duncan & Rausher, 2013; McDonald, Hansen, Mcdill, & Simpson, 2011; Chapter 2). In addition to

these floral display traits, *I. lacunosa*'s early growth is accelerated relative to *I. cordatotriloba* (Chapter 2), which conforms to the observation that selfing plants often have accelerated development across their life history (Snell and Aarssen 2005). All of these traits appear to have diverged in a manner consistent with the action of selection (Duncan & Rausher, 2013; Chapter 2). In this study, we used a QTL mapping approach to test the alternate hypotheses of (1) correlated evolution of genetically linked traits (2) independent selection on independent traits.

We group the components of the selfing syndrome into five modules: floral morphological traits, inflorescence traits, nectar traits, pollen traits, and life history traits. The degree of overlap within and between these modules affects the evolution of the syndrome as a whole. High overlap between and within modules could allow the entire suite of selfing syndrome traits to evolve in response to selection on one or few traits. At the other extreme, high independence within and between modules would require direct selection on and independent genetic changes in multiple traits to assemble the selfing syndrome.

Our results have shown that floral traits, *sensu lato*, exhibit only weak genetic correlations and low QTL overlaps, especially in comparisons of traits between modules. These patterns indicate that substitutions that affect one module most often do not affect traits in the other modules. Because between-species divergence in most of these traits

has been shown to be caused by natural selection, this genetic independence implies that their divergence is not substantially due to indirect selection arising through character correlations. Rather, it implies that divergence of traits in each module was caused by selection acting directly on one or more traits in that module. This conclusion extends to life-history traits as well: divergence in these traits was caused by selection unrelated to any selection imposed on floral traits. The picture that emerges is that the overall suite of traits constituting the selfing syndrome in *I. lacunosa* was assembled piecemeal by several independent selection pressures, perhaps operating at different times since divergence from *I. cordatotriloba*, causing independent evolution in different syndrome modules.

3.4.1 Correlated evolution within modules

Although there is a general genetic independence between modules, within modules there are quite a few examples of QTL overlap between traits. Assuming that QTL overlap indicates that the same locus controls both traits—an assumption that still requires validation—it is very possible that selection on one trait may have caused a correlated response in the other. For example, in Chapter 2 we found that cyme length is significantly diverged between the two species but did not diverge in response to selection. Cyme length is significantly positively correlated with corolla length (0.348, Figure 2), which did diverge because of selection (Chapter 2). The genetic architecture of

cyme length is consistent with correlated evolution with corolla length and/or width: we identified only a single QTL affecting cyme length, which colocalizes with QTLs for corolla length and width and explains 28% of the difference between the grandparents. In this situation, the known selection on corolla size would have caused, through the substitution associated with this QTL, a major change in cyme length.

Another example involves nectar production. Although this trait is largely genetically independent of corolla length, one QTL overlaps with a QTL for corolla length and the *I. lacunosa* allele produces a shorter corolla and reduced nectar volume. While our results indicate that selection acts directly to reduce nectar production, in the case of this QTL that selection may have been supplemented by selection acting directly to reduce corolla length.

We can extend beyond individual traits and use the genetic architecture of the selfing syndrome to evaluate hypotheses about how the selfing syndrome evolved. One possible explanation is that general selection for small size and rapid maturation led to a reduction in floral morphological traits (Snell and Aarsen 2005, Sicard and Lenhard 2011). As we found no correlation between rapid growth and floral traits, we can reject this hypothesis. Selection to increase the efficacy of selfing by reducing herkogamy has also been proposed as a possible explanation for the selfing syndrome as a whole (Sicard and Lenhard 2011). As the grandparents of our mapping population did not differ in

herkogamy, our inferences about its genetic architecture are limited, but a field study of these species found that herkogamy displayed no phenotypic correlation with any other trait, even style length (Duncan and Rausher 2013a). Thus selection for more efficient selfing, also, seems unlikely to be the sole explanation for the evolution of the selfing syndrome in *I. lacunosa*. On the other hand, selection to reduce any aspect of floral size, for example to conserve resources or avoid herbivory (Sicard and Lenhard 2011) could have affected other aspects of flower size and shape: corolla length and width are correlated with each other, with style length, with cyme length, and with nectar volume (but see above). This seems the most likely explanation, although separate changes would also have been required to reduce pollen number, increase early growth, and reduce aspects of nectar production not correlated with flower size.

The low degree of QTL overlap we identified does not support the hypothesis that the selfing syndrome of *I. lacunosa* arose through correlated change. Rather, the genetic architecture of the selfing syndrome indicates that it resulted from many independent changes across the genome. Since the component traits changed in response to selection (Chapter 2), the selfing syndrome likely resulted from multiple, independent selective events.

3.4.2 Genetic architecture of the selfing syndrome

It is a widespread observation that evolutionary and developmental changes in

floral traits are more strongly interconnected with each other than with vegetative traits (Armbruster et al. 2004; Glover et al. 2015; Ashman and Majetic 2006; Smith 2015). Most research has focused exclusively on floral dimensions, but QTL studies have also found QTL overlap between less obviously related attractant and reward traits, such as color and petal morphology (Bradshaw et al. 1998), nectar and petal morphology (Wessinger, Hileman, and Rausher 2014), and others (reviewed in (Smith 2015)). While trait colocalization does not necessarily imply pleiotropy, examples of indisputable multi-trait floral pleiotropy are known from model systems: a single *Brassica* gene has been shown to affect both petal color and volatiles (Zhang et al. 2015). Our results, of traits that are moderately correlated within modules and less correlated between modules, fit broadly into a model of floral integration that predicts overlapping QTLs with varying degrees of overlap (Smith 2015). In this context, *Ipomoea lacunosa* may fall at an extreme in a continuum of possible selfing syndrome genetic architectures. Indeed, comparing its genetic architecture to other examples of the selfing syndrome supports this interpretation.

Across angiosperms, the genetic architecture of the selfing syndrome is extremely variable. Most studies to date have focused primarily on floral traits. Only two other studies have identified QTLs for life history traits (Fishman et al. 2015; Slotte, Hazzouri, and Stern 2012). Focusing solely on floral morphological traits, in *Mimulus*,

the pattern appears to be a moderate to high degree of QTL overlap (0.29, 0.66, 0.50) with QTLs scattered across the genome (Lin and Ritland 1997; Fishman, Kelly, and Willis 2002; Fishman et al. 2015). *Leptosiphon* and *Capsella* demonstrate fairly high overlap (0.54, 0.63), while *Solanum* shows lower overlap (0.23, 0.25) concentrated in few regions (Bernacchi and Tanksley 1997; Georgiady, Whitkus, and Lord 2002; Slotte, Hazzouri, and Stern 2012; Goodwillie, Ritland, and Ritland 2006).

For systems that include traits outside of floral morphology, we compared within- and between-module overlap by assigning traits to categories of modules (floral morphology, inflorescence, life history, pollen, compatibility) and averaging overlap. *Mimulus* and *Capsella* are both exhibit higher within- than between-module correlations. In *Mimulus*, within-floral-morphology overlap was 0.29 in one study, 0.58 in another, and 0.66 in a third that did not include any non-floral morphological traits. For the two studies that included other types of traits, average between-module overlap was 0 and 0.28 respectively (Lin and Ritland 1997; Fishman et al. 2015). In *Capsella*, across multiple trait classes, within-module correlations averaged 0.57 and between-module correlations averaged 0.28 (Slotte, Hazzouri, and Stern 2012). *Solanum* appears to be an exception to this pattern, as both *Solanum* studies identified either similar amounts of overlap or higher between- than within-module correlations (0.41 within, 0.43 between (Bernacchi and Tanksley 1997); 0.58 between, 0.25 within (Georgiady, Whitkus, and Lord 2002)).

These species include a range of genome sizes and structures. However, the taxa at the high and low extremes of QTL overlap, *Mimulus* and *Ipomoea*, are similar in genome size (*M. guttatus*, 430Mb; *I. lacunosa* and *I. cordatotriloba*, 497Mb and 525Mb respectively) and chromosome number (*M. guttatus*, 2N=28, *I. lacunosa* and *I. cordatotriloba*, 2N=30), suggesting that the difference in genetic architecture is not solely due to physical differences in the structure of the genome (Institute 2017; Duncan and Rausher 2013b). That difference also cannot be explained by the number or type of traits included. While we found higher overlap among floral traits than between floral and life history traits, such that the reduced overlap we identified could be due to including additional life history traits, even the overlap we found among floral traits is lower than in some studies that focused exclusively on floral traits (Fishman, Kelly, and Willis 2002; Bernacchi and Tanksley 1997). The genetic architecture of the selfing syndrome varies across systems when only floral morphological traits are considered, when floral traits more broadly are considered, and when floral and life history traits are considered.

Differences in developmental timing have been described across the lifecycle of selfing plants in comparative studies (Snell and Aarssen 2005) and QTLs for time to flowering and developmental stage at flowering have been identified in other studies of the selfing syndrome (Slotte, Hazzouri, and Stern 2012; Fishman et al. 2015). Like other studies of the selfing syndrome that have included life history traits (Slotte, Hazzouri,

and Stern 2012; Fishman et al. 2015), we found stronger correlations among floral traits than between floral and growth traits. Across multiple systems, therefore, the tendency of selfing species to flower earlier than their outcrossing relatives appears to have a separate genetic basis from floral morphological differences, suggesting that the floral and life-history aspects of the selfing syndrome have evolved independently, and we extend that result to include QTLs that affects early growth differences as well.

In this study, most traits were affected by more than one QTL, and the range of QTLs per trait (between 1 and 7) was similar to the majority of other QTL studies of the selfing syndrome (ranging from 1-2 in (Georgiady, Whitkus, and Lord 2002) to 11-15 in (Fishman, Kelly, and Willis 2002). Both the numbers of QTL per trait and the effect sizes of those QTLs are moderate among QTL studies of the selfing syndrome: our QTL are not so numerous and of such small effect as those observed in *Mimulus* by Fishman (2002) nor as few and of such strong effect as those found in *Capsella* by Slotte et al. (2012). The strongest QTLs we detected affected flower shape, early growth and pollen production, while QTLs affecting flower size were generally of small effect individually.

The range of effect sizes and which traits tended to exhibit QTLs of stronger and weaker effect also varies widely across genetic studies of the selfing syndrome. Too few studies have included floral morphological, other floral and life history traits to generalize confidently about patterns of QTL numbers and effect sizes for these different

modules, but there appears to be a trend for floral morphological traits to be more highly polygenic and to have smaller effect sizes than other floral or life history traits (Table S5). In *Capsella, Solanum* and *Leptosiphon*, individual QTLs for floral traits had large effect sizes, while studies in *Mimulus* have found that floral traits are highly polygenic and have very small individual effects (this and following (Slotte, Hazzouri, and Stern 2012; Georgiady, Whitkus, and Lord 2002; Goodwillie, Ritland, and Ritland 2006; Fishman, Kelly, and Willis 2002; Fishman et al. 2015). The pattern is reversed for phenological traits: in *Capsella* QTLs affecting phenology were numerous and had small effect sizes, but in *Mimulus* phenology is governed by few QTLs of large effect.

In the context of other QTL studies of the selfing syndrome, our results indicate that the genetic architecture of the selfing syndrome is highly variable across different systems. Selfing species vary as to which traits are polygenic, which traits are affected by stronger and weaker QTLs, and the extent to which QTLs overlap. However, all demonstrate some correlation among selfing syndrome traits, and all arrive at a similar phenotype despite the diversity of their genetic architecture. It is possible to draw two general conclusions from the diversity of genetic architectures identified in the selfing syndrome. First, although the degree of correlation varies widely, no study has identified a simple genetic basis for the traits involved. Rather, the selfing syndrome is assembled from multiple, genetically distinct components. This multiplicity of genetic

components applies both within and between the floral morphology, other floral, and life history aspects of the selfing syndrome. Second, the phenotypic changes observed are similar despite the variety of genetic architectures that underlie them. This suggests that, like biotic pollination syndromes, the convergent, correlated evolutionary response that constitutes the selfing syndrome is evolving repeatedly in response to similar selective pressures. In *I. lacunosa*, we show both a high level of genetic independence and evidence from Fst-Qst analysis (Chapter 2) that these genetically distinct traits evolved as a result of selection. Together, these studies show that independent selection on multiple traits assembles the selfing syndrome.

3.4.3 Transmission ratio distortion and dominance

We used a reduced-representation next-generation sequencing approach to generate markers, construct a linkage map, and identify QTLs. Our linkage map was broadly consistent with our expectations for number of linkage groups, and although it suffered from long linkage groups and large gaps, it was sufficient for QTL mapping. The linkage map showed significant evidence of transmission ratio distortion in several regions. Transmission ratio distortion is common in interspecific crosses, and may increase with degree of distance between species (Hall and Willis 2005). More of the distorted markers demonstrated a deficit of the *I. lacunosa* homozygous genotype than of the *I. cordatotriloba* homozygous genotype. This pattern has been noted across other

mapping populations between large-flowered outcrossers and small-flowered selfers (Slotte, Hazzouri, and Stern 2012; Fishman et al. 2015; Hall and Willis 2005; Goodwillie, Kalisz, and Eckert 2005). However, it is not universal: in tomato, distortion in favor of the selfing parent is more common (Bernacchi and Tanksley 1997; Georgiady, Whitkus, and Lord 2002). Many factors can contribute to a pattern of transmission ratio distortion: pollen performance, pollen-style interactions, hybrid incompatibilities, inbreeding depression, and cytoplasmic incompatibilities are all possible causes (Slotte, Hazzouri, and Stern 2012; Goodwillie, Ritland, and Ritland 2006; Moyle and Graham 2006; Fishman, Aagaard, and Tuthill 2008). A distorting event involving pollen production—if, for example, fewer pollen bearing the distorted loci are produced in the anther, or pollen bearing the distorted loci produce less effective pollen tubes and fertilize fewer ovules—is plausible given the distribution of pollen count QTLs: one mapped to a highly distorted region on LG11. That the transmission ratio distortion we observed results from pollination rather than an intrinsic inviability or inbreeding depression is also plausible based on patterns of seed germination and development: in general, we observed very high germination rates (>95% of seeds germinated) and generally healthy plants. Therefore if an intrinsic inviability was responsible for the pattern of transmission ratio distortion we observed, it presumably acted early in seed development. A cytoplasmic incompatibility could also have been involved, since we

performed the initial cross only in one direction (with an *I. cordatotriloba* female).

Although the phenotypic distributions we observed were consistent with dominance of *I. cordatotriloba* trait values for most floral traits and of *I. lacunosa* for early growth traits, the QTLs we identified for these traits did not match this pattern of dominance. QTLs identified for corolla length and width were mostly partially dominant toward the *I. lacunosa* allele, while internode QTLs were mostly partially dominant toward the *I. cordatotriloba* allele. This pattern, where the hybrids exceed the midparent value in a way consistent with dominance of the larger-flowered species but QTL effects are not consistent with the phenotypic estimates, has now been observed for floral trait QTL between outcrossing and selfing species in both *Leptosiphon* and Mimulus (Fishman, Kelly, and Willis 2002; Goodwillie, Ritland, and Ritland 2006). In our study it appears to also apply to another size difference - early growth - and to again favor larger trait values. However, we did not identify significant differences between the F1 and the midparent value.

Several possibilities could account for this disparity. In general, F2 individuals that produced more flowers produced larger flowers. The appearance of phenotypic dominance could thus result from bias in which plants ever flowered at all: the distribution of flowers measured was skewed toward large-flowered plants, and the smaller, "missing" F2 flowers would have brought the distribution closer to the

midparent value. Consistent with this, we identified a locus with strong *I. cordatotriloba* dominance for whether plants ever flowered. This seems the most likely explanation, given that the F1 did not also exhibit dominance of the larger-flowered individual, as it did in *Leptosiphon* and *Mimulus*. Other proposed explanations that could also apply here include maternal effects, since we used *I. cordatotriloba* as the female, and loci of small effect that were not detected as QTL. In the former case, if maternally-inherited elements had a strong effect on phenotype, the presence of *I. cordatotriloba* cytoplasm only could bias offspring towards an *I. cordatotriloba* phenotype. Maternal effects are important in plant evolution, particularly in life history traits (Donohue 2009), and it is possible that maternally inherited growth patterns could result in generally larger plants with larger flowers. In the latter, if very numerous loci of small effect with *I. cordatotriloba* dominance were responsible for much of the phenotypic difference, the patterns of QTL dominance may in fact be consistent with the pattern of phenotypic dominance. QTLs of small effect are likely to be overlooked, a phenomenon known as the Beavis effect (Xu 2003). However, both maternal effects and small QTLs would have had a visible effect on the F1 phenotype as well. The underrepresentation of F2s that did not flower is thus in this case the most likely explanation for the disparity between phenotypic dominance patterns and QTL dominance patterns.

3.4.4 QTL sign test and directional selection

The majority of selfing syndrome components in *I. lacunosa* diverged in a manner consistent with selection, according to an Fst-Qst analysis (Chapter 2). However, although the majority of allelic effects we identified are in the same direction as the difference between species, a QTL sign test (Orr 1998; Muir, Pease, and Moyle 2014) found evidence of a history of directional suggestion only on corolla length. In addition, five QTLs affecting three traits exhibited effects in the direction opposite to the difference between the species. Taken alone, the QTL sign test results would suggest that the traits comprising the selfing syndrome in *I. lacunosa* may have diverged through a mix of drift and selection. However, the failure of all QTL effects to align with the interspecific difference does not necessarily mean that selection was not responsible for the differences: the fixation of antagonist allelic effects could also result from initial changes that "overshot" an optimum, followed by compensatory changes in the opposite direction (Orr 1998).

3.4.5 Conclusions and future directions

Future research in the selfing syndrome should aim to disentangle the different evolutionary routes by which highly selfing species arrive at their new morphologies. Two areas, in particular, offer valuable opportunities for further research.

Several selective explanations have been proposed for the selfing syndrome, and they lead to distinct evolutionary predictions. For example, selection for more efficient

selfing is unlikely to include selection for nectar reduction, while selection to evade insect predators might. Now that there is clear evidence that multiple aspects of the selfing syndrome do evolve in response to natural selection and are genetically independent, field tests of these and other predictions are a high priority. Secondly, how the genetic structures that underlie the selfing syndrome evolve may represent a fruitful avenue for inquiry. Genetic architecture, while often treated as a fixed structural element, can itself evolve (Hansen 2013), and one study has found that floral trait integration is reduced in self-compatible species compared to their self-incompatible congeners (Anderson and Busch 2006). Extending research in the genetic consequences of the transition to selfing to the genetic architecture of floral traits thus represents another goal for the field moving forward.

I. lacunosa and *I. cordatotriloba* continue to present an excellent system to explore these questions, as the high degree of genetic independence of selfing syndrome traits in *I. lacunosa* will be valuable for separating traits from their genetic backgrounds to test predictions about selective pressures. These investigations should also extend into other systems to determine what selective forces are responsible for selfing syndrome components, whether large-scale patterns exist in the order of trait changes and nature of selection, which changes tend to result from selection and from drift, and where along a continuum of genetic architecture most selfing species tend to fall.

Conclusions and Future Directions

Self-pollination evolves frequently in flowering plants. (Barrett et al. 2014). Theoretical predictions about the genetic consequences of selfing and observations of morphological trends in selfing species have a long lineage in evolutionary biology (e.g. (Fisher 1941; Ornduff 1969; Stebbins 1957)). In the past decade the availability of genetic tools in non-model systems has led to increased interest in the genomic consequences of self-pollination and the selective and genetic basis of the suite of traits called the "selfing syndrome" (Brandvain et al. 2013; Foxe et al. 2009; Slotte, Hazzouri, and Stern 2012; Qiu et al. 2011; Fishman, Kelly, and Willis 2002; Fishman et al. 2015; Brandvain et al. 2014).

In this dissertation, I used the highly selfing morning glory *Ipomoea lacunosa* and its mixed-mating sister species *I. cordatotriloba* as a model for the transition to self-pollination and the evolution of the selfing syndrome. This system is a rich resource for these questions. *I. lacunosa* is highly selfing displays a typical selfing syndrome, while *I. cordatotriloba* is self-compatible but selfs on average only 0.5 of the time and spans a wide range of selfing rates in its natural populations (Duncan and Rausher 2013). Both species can be grown as annuals in greenhouse conditions, and artificial hybrids can be generated. *I. lacunosa's* small genome has been sequenced, and I have developed genetic resources for the system.

In the first chapter of this dissertation, I report tests of theoretical predictions of

the genomic consequences of self-pollination. In this chapter, we used a SNP panel generated from transcriptome sequencing to investigate genetic diversity, linkage decay, synonymous vs. non-synonymous mutation and codon usage bias. We also used demographic modeling for additional estimates of historical population sizes. We found that the genetic diversity and effective population size of *I. lacunosa* were even more reduced than predicted by theoretical expectations, suggesting that the species had experienced a bottleneck in its evolution. The demographic modeling was consistent with this. *I. lacunosa* displayed increased non-synonymous polymorphism compared to *I. cordatotriloba*, suggesting relaxation of natural selection. However, we found no evidence of decreased codon usage bias, another consequence of relaxed selection.

The population genetics of *I. lacunosa* tell a story of a highly selfing species that went through a bottleneck in its evolutionary history and experienced many consequences of the reduction in population size, including loss of genetic diversity and relaxation of the efficacy of natural selection. However, this chapter leaves open questions about the morphological changes associated with selfing. Although the selfing syndrome is frequently assumed to be the result of selection to reduce resource use, develop faster or avoid herbivory, few studies have tested for evidence of the action of natural selection in the trait changes between selfing and outcrossing species. This is an important distinction to establish: the effects of genetic drift are magnified when the

effective population size is small, as it is in many selfing species and as Chapter One demonstrated it is in *I. lacunosa*. In Chapter Two, therefore, I used a Qst-Fst comparison to test whether the relative patterns of genetic and phenotypic differentiation between I. lacunosa and I. cordatotriloba are consistent with genetic drift or natural selection.

Qst-Fst comparisons are a valuable way to test for the past action of selection in natural populations (Leinonen et al. 2013). By comparing the divergence of phenotypic traits and neutral genetic markers, we can determine if the differentiation in phenotypes is larger than can be explained by neutral processes. We used synonymous SNPs from the genetic dataset in Chapter One in combination with phenotypic measurements from individuals from the same populations in a shared greenhouse environment. In addition to providing phenotypic data for the Qst-Fst comparison, this allowed us to more precisely describe phenotypic differences between the two species across their ranges. Our Qst-Fst comparison was consistent with a role for natural selection in early growth traits, nectar traits, and floral traits in the selfing syndrome of *I. lacunosa*.

Chapter Three demonstrates that in addition to having evolved because of natural selection, the differences between *I. lacunosa* and *I. cordatotriloba* are the result of multiple independent changes. Across different categories of traits—floral morphology, inflorescence structure, nectar, pollen, and life history—we found low levels of correlation and of QTL overlap. Even within traits, the structure of trait correlation

allowed considerable latitude for traits to evolve independently. This suggests that at least in *I. lacunosa,* explanations for the selfing syndrome that group all the changes as correlated evolution driven by one selective pressure do not fit the genetic architecture observed. We then review the literature of other QTL studies of the selfing syndrome and find a similar pattern of higher but not complete within-module phenotypic correlation and QTL overlap and lower between-module correlation and overlap.

This dissertation raises many questions for future research both into the selfing syndrome generally and into the biology of these species. Our result in Chapter One of a larger-than-expected bottleneck is consistent with some studies of highly selfing plants, but not others. In *Mimulus*, *Capsella*, *Collinsia*, and *Clarkia* (Guo et al. 2009; Brandvain et al. 2013; Brandvain et al. 2014; Hazzouri et al. 2013; Pettengill and Moeller 2012), the larger-bottleneck pattern appears. On the other hand, other taxa, such as *Eichhornia paniculata* (Ness, Wright, and Barrett 2010); *Leavenworthia alabamica* (Busch, Joly, and Schoen 2011); and *Arabidopsis lyrata* (Foxe et al. 2010), do not exhibit this evidence of population reduction. These species do not follow an obvious pattern of, for example, the correspondence between the loss of self-incompatibility and the adoption of a highly selfing lifestyle: in *Capsella*, *Leavenworthia* and *Arabidopsis*, the transition to selfing appears to have coincided with the loss of self-incompatibility, whereas *Mimulus*, *Clarkia*, *Collinsia*, *Eichornia* and *Ipomoea* it did not. Exploring the roles of divergence time, ongoing interbreeding between selfing and non-selfing populations, and other

factors determining the effective population sizes of highly selfing species should be an ongoing goal.

Various traits have been included in the selfing syndrome. At its most expansive it includes floral morphology traits like petal size, attractant traits like nectar, reproductive allocation traits including pollen:ovule ratio and number of flowers, floral-associated vegetative traits like sepal and pedicel size and inflorescence structure, growth and life history traits, and even range size (Ornduff 1969; Snell and Aarssen 2005; Sicard and Lenhard 2011). However, to our knowledge *I. lacunosa* is the only system in which any of these traits have been subject to direct tests for selection. Determining whether general patterns exist in which traits did and did not diverge because of selection is obviously an important future aim. Qst-Fst comparisons, field tests of selection and molecular signatures of natural selection will all be valuable in this endeavor. Relatedly, research should aim to determine in which order these changes typically occur, whether certain trait changes are valuable only in the context of other previous trait changes, and if these patterns are consistent across different taxa.

We have shown that the genetic architecture of the selfing syndrome is not consistent with correlated evolution in response to a single selective pressure, and that genetic architecture varies across systems. Therefore a clear next step is expanding the study of comparative genetic architecture of selfing syndrome traits beyond floral

morphology. In addition, as genetic architecture can itself evolve and may be differentiated between selfing and outcrossing species (Hansen 2013; Anderson and Busch 2006), investigating how the genetic architecture of floral traits evolves in response to different selective regimes is another valuable area of research.

Appendix A: Chapter 1 supplemental materials

Table S1.1. Samples and collection sites. Under "Species," "C" indicates *I. cordatotriloba* and "L" indicates *I. lacunosa*.

Sympatric (dark grey)

Population (Location)	Source	Latitude	Longitude	Species	Individuals
CCR	Rausher Lab	33.48225	-79.26995	C	PC_10_CCR_1
					PC_11_CCR_2
					PC_12_CCR_5
				L	L_13_CCR_1
					L_14_CCR_2
					L_15_CCR_3
Site1	Rausher Lab	33.95844	-78.99154	C	PC_19_Site1_2
					PC_20_Site1_3
				L	L_16_Site1_1
					L_17_Site1_2
					L_18_Site1_4
					L_21_Site1_4
CHAD	Rausher Lab	34.31021	-78.82640	C	PC_25_CHAD_1
					PC_26_CHAD_2
					PC_27_CHAD_3
				L	L_22_CHAD_2
					L_23_CHAD_3
					L_24_CHAD_4
POL	Rausher Lab	34.94662	-77.24104	C	PC_67_POL_3
					PC_68_POL_5
					PC_69_POL_6
				L	L_64_POL_1
					L_65_POL_3
					L_66_POL_4

Allopatric - I. cordatotriloba (purple)

Population (Location)	Source	Latitude	Longitude	Species	Individuals
SOS	Rausher Lab	34.57249	-77.94510	C	PC_28_SOS_1
					PC_29_SOS_2
					PC_30_SOS_3
Ocean	Rausher Lab	34.48462	-77.56372		PC_31_Ocean_1
					PC_32_Ocean_2
					PC_33_Ocean_3
Site5	Rausher Lab	34.37725	-77.89655		PC_34_Site5_1
					PC_35_Site5_4
					PC_36_Site5_7
PI_518494_01_SD	USDA	18.20000	-93.08306		PC_44_PI_518494_01_SD_MX
					PC_71_PI_518494_01_SD_1
PI_518495_02_SD	USDA	18.06667	-92.91667		PC_51_PI_518495_02_SD_MX
					PC_72_PI_518495_02_SD_2
PI_675061	USDA	30.18305	-97.87615		PC_73_PI_675061_1

Allopatric - I. lacunosa (red)

Population (Location)	Source	Latitude	Longitude	Species	Individuals
LTS	Rausher Lab	35.33599	-79.35348	L	L_1_LTS_3
					L_2_LTS_5
					L_3_LTS_6
CCF	Rausher Lab	34.89755	-79.80206		L_4_CCF_1
					L_5_CCF_2
					L_6_CCF_4
PCB	Rausher Lab	34.22187	-80.40211		L_7_PCB_1
					L_8_PCB_2
					L_9_PCB_6
Spindletop_Farm	Baskin Lab	38.01565	-84.50518		L_50_Spindletop_Farm _KY
Kent	Rausher Lab	37.19345	-80.57372		L_53_Kent_6
					L_54_Kent_7
					L_55_Kent_8
KS	Rausher Lab	38.88445	-95.32063		L_56_KS_4

				L_57_KS_5
				L_58_KS_8

Unknown - I. cordatotriloba (light purple)

Population (Location)	Source	General Location	Species	Individuals
PI_645624_01_SD	USDA	LSU Campus, Baton Rouge, LA	C	PC_37_PI_645624_01_SD_LA
PI_645625_01_SD	USDA	LSU Campus, Baton Rouge, LA		WC_43_PI_645625_01_SD_LA
GRIF_6183_01_SD	USDA	TX		PC_46_GRIF_6183_01_SD_TX
PI_645627_02_SD (Tift County, GA)	USDA	Tift County, GA		PC_48_PI_645627_02_SD_GA
Grif_15931_01_SD	USDA	TX		PC_52_Grif_15931_01_SD_TX

Unknown - I. lacunosa (light red)

Population (Location)	Source	General Location	Species	Individuals
PI_645623_01_SD	USDA	Texas	L	L_38_PI_645623_01_SD_TX
PI_645621_02_SD	USDA	Tift County, GA		L_39_PI_645621_02_SD_GA

Table S1.2. Codon usage in *I. lacunosa.* Numbers are number of codons used in annotated coding sequence features of the *I. lacunosa* genome.

TTT PHE 325,346
TTC PHE 328,239

TTA LEU 117,788
TTG LEU 295,329
CTT LEU 340,479
CTC LEU 279,185
CTA LEU 102,427
CTG LEU 220,913

TCT SER 274,064
TCC SER 227,469
TCA SER 204,775
TCG SER 109,560
AGT SER 193,973
AGC SER 219,881

TAT TYR 191,741
TAC TYR 139,111

TAA STP 82,514
TAG STP 72,870
TGA STP 131,898

TGT CYS 167,849
TGC CYS 186,417
TGG TRP 194,668

CCT PRO 216,285
CCC PRO 153,892
CCA PRO 229,386
CCG PRO 107,754

CAT HIS 250,790
CAC HIS 181,606

CAA GLN 289,499
CAG GLN 225,144

CGT ARG 85,708
CGC ARG 93,552
CGA ARG 102,090
CGG ARG 101,099

ATT ILE 307,879
ATC ILE 279,055
ATA ILE 186,601

ATG MET 261,469

ACT THR 202,906
ACC THR 174,995
ACA THR 187,644
ACG THR 83,692

AAT ASN 317,357
AAC ASN 259,089

AAA LYS 330,375
AAG LYS 343,609

AGA ARG 240,522
AGG ARG 191,781

GTT VAL 273,624
GTC VAL 171,199
GTA VAL 138,664
GTG VAL 203,761

GCT ALA 259,038
GCC ALA 189,663
GCA ALA 228,468
GCG ALA 100,202

GAT ASP 333,143
GAC ASP 188,747

GAA GLU 360,814
GAG GLU 314,881

GGT GLY 188,584
GGC GLY 181,920
GGA GLY 228,443
GGG GLY 161,073

Table S1.3. Proportion of change in efficacy of selection between *I. cordatotriloba* and *I. lacunosa* due solely to switch from mixed mating to selfing. *s*: selection coefficient acting against deleterious allele.

	Mutation rate (μ)			
s	5×10^{-8}	1.5×10^{-8}	7.5×10^{-9}	2×10^{-9}
0.000001	0.394	0.393	0.392	0.384
0.000005	0.392	0.388	0.381	0.336
0.00001	0.390	0.381	0.366	0.262
0.00005	0.374	0.313	0.210	0.0037
0.0001	0.350	0.210	0.061	0.0
0.0005	0.120	0.0003	0.0	0.0
0.001	0.012	0.0	0.0	0.0
0.005	0.0	0.0	0.0	0.0
0.01	0.0	0.0	0.0	0.0

S1.4 The following script was used to filter called SNPs.

Variant filtration script

```
#!/bin/bash
#
#SBATCH --mem=50000
#SBATCH --job-name=Hard_filter_joint_genotyped_file_QD_DP
#SBATCH -p rausherlab
#SBATCH --mail-user=jlr42
#SBATCH --mail-type=ALL
export PATH=/opt/apps/slurm/java/jdk1.8.0_60/jre/bin:$PATH
java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
    -T VariantFiltration \
    -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
    -V
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/All_transc
riptomes_jointgenotyped_12-21_raw_snps.vcf \
```

```
    -window 35 -cluster 3 -filterName FS -filter "FS > 30.0" -filterName QD -filter "QD
< 2.0" \
     --genotypeFilterExpression "DP < 10" \
    --genotypeFilterName "lowDP" \
   -o
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_raw_snpstempDP1.vcf


java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
    -T VariantFiltration \
    -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
    -V
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_raw_snpstempDP1.vcf \
    --setFilteredGtToNocall \
    -o
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_raw_snpstempDP2.vcf
```

```
java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
    -T VariantFiltration \
    -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
    -V /work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenotyped_1-13_raw_snpstempDP2.vcf \
     --filterExpression "AN < 120" \
    --filterName "lowAN" \
     -o /work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenotyped_1-13_raw_snpstempDP3.vcf
java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
    -T SelectVariants \
    -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
    -V /work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenotyped_1-13_raw_snpstempDP3.vcf  \
```

```
    -select 'vc.isBiallelic() ? AF > 0.0 : vc.hasGenotypes()  &&
vc.getCalledChrCount(vc.getAltAlleleWithHighestAlleleCount())/(1.0*vc.getCalledChrC
ount()) > 0.0 ' -o
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_raw_snpstempDP4.vcf   -env    -ef
java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
        -T SelectVariants \
        -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
        -V
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_raw_snpstempDP4.vcf \
    -select 'vc.isBiallelic() ? AF < 1.0 : vc.hasGenotypes() &&
vc.getCalledChrCount(vc.getAltAlleleWithHighestAlleleCount())/(1.0*vc.getCalledChrC
ount()) < 1.0' -o
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_QD_DP10_AF_NCalled60.vcf    -env    -ef
java -jar /dscrhome/jlr42/GenomeAnalysisTK.jar \
        -T VariantsToTable \
        -R /work/rausher/lacunosa_ref/lacunosa_genome_070416.fasta \
```

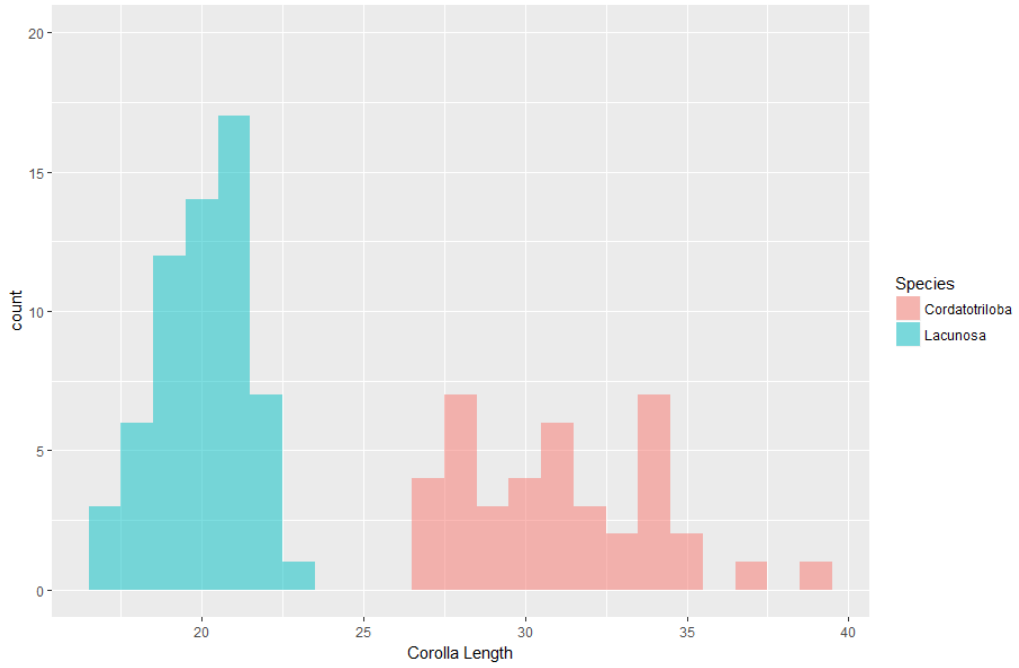
```
    -V
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/VCFs/Filtered/Jointgenot
yped_1-13_QD_DP10_AF_NCalled60.vcf \
    -F CHROM -F POS -F TYPE -F MULTI-ALLELIC -F REF -F ALT \
    -F NCALLED -F AC -F AF -F HOM-REF -F HOM-VAR -F HET \
    -F QUAL -F MQ  \
    -GF GT \
    --allowMissingData \
    -o
/work/rausher/Transcriptome_data_May_2016/GATK_pipeline/SNPTables/Jointgenotyp
ed_1-13_QD_DP10_AF_NCalled60.table
```

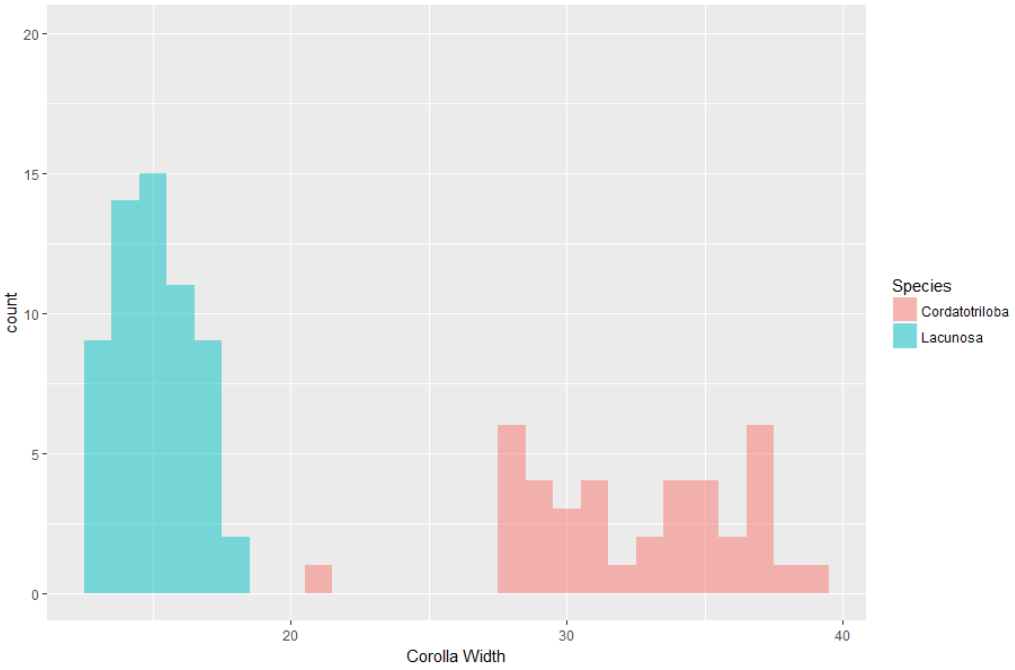
Appendix B: Chapter 2 supplemental materials

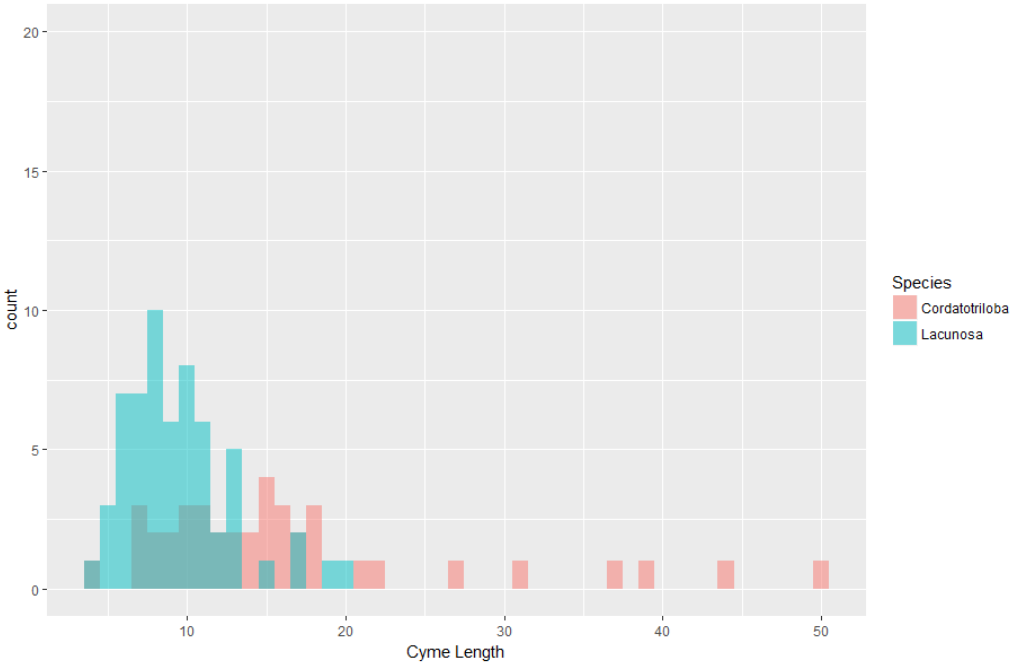
Table S2.1 Restricted estimation maximum likelihood effects for the site and female parent random effects from the nested ANOVA

Trait	Nested ANOVA, site effect. Variance ratio, variance component estimate (95% CI), Wald P value	Nested ANOVA, accession effect. Variance component estimate (95% CI), Wald P value
Internode 1	0.0691, 0.321 (-2.076, 2.719), P=0.7928	1.1195, 5.202 (1.411, 8.994), P=0.0072
Internode 2	0.1569, 3.264 (-3.910, 10.439), P=0.3725	0.3604, 7.497 (-2.159, 17.154), P=0.1281
Internode 3	-0.0397, -6.287 (-79.171, 66.596), P=0.8657	1.1945, 189.358 (58.175, 320.541), P=0.0047
Flowers per day	0.3594, 0.745 (-0.282, 1.771), P=0.1551	-0.0098, -0.020 (-0.731, 0.690), P=0.955
Flowers on inflorescence	0.2152, 0.160 (-0.087, 0.407), P=0.2050	0.1572, 0.117 (-0.189, 0.422), P=0.454
Cyme length	2.6362, 35.039 (7.373, 62.704), P=0.0131	0.7253, 9.640 (0.691, 18.590), P=0.0348
Corolla length	1.5244, 2.575 (0.266, 4.884), P=0.0288	0.5766, 0.974 (-0.028, 1.976), P=0.567
Corolla width	0.4013, 1.141 (-1.228, 3.510), P=0.3451	1.1471, 3.262 (0.081, 5.715), P=0.0092
Nectar volume	1.0909, 0.162 (-0.036, 0.359). P=0.1084	0.7296, 0.108 (-0.010, 0.226), P=0.0729
Nectar sugar concentration	-0.0134, -45.118 (-713.254, 623.019), P=0.8947	0.1848, 622.599 (-721.087, 1966.285), P=0.3637
Leaf length/width	-0.0099, 0.055 (-0.780, 0.671), P=0.8829	0.0147, 0.0865 (-1.613, 1.776), P=0.9249
Leaf dissection	0.8869, 0.025 (0.005, 0.045), P=0.0145	-0.0667, -0.002 (-0.011, 0.007), P=0.6772
Sepal length	0.5227, 0.330 (0.047, 0.612), P=0.0223	-0.2152, 0.136 (-0.305, 0.034), P=0.1160

Figure S2.1. Distributions of corolla length, corolla width and cyme length in *I. cordatotriloba* and *I. lacunosa*.




20
15
10
5
0
count
20
30
40
Corolla Width
Species
Cordatotriloba
Lacunosa


20
15
10
5
0
count
10
20
30
40
50
Cyme Length
Species
Cordatotriloba
Lacunosa

Appendix C: Chapter 3 supplemental materials

Table S3.1. Traits differed between plants that produced many and few flowers. Depending on whether variances were equal (determined by a Bartlett test) we performed either Welch T-tests or T-tests (Test: W=Welch, T=T-test). Traits in bold differed significantly between low- and high-flowering plants after performing a sequential Bonferroni correction for multiple comparisons.

Trait	Under 3 Mean	Over 3 Mean	T	P	Test
Day Measured	182.9773	147.8181	5.1459	8.99E-07	W
Cyme length	20.04755	24.31953	-5.4501	9.44E-08	T
Flowers per day	1.1	1.377983	-6.6636	1.2E-10	W
Flowers per inflorescence	1.286364	1.48034	-3.3062	0.001156	W
Corolla length	26.26545	26.82025	-2.6672	0.00844	W
Corolla tissue length	30.14818	30.98614	-3.7697	0.000224	W
Corolla width	20.70455	21.60243	-4.3977	1.45E-05	T
Corolla length / width	1.275313	1.247099	2.3199	0.02148	W
Corolla tissue length / corolla length	1.148887	1.155566	-1.8272	0.06939	W
Stigma length	15.59955	16.0456	-3.2458	0.001283	T
Nectar volume in uL	0.120009	0.247858	-5.6011	5.38E-08	W
Leaf 1 open	11.23853	11.7572	-1.5662	0.1182	T
Leaf 2 open	13.23853	13.83817	-1.6757	0.09469	T
Leaf 3 open	15.66038	16.11588	-1.3509	0.178	W
Height 21 days	474.7523	427.7692	1.6981	0.09077	W
Internode 1 at 21 days	12.45872	11.50607	2.0845	0.03813	W
Internode 2 at 21 days	18.93578	15.84615	1.8036	0.07335	W
Internode 3 21 days	70.36697	61.74494	2.2313	0.02628	T
Number of leaves 21 days	6.311927	5.821862	1.1941	0.2333	T

Table S3.2. F1 and grandparent early growth measurements. Data from 13 F1s from 4 seed families, all produced from *I. cordatotriloba* ovules and *I. lacunosa* pollen, and 58 each *I. lacunosa* and *I. cordatotriloba* derived from the grandparents of the mapping population. The F1 values were compared to the midparent average with a one-sample T-test. For all traits except number of leaves at 23 days, the F1 differed significantly from the mid-parent value. Faster growth of F1 hybrids could reflect *I. lacunosa* dominance, heterosis, or both.

	F1 Avg.	**Cdt. Avg.**	**Lac. Avg**	**Mid.Avg.**	**T**	**P**
Leaf 1 open	12	12.86207	13.53571	13.19889	21.67993	5.43E-11
Leaf 2 open	14.30769	15.08621	15.94643	15.51632	22.1109	4.31E-11
Leaf 3 open	17	18.10345	18.64286	18.37315	21.49115	6.01E-11
Height 21 days	561.9231	334.8966	492.9138	413.9052	7.758091	5.14E-06
Internode 1 at 21 days	10.15385	6.327586	13.03448	9.681034	5.394563	0.000161
Internode 2 at 21 days	17.84615	6.862069	22.36207	14.61207	3.726109	0.002894
Internode 3 at 21 days	78.30769	26.39655	79.48276	52.93966	6.847195	1.78E-05
Number of leaves at 23 days	9.923077	8.586207	7.362069	7.974138	0.762366	0.460565

Table S3.3. Pearson correlation coefficients for all traits.

L1, Leaf 1 opening, L2, Leaf 2 opening, L3, Leaf 3 opening, H, Height day 21, I1, Internode 1, I2, Internode 2, I3, Internode 3, NL, Number of leaves day 21, D, Flower date, FD, Flowers per day, CyL, Cyme length, FI, Flowers on inflorescence, C, Color, CL, Corolla length, CTL, Corolla tissue length, CW, Corolla width, SL, Style plus ovary length, NV, Nectar volume, PCC, Pollen count Coulter, PCI, Pollen count ImageJ, PD, Pollen diameter.

	L1	L2	L3	H	I1	I2	I3	NL	D	FD	CyL	FI	C	CL	CTL	CW	SL	NV	PCC	PCI	PD
L1	1	0.96	0.92	-0.77	0.05	-0.05	-0.39	-0.83	0.27	0.24	0	-0.01	-0.04	-0.06	-0.04	0.05	-0.12	-0.02	-0.28	-0.09	-0.46
L2	0.96	1	0.94	-0.75	0.1	-0.04	-0.37	-0.84	0.26	0.24	0.02	-0.02	-0.02	-0.04	-0.02	0.07	-0.13	0.01	-0.24	-0.06	-0.42
L3	0.92	0.94	1	-0.75	0.16	0.05	-0.21	-0.85	0.27	0.29	-0.02	-0.03	-0.02	-0.05	-0.04	0.05	-0.09	-0.03	-0.22	-0.07	-0.36
H	-0.77	-0.75	-0.75	1	0.14	0.16	0.52	0.77	-0.17	-0.22	0.01	-0.02	-0.01	0.04	0.04	-0.05	0.06	0.01	0.22	0.04	0.42
I1	0.05	0.1	0.16	0.14	1	0.59	0.58	-0.07	0.19	-0.06	-0.03	-0.06	-0.01	-0.02	-0.03	-0.03	-0.05	-0.12	0.14	0.05	0.16
I2	-0.05	-0.04	0.05	0.16	0.59	1	0.62	0.01	0.06	-0.09	-0.05	0.02	0.05	0.06	0.03	0.01	0.06	-0.09	0.09	-0.02	0.12
I3	-0.39	-0.37	-0.21	0.52	0.58	0.62	1	0.25	-0.01	-0.13	0	-0.03	0.06	-0.01	-0.02	-0.03	0.01	-0.06	0.32	0.02	0.31
NL	-0.83	-0.84	-0.85	0.77	-0.07	0.01	0.25	1	-0.24	-0.24	-0.05	-0.02	-0.01	0.03	0.04	-0.02	0.1	0.01	0.21	0.1	0.4
D	0.27	0.26	0.27	-0.17	0.19	0.06	-0.01	-0.24	1	0.04	-0.36	-0.21	-0.12	-0.16	-0.21	-0.14	-0.24	-0.22	-0.21	-0.11	-0.3
FD	0.24	0.24	0.29	-0.22	-0.06	-0.09	-0.13	-0.24	0.04	1	0.04	0.09	0	-0.04	-0.03	0.08	-0.04	0.11	-0.28	-0.21	-0.18
CyL	0	0.02	-0.02	0.01	-0.03	-0.05	0	-0.05	-0.36	0.04	1	0.54	-0.11	0.35	0.39	0.34	0.23	0.35	-0.01	0.21	0.03
FI	-0.01	-0.02	-0.03	-0.02	-0.06	0.02	-0.03	-0.02	-0.21	0.09	0.54	1	-0.01	0	-0.02	0.08	-0.02	0.12	0.03	0.16	-0.07
C	-0.04	-0.02	-0.02	-0.01	-0.01	0.05	0.06	-0.01	-0.12	0	-0.11	-0.01	1	0	-0.01	0.02	-0.01	0.14	-0.05	-0.04	-0.22
CL	-0.06	-0.04	-0.05	0.04	-0.02	0.06	-0.01	0.03	-0.16	-0.04	0.35	0	0	1	0.91	0.52	0.67	0.36	-0.13	0.01	-0.1
CTL	-0.04	-0.02	-0.04	0.04	-0.03	0.03	-0.02	0.04	-0.21	-0.03	0.39	-0.02	-0.01	0.91	1	0.72	0.67	0.45	-0.05	-0.01	-0.01
CW	0.05	0.07	0.05	-0.05	-0.03	0.01	-0.03	-0.02	-0.14	0.08	0.34	0.08	0.02	0.52	0.72	1	0.5	0.42	-0.16	-0.03	-0.09
SL	-0.12	-0.13	-0.09	0.06	-0.05	0.06	0.01	0.1	-0.24	-0.04	0.23	-0.02	-0.01	0.67	0.67	0.5	1	0.26	0.06	0.1	0.11
NV	-0.02	0.01	-0.03	0.01	-0.12	-0.09	-0.06	0.01	-0.22	0.11	0.35	0.12	0.14	0.36	0.45	0.42	0.26	1	-0.01	0.02	0.1
PCC	-0.28	-0.24	-0.22	0.22	0.14	0.09	0.32	0.21	-0.21	-0.28	-0.01	0.03	-0.05	-0.13	-0.05	-0.16	0.06	-0.01	1	0.36	0.72
PCI	-0.09	-0.06	-0.07	0.04	0.05	-0.02	0.02	0.1	-0.11	-0.21	0.21	0.16	-0.04	0.01	-0.01	-0.03	0.1	0.02	0.36	1	0.32
PD	-0.46	-0.42	-0.36	0.42	0.16	0.12	0.31	0.4	-0.3	-0.18	0.03	-0.07	-0.22	-0.1	-0.01	-0.09	0.11	0.1	0.72	0.32	1

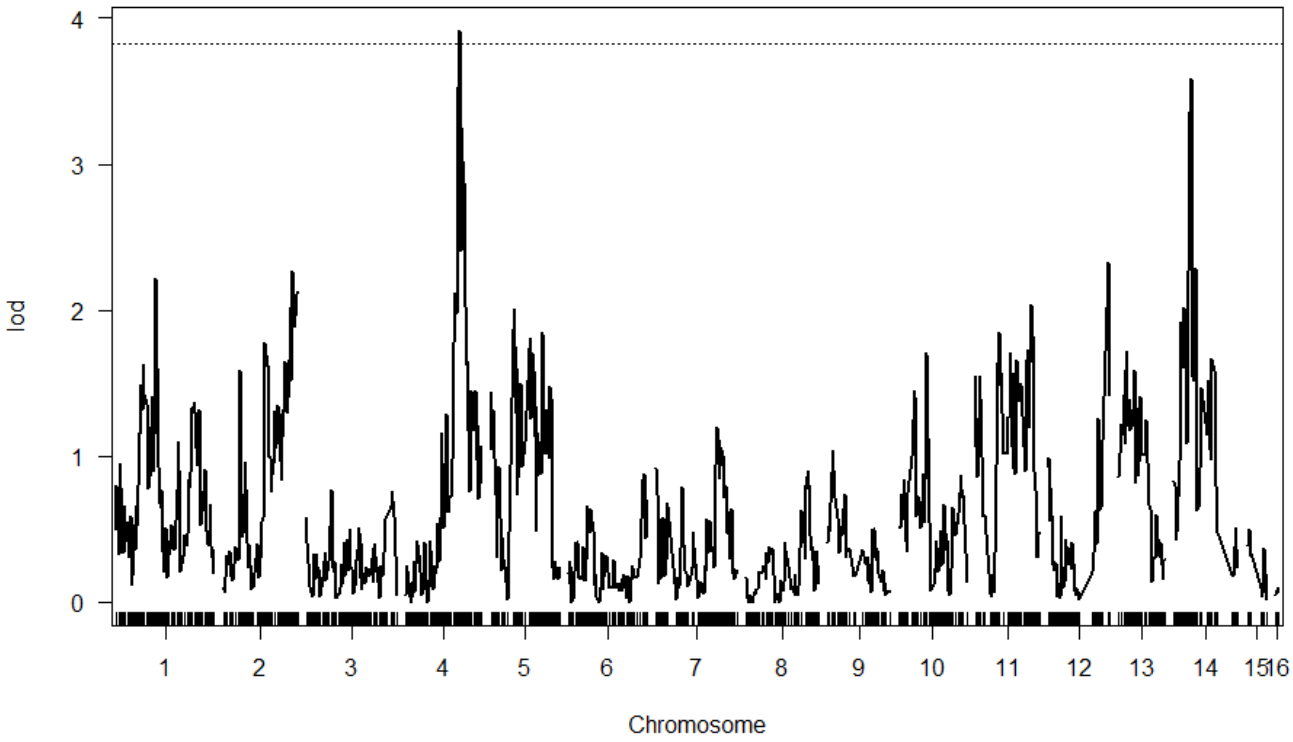


Figure S3.1

LOD plot of Internode 1 on day 21. Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG14 is significant at chromosome-wide level.

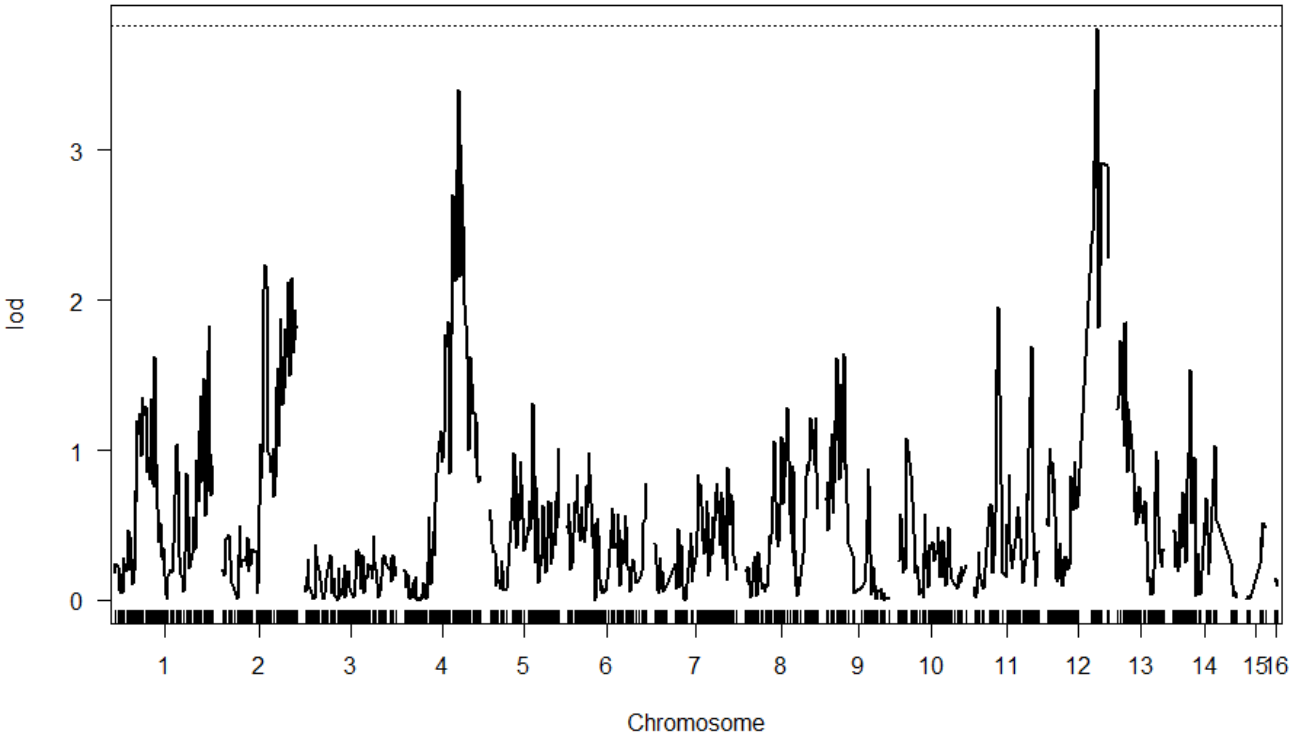


Figure S3.2

LOD plot of Internode 2 on day 21. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG4 and LG12 significant at chromosome-wide level.

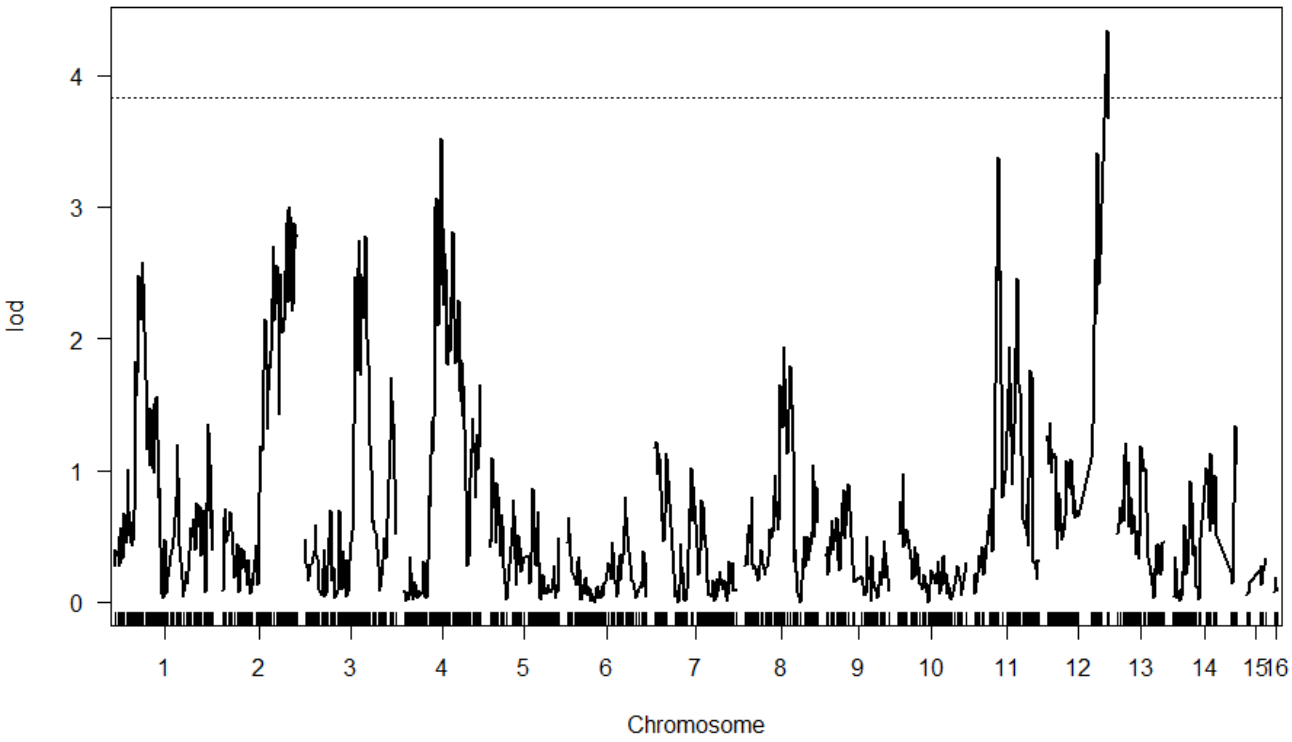


Figure S3.3

LOD plot of Internode 3 on day 21. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG 4 and 11 significant at chromosome-wide level.

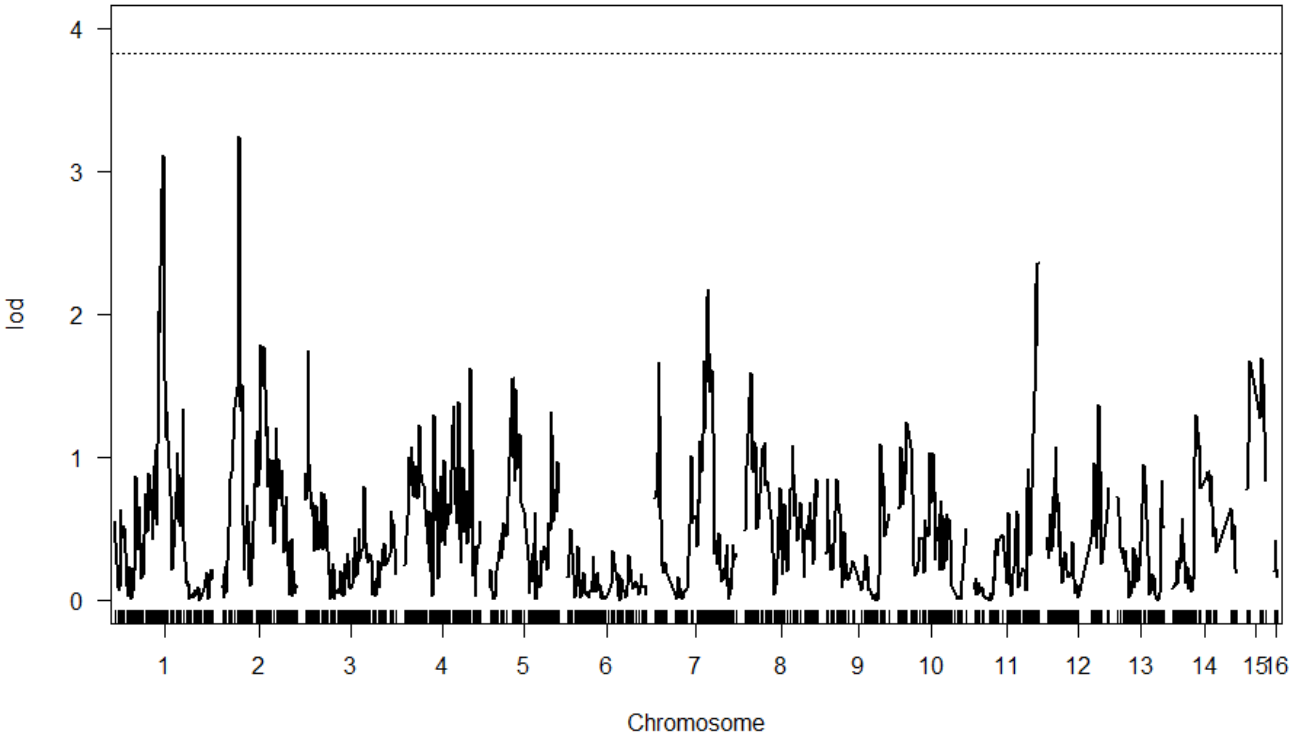


Figure S3.4

LOD plot of opening of leaf 3. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG1 and LG2 significant at chromosome-wide level.

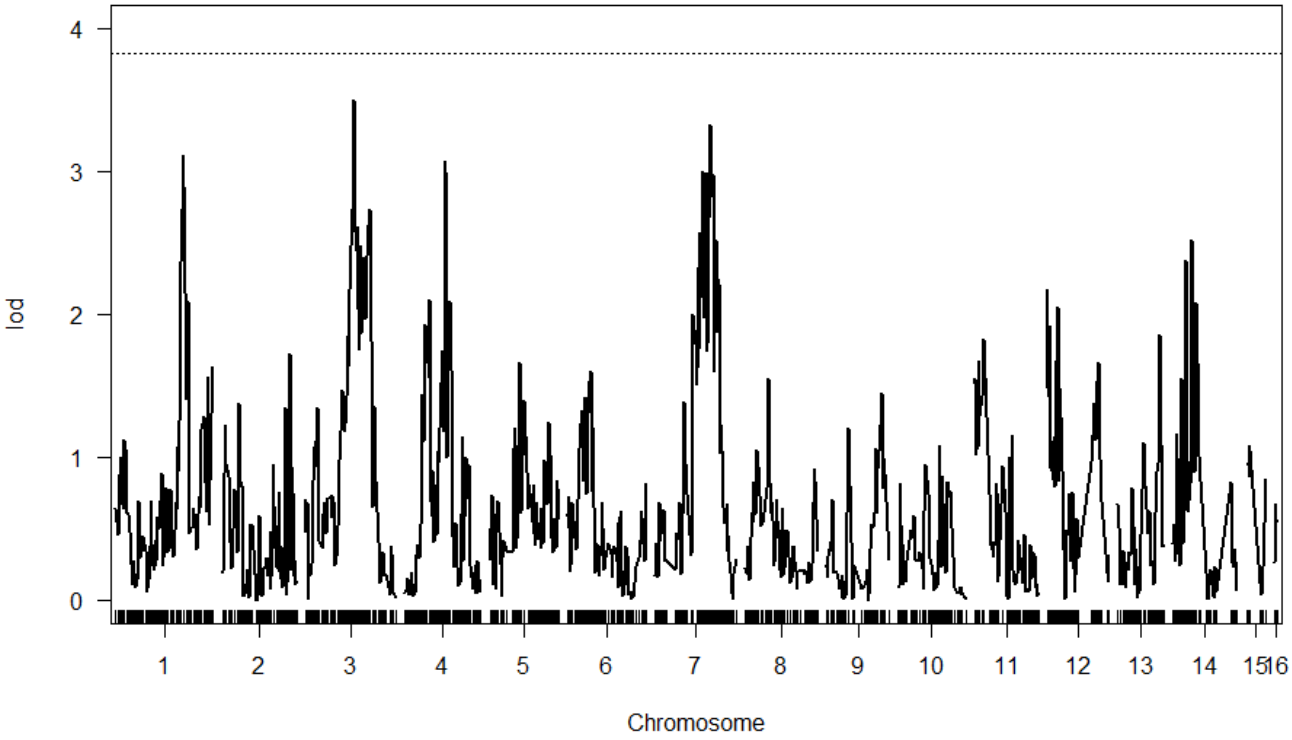


Figure S3.5

LOD plot of flower measurement date. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG3 and LG7 significant at chromosome-wide level.

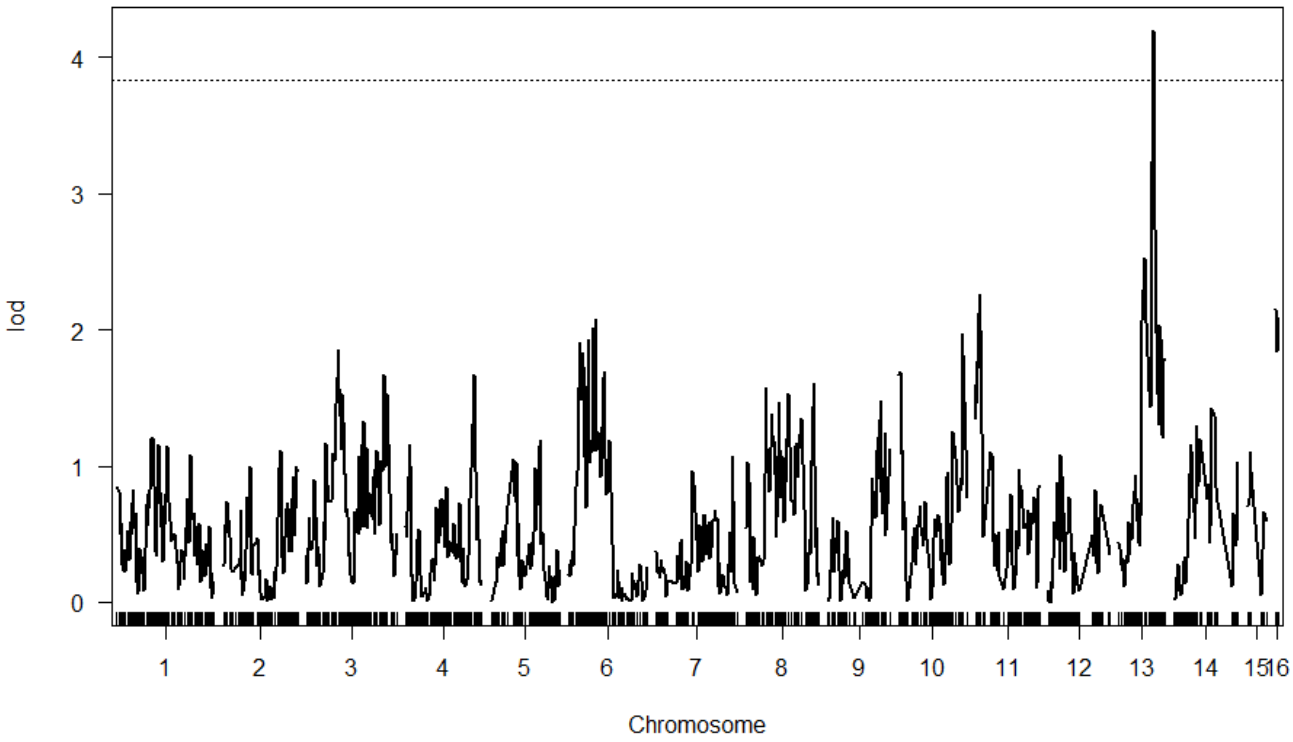


Figure S3.6

LOD plot of whether flowers were produced Dashed line indicates genome-wide significance at the 0.05 level.

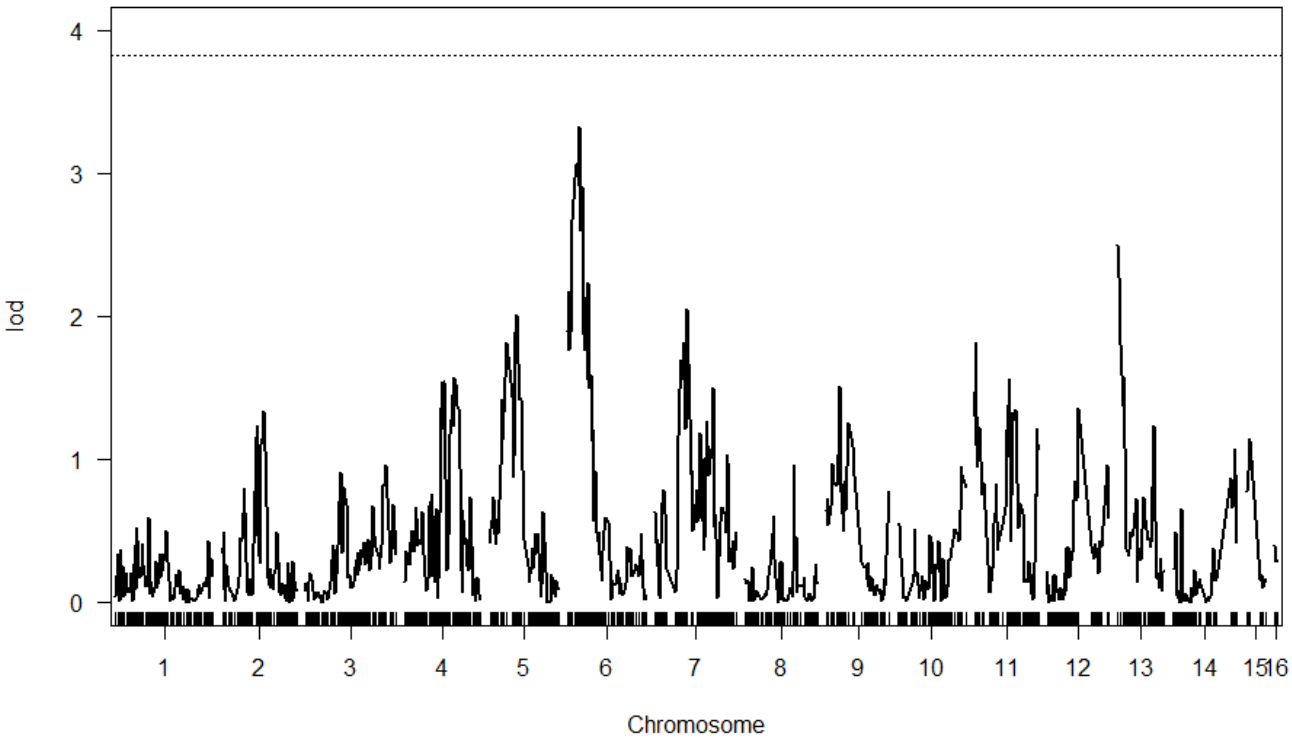


Figure S3.7

LOD plot of number of flowers produced per day. Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG6 is significant at chromosome-wide level.

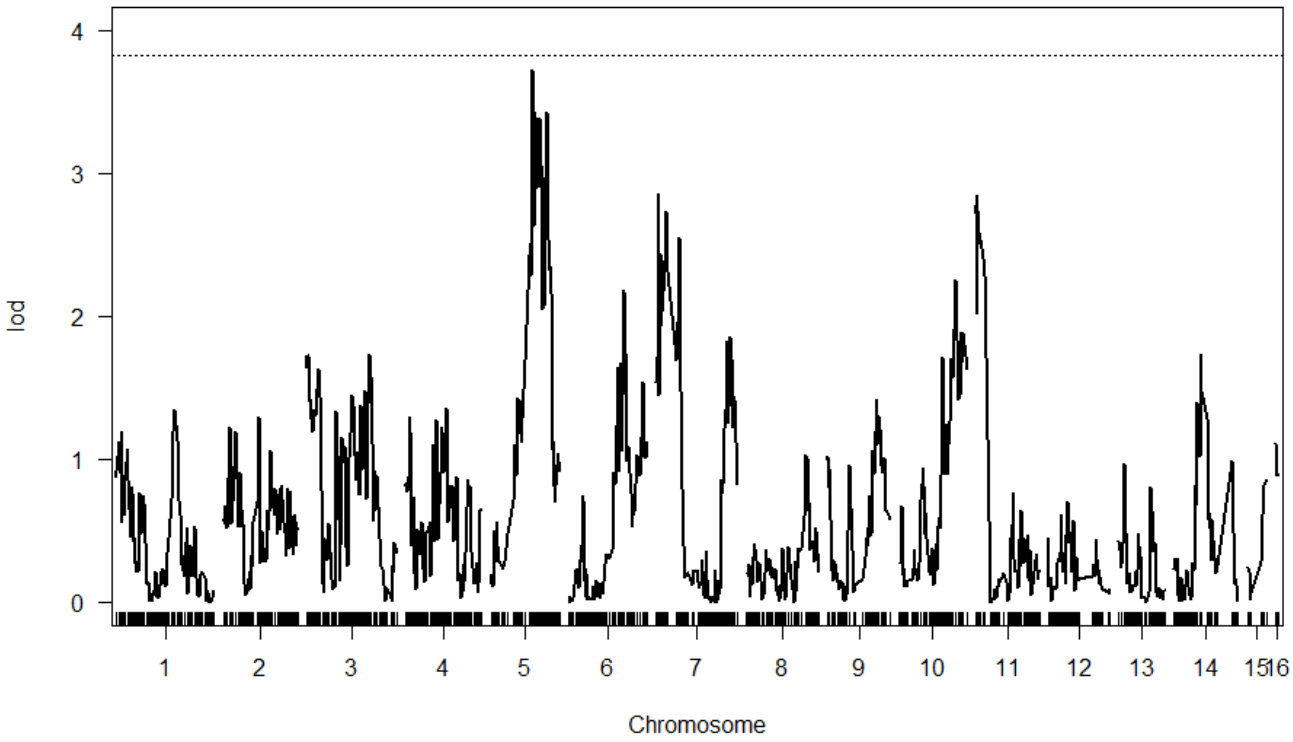


Figure S3.8

LOD plot of cyme length. Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG5 is significant at chromosome-wide level.

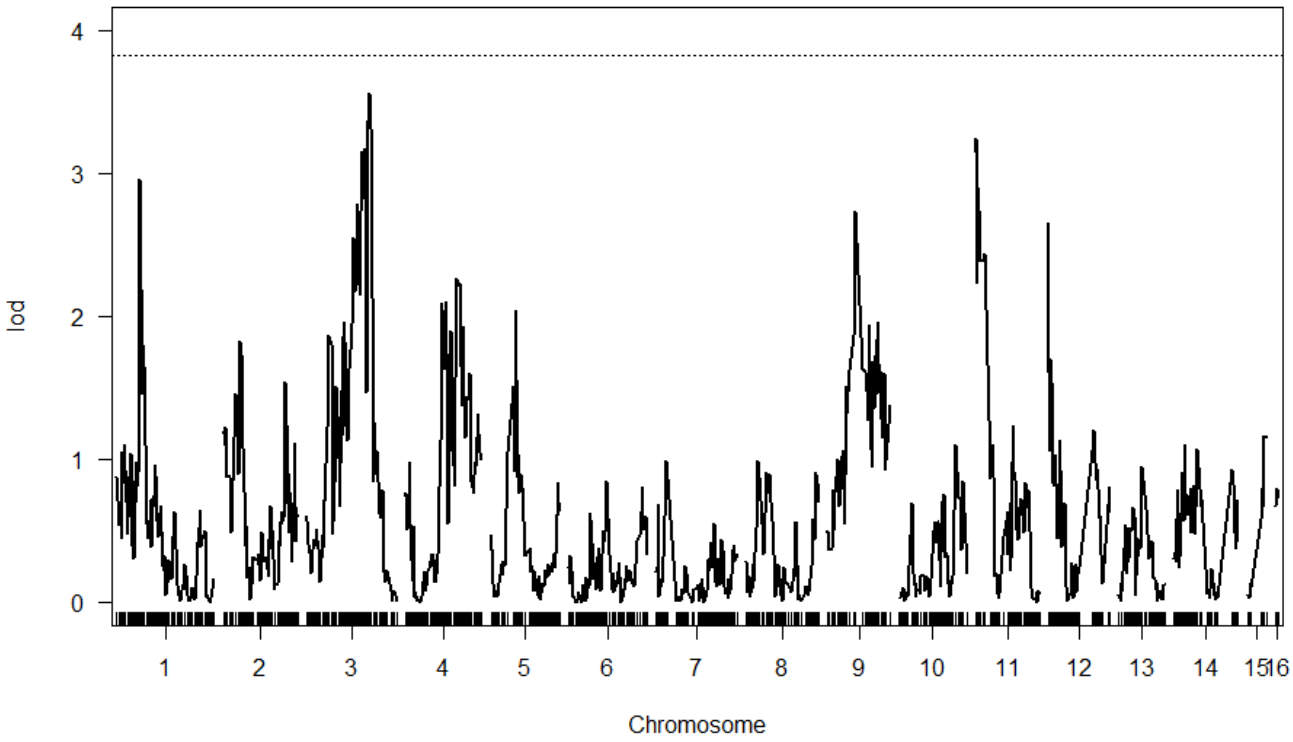


Figure S3.9

LOD plot of flowers per inflorescence. Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG11 is significant at chromosome-wide level.

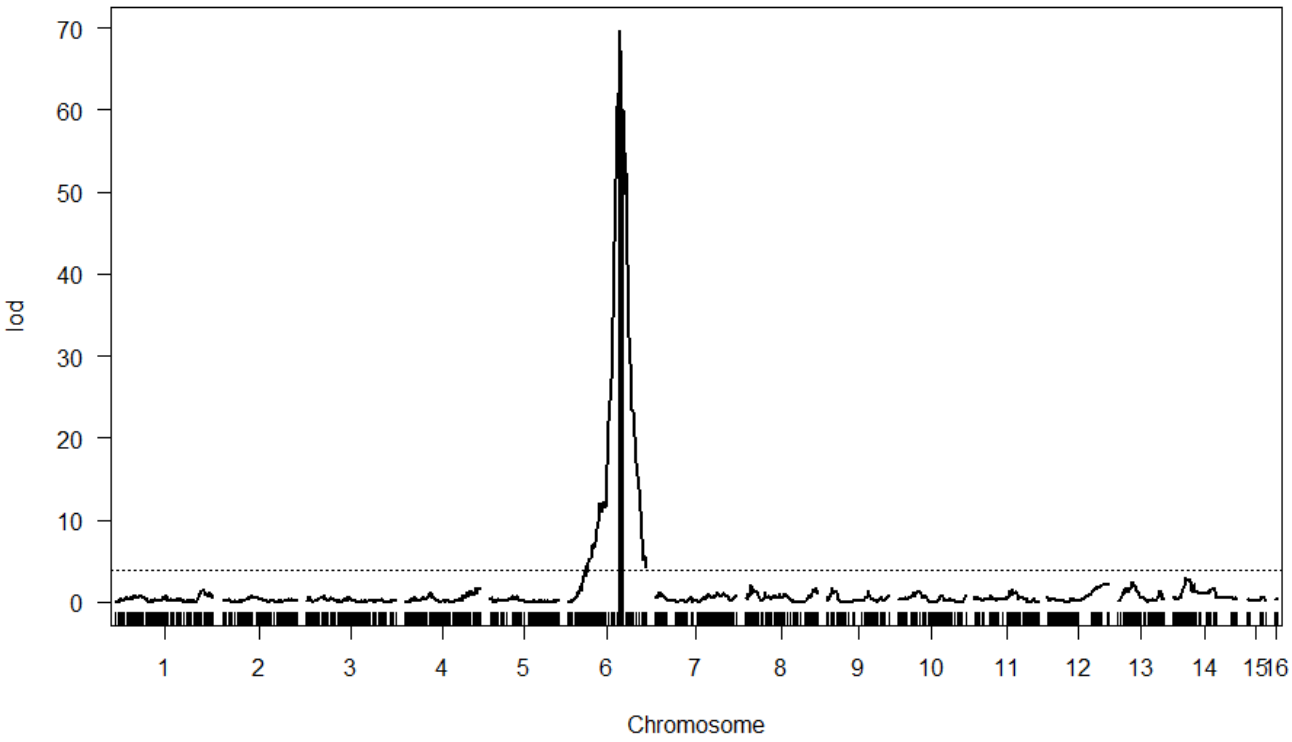


Figure S3.10

LOD plot of flower color. Dashed line indicates genome-wide significance at the 0.05 level. A candidate locus maps to the same position.

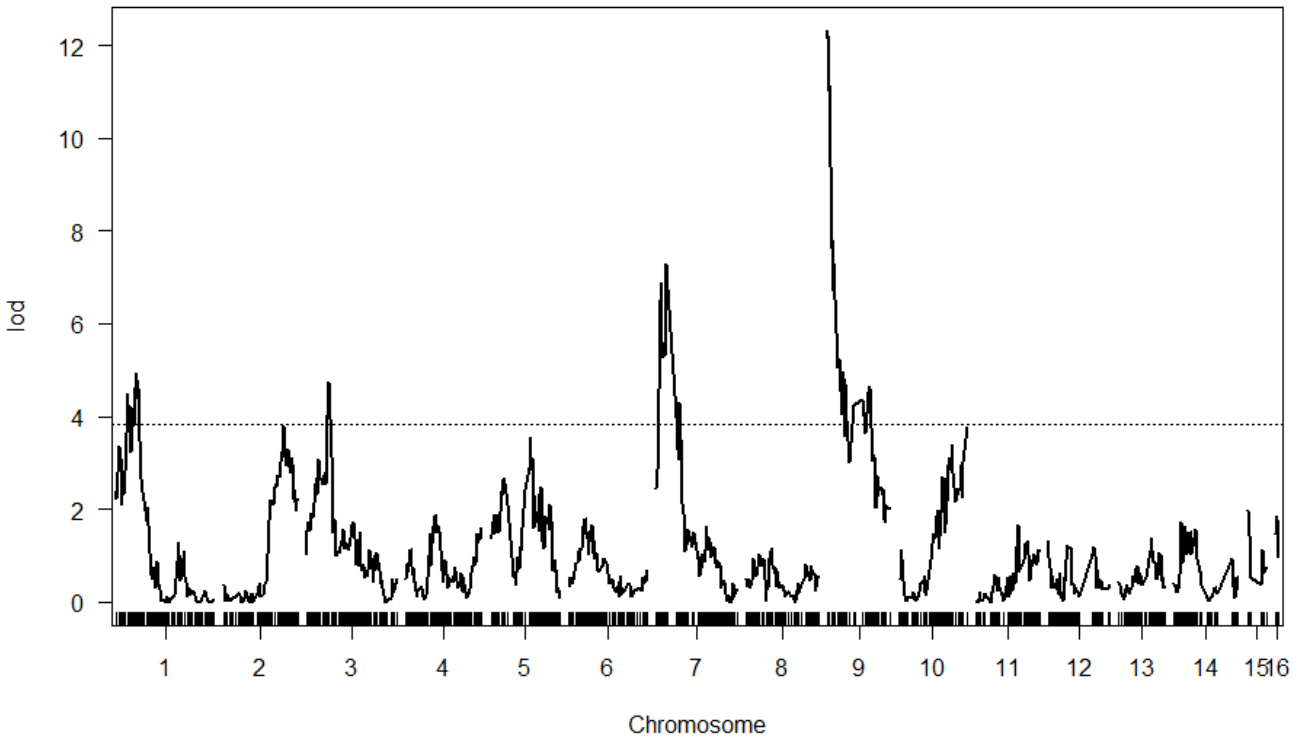


Figure S3.11

LOD plot of corolla length. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG2, LG5 and LG10 are significant at the chromosome-wide level.

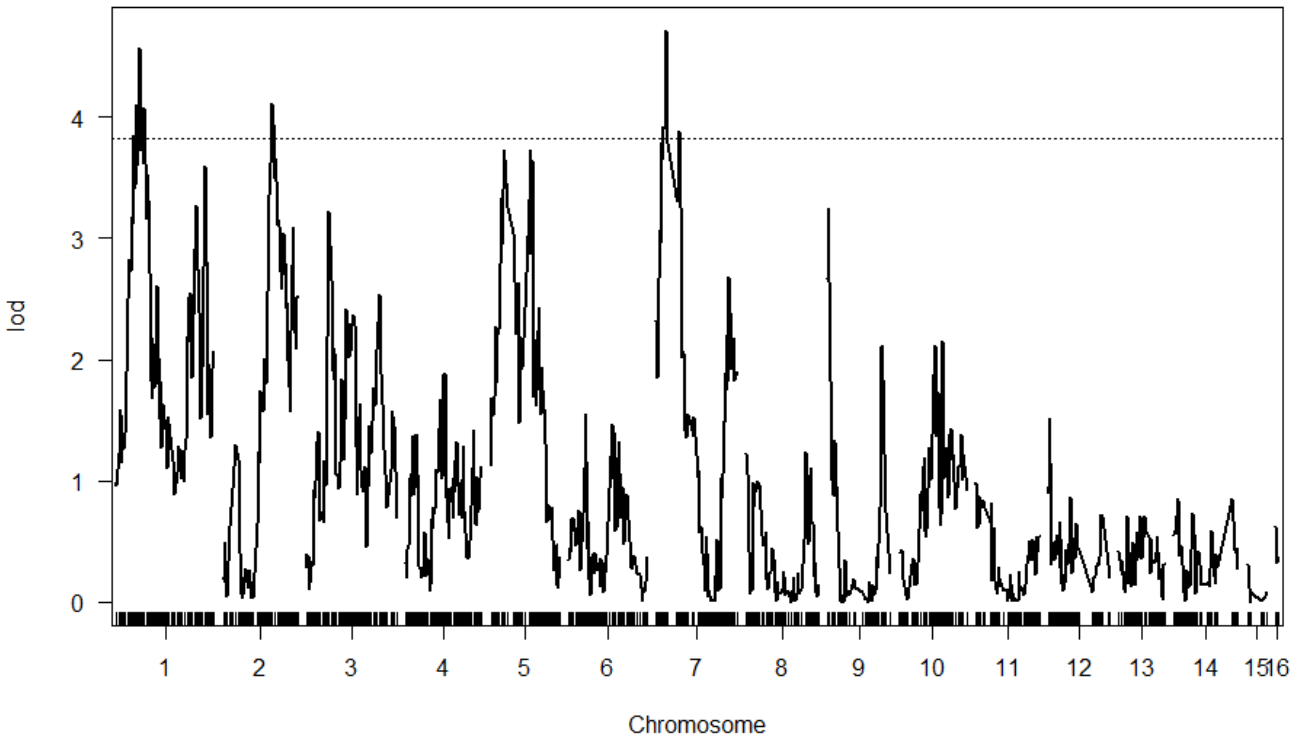


Figure S3.12

LOD plot of corolla width. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG3 and LG9 are significant at the chromosome-wide level.

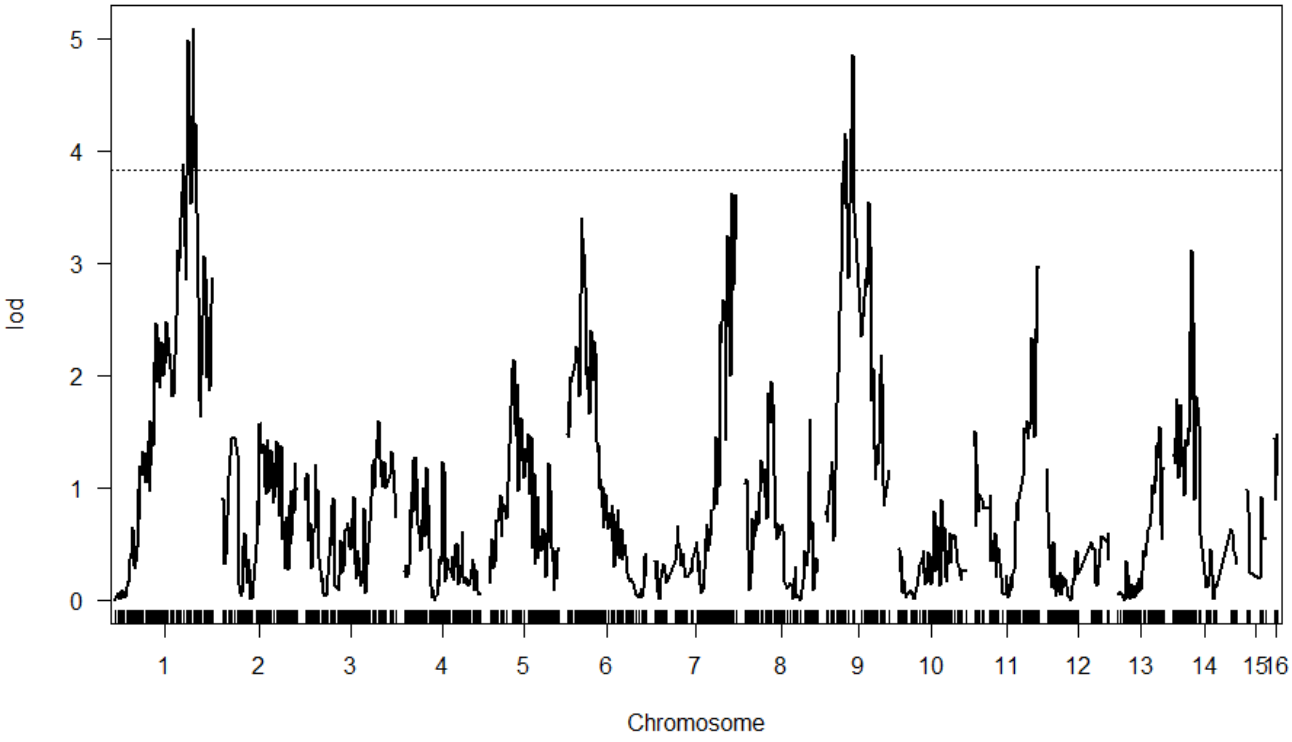


Figure S3.13

LOD plot of corolla shape (length / width). Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG6, LG7 and LG14 are significant at the chromosome-wide level.

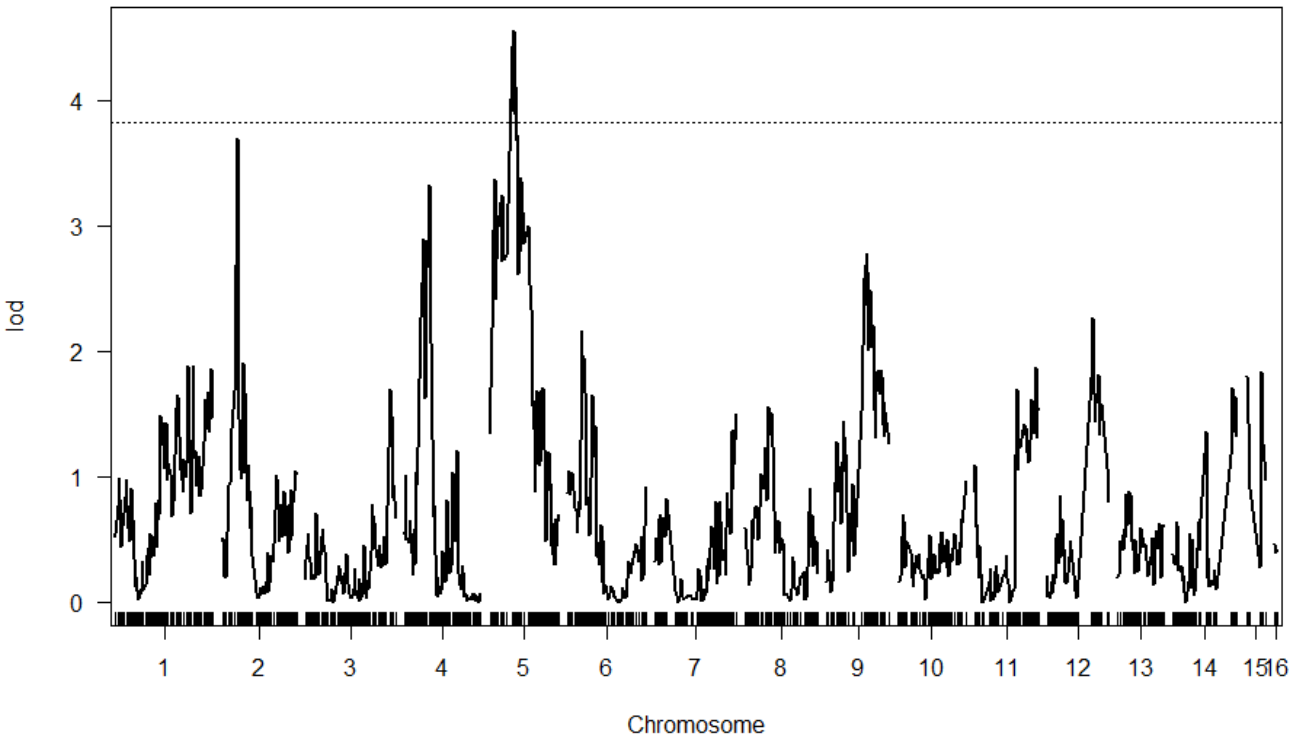


Figure S3.14

LOD plot of corolla shape (tissue length / width). Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG4 and LG2 are significant at the chromosome-wide level.

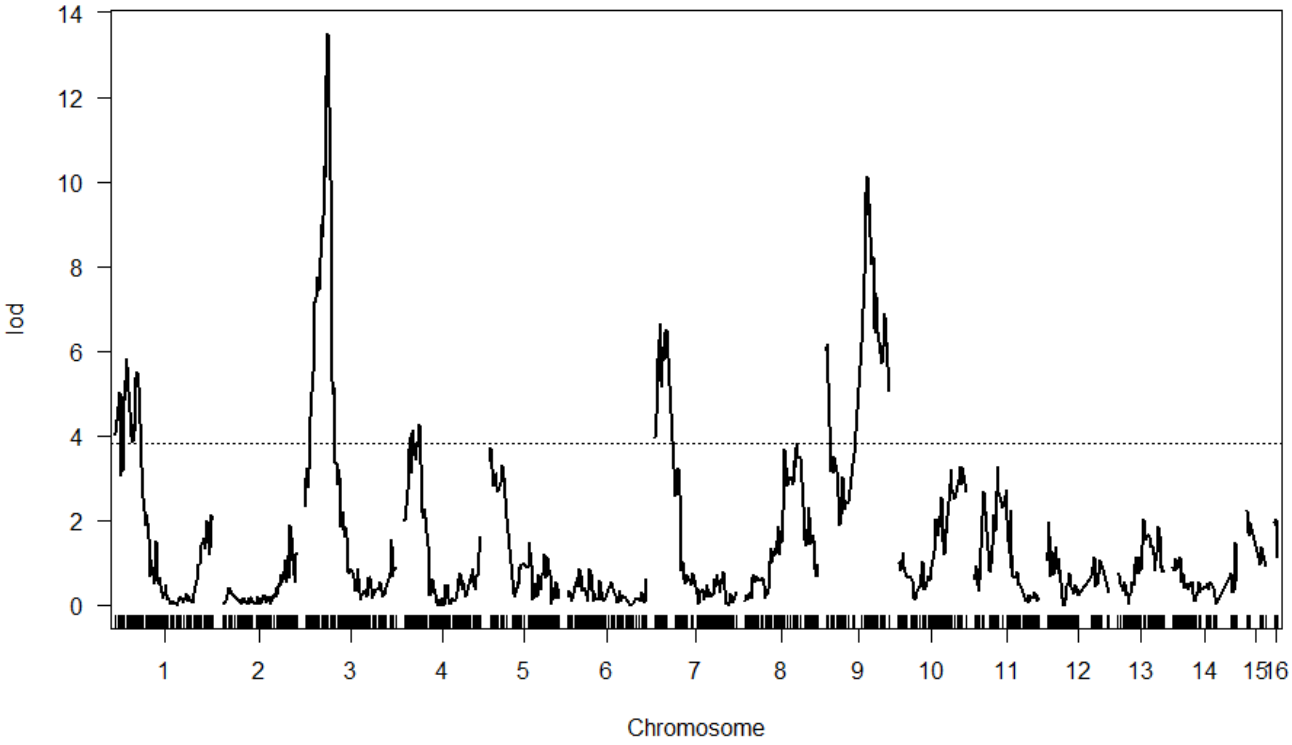


Figure S3.15

LOD plot of style-plus-ovary length. Dashed line indicates genome-wide significance at the 0.05 level. Peaks on LG5, LG10 and LG11 are significant at the chromosome-wide level.

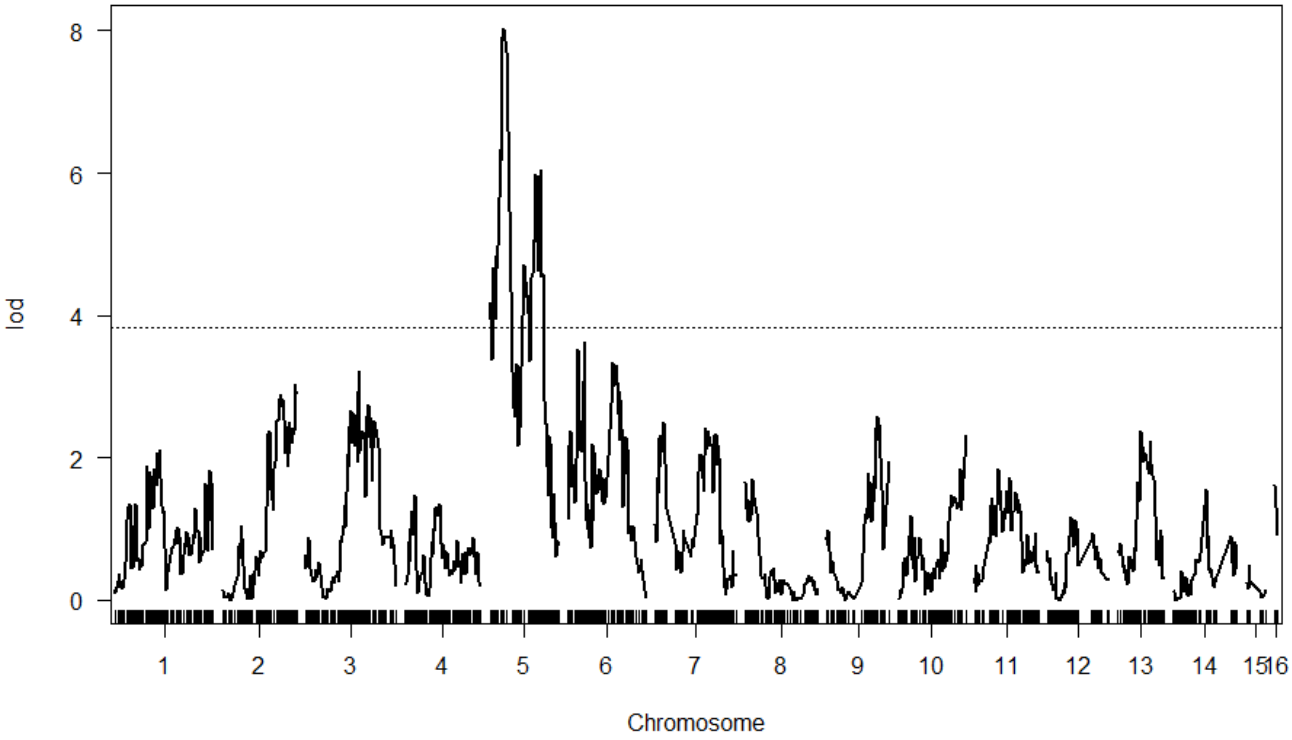


Figure S3.16

LOD plot of nectar volume. Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG13 is significant in non-parametric analysis.

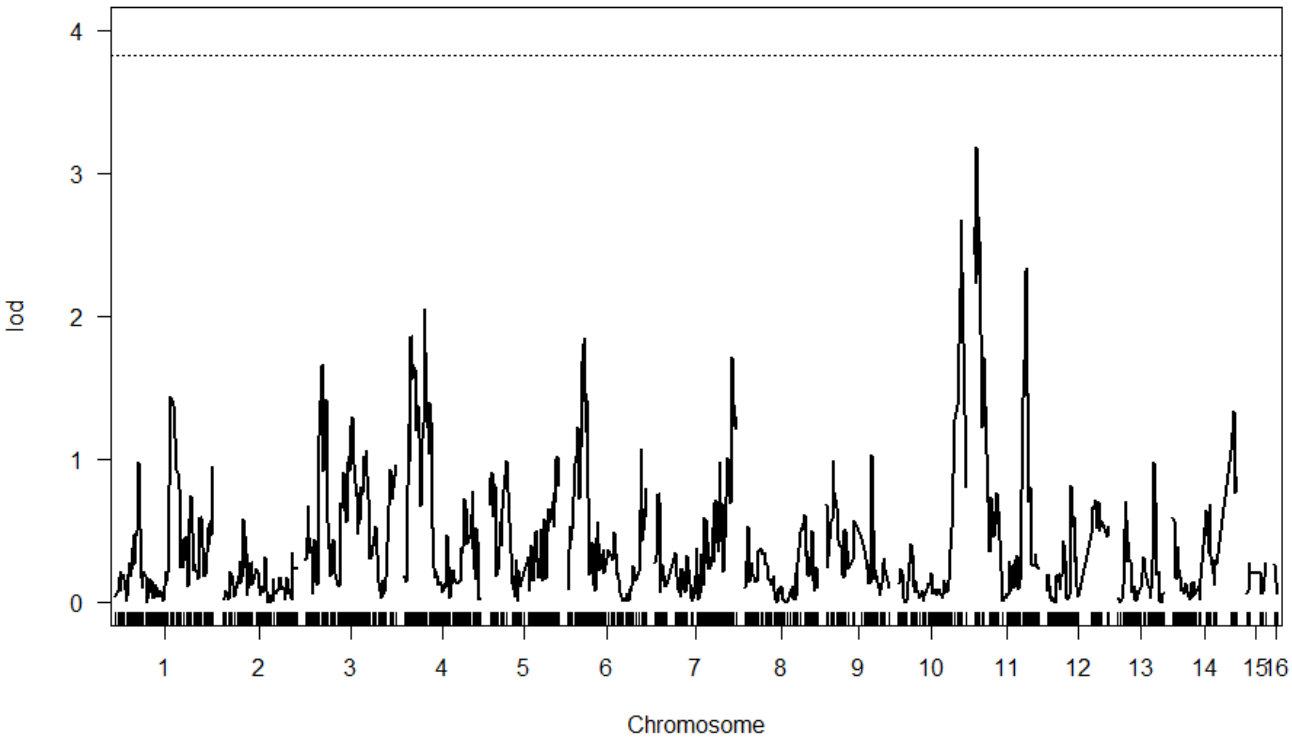


Figure S3.17

LOD plot of pollen count (ImageJ) Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG11 is significant at the chromosome-wide level.

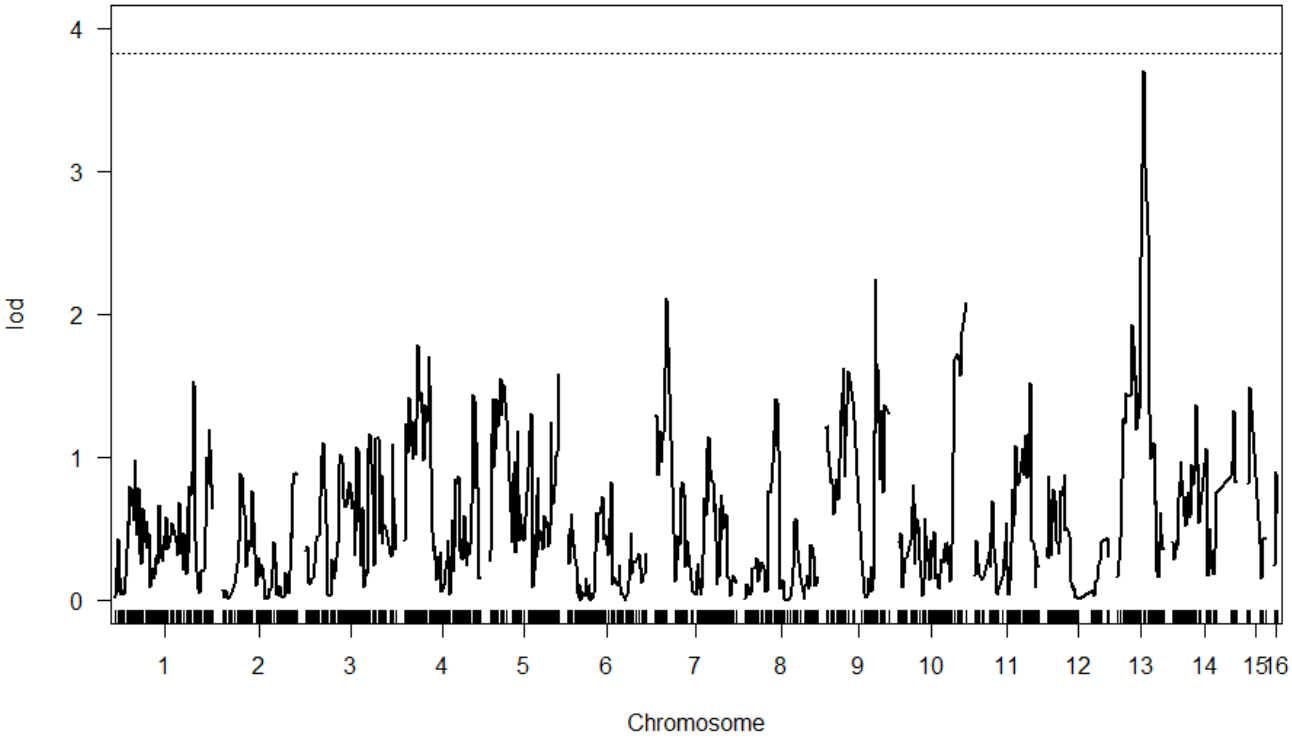


Figure S3.18

LOD plot of pollen count (Coulter counter) Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG13 is significant at the chromosome-wide level.

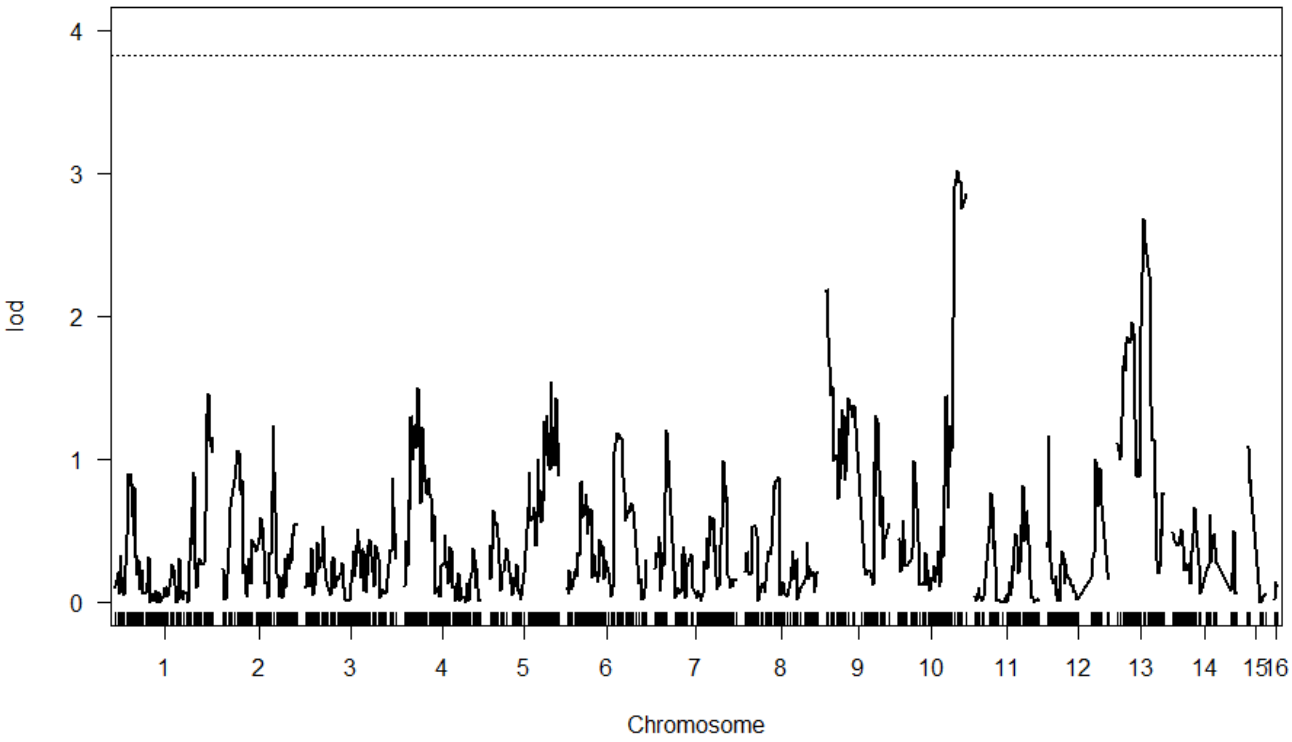


Figure S3.19

LOD plot of pollen diameter (Coulter counter) Dashed line indicates genome-wide significance at the 0.05 level. Peak on LG10 is significant at the chromosome-wide level.

Table S3.4. QTL sign test results. Note that for nectar volume, only the *I. cordatotriloba* standard deviations were used because *I. lacunosa* produced no nectar. LNorm: log normal. Exp: Exponential. RHE: relative homozygous effect (additive effect divided by difference between grandparents of the mapping population). FitQTL: variance explained, using FitQTL in R/QTL.

Trait	QTL measure	Grandparent St. Dev. Used	Distribution of QTL effects	P
Corolla length	RHE	*I. cordatotriloba*	Lnorm	1.83E-02
Corolla length	RHE	*I. lacunosa*	Lnorm	1.99E-02
Corolla length	RHE	Average	Lnorm	0.01905
Corolla length	FitQTL	*I. cordatotriloba*	Lnorm	0.02381
Corolla length	FitQTL	*I. lacunosa*	Lnorm	0.0304
Corolla length	FitQTL	Average	Lnorm	0.02685
Shape (length/width)	RHE	*I. cordatotriloba*	Lnorm	0.7867
Shape (length/width)	RHE	*I. lacunosa*	Lnorm	0.7863
Shape (length/width)	RHE	Average	Lnorm	0.7865
Shape (length/width)	FitQTL	*I. cordatotriloba*	Lnorm	0.7991
Shape (length/width)	FitQTL	*I. lacunosa*	Lnorm	0.7956
Shape (length/width)	FitQTL	Average	Lnorm	0.7973
Corolla width	RHE	*I. cordatotriloba*	Lnorm	0.1554
Corolla width	RHE	*I. lacunosa*	Lnorm	0.169
Corolla width	RHE	Average	Lnorm	0.1621
Corolla width	FitQTL	*I. cordatotriloba*	Lnorm	0.1904
Corolla width	FitQTL	*I. lacunosa*	Lnorm	0.2203
Corolla width	FitQTL	Average	Lnorm	0.2052
Nectar volume	RHE	*I. cordatotriloba*	Lnorm	0.5782
Nectar volume	FitQTL	*I. cordatotriloba*	Exp	0.6018
Style plus ovary	RHE	*I. cordatotriloba*	Exp	0.2844
Style plus ovary	RHE	*I. lacunosa*	Exp	0.2925
Style plus ovary	RHE	Average	Exp	0.2884
Style plus ovary	FitQTL	*I. cordatotriloba*	Exp	0.3008
Style plus ovary	FitQTL	*I. lacunosa*	Exp	0.3244
Style plus ovary	FitQTL	Average	Exp	0.3124

Table S3.5 QTL numbers and effect sizes for floral morphology traits, other floral traits, life history traits, and other categories of traits in other QTL studies of the selfing syndrome

Study	**Floral morphology (n QTLs)**	**Effect sizes (avg. by trait)**	**Other floral (n QTLs)**	**Effect sizes (avg. by trait)**	**Life history (n QTLs)**	**Effect sizes (avg. by trait)**	**Other (n QTLs)**	**Effect sizes (avg. by trait)**	**Notes**
Bernacchi and Tanksley 1997	Stigma exsertion (1), bud type (7), corolla indentation (2), flower size (3),	0.197, 0.137, 0.116, 0.099	Rachis length (3), inflorescence vegetative meristem (2)	0.207, 0.090	Number of flowers (1)	0.284	Self-incompatibility (4), unilateral incompatibility (6)	0.117, 0.121	Effect sizes reported as R2
Fishman 2002	Throat width (14), corolla width (14), tube length (13), corolla length (11), style length (12), stamen length (13), stigma-anther position (15)	0.231, 0.701, 0.350, 0.788, 0.357, 0.318, 0.151	NA	NA	NA	NA	NA	NA	Effect sizes reported as relative homozygous effect. Lower effect sizes reported from composite interval mapping.

Fishman 2015	Corolla length (6), pistil length (7), tube length (7), stigma-anther separation (3), long stamen length (6), corolla width (3), throat width (3)	0.143, 0.123, 0.137, 0.513, 0.138, 0.187, 0.19	Calyx length (4)	0.133	Time to flower (4)	0.555	Anther sterility (3)	NA	Effect sizes as relative homozygous effect.
Georgiady et al. 2002	Anther total length (1), anther sterile length (2), style length (1)	0.352, 0.199, 0.422	Flowers per inflorescence (1)	0.298	NA	NA	NA	NA	Significant QTL only. 13 non-significant QTL were also reported. Effect sizes as relative homozygous effect.
Goodwillie 2006	Corolla tube (8), corolla lobe length (8), corolla lobe width (6), anther length (7), stigma length (4)	0.201, 0.140, 0.188, 0.239, 0.376	NA	NA	NA	NA	NA	NA	Effect sizes as relative homozygous effect.

Lin and Ritland 1997	Flower length (1), pistil length (1), long stamen length (3), short stamen length (3)	0.078, 0.286, 0.128, 0.112	Sepal length (2)	0.106	NA	NA	NA	NA	Effect sizes as relative homozygous effect.
Slotte 2012	Lateral stamen length (5), median stamen length (5), petal length (3), petal width (5), style+gynoecium length (2)	0.089, 0.098, 0.201, 0.129, 0.070	Lateral sepal length (2), median sepal length (2), ovule number (2), pollen number (3)	0.114, 0.125, 0.048, 0.065	Days to flowering (3), leaf number at flowering (4)	0.053, 0.098	Homeotic floral abnormalities (4), self-compatibility (1)	0.033, NA	Effect sizes as relative homozygous effect.

Appendices go here, after the text, before the references & biography.

Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography. Appendices go here, after the text, before the references & biography.

References

Anderson, Ingrid, and Jeremiah Busch. 2006. "Pollinator - Mediated Selection Weakens Floral Integration in Self - Compatible." *American Journal of Botany* 93 (6): 860–67.

Armbruster, WS, C Pélabon, TF Hansen, and CPH Mulder. 2004. "Floral Integration, Modularity, and Accuracy." In *Phenotypic Integration*, 23–49. Oxford University Press.

Arunkumar, Ramesh, Rob W. Ness, Stephen I. Wright, and Spencer C Barrett. 2015. "The Evolution of Selfing Is Accompanied by Reduced Efficacy of Selection and Purging of Deleterious Mutations." *Genetics* 199 (3): 817–29. doi:10.1534/genetics.114.172809.

Ashman, T-L, and C J Majetic. 2006. "Genetic Constraints on Floral Evolution: A Review and Evaluation of Patterns." *Heredity* 96 (5): 343–52. doi:10.1038/sj.hdy.6800815.

Austin, Daniel F. 1978. "The Ipomoea Batatas Complex-I. Taxonomy." *Bulletin of the Torrey Botanical Club* 105 (2): 114–29.

Barrett, Spencer C. 2002. "The Evolution of Plant Sexual Diversity." *Nature Reviews. Genetics* 3 (4): 274–84. doi:10.1038/nrg776.

Barrett, Spencer C, Ramesh Arunkumar, and Stephen I Wright. 2014. "The Demography and Population Genomics of Evolutionary Transitions to Self-Fertilization in Plants." *Proceedings of the Royal Society B.* 369 (1648): 1–9. http://rstb.royalsocietypublishing.org/content/369/1648/20130344.short.

Barrett, Spencer C, and LD D Harder. 1996. "Ecology and Evolution of Plant Mating." *Trends in Ecology & Evolution* II (2): 73–79. http://www.sciencedirect.com/science/article/pii/0169534796810469.

Bernacchi, Dario, and SD Tanksley. 1997. "An Interspecific Backcross of Lycopersicon Esculentum X L. Hirsutum: Linkage Analysis and a QTL Study of Sexual Compatibility Factors and Floral Traits." *Genetics* 14850: 861–77. http://www.genetics.org/content/147/2/861.short.

Bolten, Alan B., Peter Feinsinger, Herbert G. Baker, and Irene Baker. 1979. "On the Calculation of Sugar Concentration in Flower Nectar." *Oecologia* 18: 301–4.

Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. Buckler. 2007. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics* 23 (19): 2633–35.

doi:10.1093/bioinformatics/btm308.

Bradshaw, H. D., Kevin G. Otto, Barbara E. Frewen, John K. McKay, and Douglas W. Schemske. 1998. "Quantitative Trait Loci Affecting Differences in Floral Morphology between Two Species of Monkeyflower (Mimulus)." *Genetics* 149 (1): 367–82.

Brandvain, Yaniv, Amanda M. Kenney, Lex Flagel, Graham Coop, and Andrea L. Sweigart. 2014. "Speciation and Introgression between Mimulus Nasutus and Mimulus Guttatus." *PLoS Genetics* 10 (6). doi:10.1371/journal.pgen.1004410.

Brandvain, Yaniv, Tanja Slotte, Khaled M Hazzouri, Stephen I Wright, and Graham Coop. 2013. "Genomic Identification of Founding Haplotypes Reveals the History of the Selfing Species Capsella Rubella." *PLoS Genetics* 9 (9): e1003754. doi:10.1371/journal.pgen.1003754.

Broman, Karl W. 2010. "Genetic Map Construction with R/qtl," 1–41.

Broman, Karl W., and Śaunak Sen. 2009. *A Guide to QTL Mapping with R/qtl*. New York: Springer. doi:10.1007/978-0-387-92125-9.

Broman, Karl W, Hao Wu, and Gary A Churchill. 2003. "R / Qtl®: QTL Mapping in Experimental Crosses." *Bioinformatics (Oxford, England)* 19 (7): 889–90. doi:10.1093/bioinformatics/btg112.

Busch, Jeremiah W., Simon Joly, and Daniel J. Schoen. 2011. "Demographic Signatures Accompanying the Evolution of Selfing in Leavenworthia Alabamica." *Molecular Biology and Evolution* 28 (5): 1717–29. doi:10.1093/molbev/msq352.

Cantarel, Brandi L., Ian Korf, Sofia M C Robb, Genis Parra, Eric Ross, Barry Moore, Carson Holt, Alejandro Sánchez Alvarado, and Mark Yandell. 2008. "MAKER: An Easy-to-Use Annotation Pipeline Designed for Emerging Model Organism Genomes." *Genome Research* 18 (1): 188–96. doi:10.1101/gr.6743907.

Cao, Jun, Korbinian Schneeberger, Stephan Ossowski, Torsten Günther, Sebastian Bender, Joffrey Fitz, Daniel Koenig, et al. 2011. "Whole-Genome Sequencing of Multiple Arabidopsis Thaliana Populations." *Nature Genetics* 43 (10): 956–63. doi:10.1038/ng.911.

Chang, Shu-Mei, and Mark D. Rausher. 1998. "Frequency-dependent Pollen Discounting Contributes to Maintenance of a Mixed Mating System in the Common Morning Glory Ipomoea Purpurea." *The American Naturalist* 152 (5): 671–83.

doi:10.1086/286198.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. "The Effect of Deleterious Mutations on Neutral Molecular Variation." *Genetics* 134 (4): 1289–1303. doi:10.1111/j.0014-3820.2002.tb00188.x.

Charlesworth, Brian. 2009. "Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10 (3): 195–205. doi:10.1038/nrg2526.

Charlesworth, Deborah, and Stephen I Wright. 2001. "Breeding Systems and Genome Evolution." *Current Opinion in Genetics & Development* 11 (6): 685–90. doi:10.1016/S0959-437X(00)00254-9.

Coffman, Alec J., Ping Hsun Hsieh, Simon Gravel, and Ryan N. Gutenkunst. 2015. "Computationally Efficient Composite Likelihood Statistics for Demographic Inference." *Molecular Biology and Evolution* 33 (2): 591–93. doi:10.1093/molbev/msv255.

Darwin, Charles. 1877. *The Various Contrivances by Which Orchids Are Fertilised by Insects.* London: John Murray.

Delignette-Muller, Marie Laure, and Christophe Dutang. 2015. "{fitdistrplus}: An {R} Package for Fitting Distributions." *Journal of Statistical Software* 64 (4): 1–34. http://www.jstatsoft.org/v64/i04/.

Diaz, Jaime, Peter Schmiediche, and DF Austin. 1996. "Polygon of Crossability between Eleven Species of Ipomoea: Section Batatas (Convolvulaceae)." *Euphytica*, no. 1979: 189–200. http://link.springer.com/article/10.1007/BF00023890.

Dobin, Alexander, and Thomas R. Gingeras. 2015. "Mapping RNA-Seq Reads with STAR." *Current Protocols in Bioinformatics* 51: 11.14.1-11.14.9. doi:10.1177/0963721412473755.Surging.

Donohue, Kathleen. 2009. "Completing the Cycle: Maternal Effects as the Missing Link in Plant Life Histories." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1520): 1059–74. doi:10.1098/rstb.2008.0291.

Duncan, Tanya M, and Mark D Rausher. 2013a. "Evolution of the Selfing Syndrome in Ipomoea." *Frontiers in Plant Science* 4 (August): 1–8. doi:10.3389/fpls.2013.00301.

———. 2013b. "Morphological and Genetic Differentiation and Reproductive Isolation

among Closely Related Taxa in the Ipomoea Series Batatas." *American Journal of Botany* 100 (11): 2383–2193. doi:10.3732/ajb.1200467.

Escobar, Juan S., Alberto Cenci, Jeremy Bolognini, Annabelle Haudry, Stefan Laurent, Jacques David, and Sylvain Glémin. 2010. "An Integrative Test of the Dead-End Hypothesis of Selfing Evolution in Triticeae (Poaceae)." *Evolution* 64 (10): no-no. doi:10.1111/j.1558-5646.2010.01045.x.

Fenster, CB, and WS Armbruster. 2004. "Pollination Syndromes and Floral Specialization." *Annual Review of Ecology, Evolution, and Systematics* 35 (2004): 375–403. doi:10.2307/annurev.ecolsys.34.011802.30000015.

Fisher, RA. 1941. "Average Excess and Average Effect of a Gene Substitution." *Annals of Human Genetics*. http://onlinelibrary.wiley.com/doi/10.1111/j.1469-1809.1941.tb02272.x/abstract.

Fishman, Lila, Jan Aagaard, and John C. Tuthill. 2008. "Toward the Evolutionary Genomics of Gametophytic Divergence: Patterns of Transmission Ratio Distortion in Monkeyflower (Mimulus) Hybrids Reveal a Complex Genetic Basis for Conspecific Pollen Precedence." *Evolution* 62 (12): 2958–70. doi:10.1111/j.1558-5646.2008.00475.x.

Fishman, Lila, Paul M Beardsley, Angela Stathos, Charles F Williams, and Jeffrey P Hill. 2015. "The Genetic Architecture of Traits Associated with the Evolution of Self-Pollination in Mimulus." *New Phytologist* 205: 907–17.

Fishman, Lila, Alan J Kelly, and John H Willis. 2002. "Minor Quantitative Trait Loci Underlie Floral Traits Associated with Mating System Divergence in Mimulus." *Evolution* 56 (11): 2138–55. http://www.ncbi.nlm.nih.gov/pubmed/12487345.

Foxe, John Paul, Tanja Slotte, Eli A Stahl, Barbara Neuffer, Herbert Hurka, and Stephen I Wright. 2009. "Recent Speciation Associated with the Evolution of Selfing in Capsella." *Proceedings of the National Academy of Sciences of the United States of America* 106 (13): 5241–45. doi:10.1073/pnas.0807679106.

Foxe, John Paul, Marc Stift, Andrew Tedder, Annabelle Haudry, Stephen I. Wright, and Barbara K. Mable. 2010. "Reconstructing Origins of Loss of Self-Incompatibility and Selfing in North American Arabidopsis Lyrata: A Population Genetic Context." *Evolution* 64 (12): 3495–3510. doi:10.1111/j.1558-5646.2010.01094.x.

Garner, Austin G, Amanda M Kenney, Lila Fishman, and Andrea L Sweigart. 2015. "Genetic Loci with Parent of Origin Effects Cause Hybrid Seed Lethality between

Mimulus Species." doi:10.1111/nph.13897.

Georgiady, MS, RW Whitkus, and EM Lord. 2002. "Genetic Analysis of Traits Distinguishing Outcrossing and Self-Pollinating Forms of Currant Tomato, Lycopersicon Pimpinellifolium (Jusl.) Mill." *Genetics* 344 (May): 333–44. http://www.genetics.org/content/161/1/333.short.

Glémin, Sylvain. 2007. "Mating Systems and the Efficacy of Selection at the Molecular Level." *Genetics* 177 (2): 905–16. doi:10.1534/genetics.107.073601.

Glover, Beverley J., Chiara A. Airoldi, Samuel F. Brockington, Mario Fernández-Mazuecos, Cecilia Martínez-Pérez, Greg Mellers, Edwige Moyroud, and Lin Taylor. 2015. "How Have Advances in Comparative Floral Development Influenced Our Understanding of Floral Evolution?" *International Journal of Plant Sciences* 176 (4): 307–23. doi:10.1086/681562.

Goldberg, Emma E, and Boris Igić. 2012. "Tempo and Mode in Plant Breeding System Evolution." *Evolution; International Journal of Organic Evolution* 66 (12): 3701–9. doi:10.1111/j.1558-5646.2012.01730.x.

Goodwillie, Carol, Susan Kalisz, and Christopher G. Eckert. 2005. "The Evolutionary Enigma of Mixed Mating Systems in Plants: Occurrence, Theoretical Explanations, and Empirical Evidence." *Annual Review of Ecology, Evolution, and Systematics* 36 (1): 47–79. doi:10.1146/annurev.ecolsys.36.091704.175539.

Goodwillie, Carol, C Ritland, and Kermit Ritland. 2006. "The Genetic Basis of Floral Traits Associated with Mating System Evolution in Leptosiphon (Polemoniaceae): An Analysis of Quantitative Trait Loci." *Evolution* 60 (3): 491–504. http://www.ncbi.nlm.nih.gov/pubmed/16637495.

Goodwillie, Carol, Risa D Sargent, Christopher G Eckert, Elizabeth Elle, Monica A Geber, Mark O Johnston, Susan Kalisz, et al. 2010. "Correlated Evolution of Mating System and Floral Display Traits in Flowering Plants and Its Implications for the Distribution of Mating System Variation." *The New Phytologist* 185 (1): 311–21. doi:10.1111/j.1469-8137.2009.03043.x.

Guo, Ya-Long, Jesper S Bechsgaard, Tanja Slotte, Barbara Neuffer, Martin Lascoux, Detlef Weigel, and Mikkel H Schierup. 2009. "Recent Speciation of Capsella Rubella from Capsella Grandiflora, Associated with Loss of Self-Incompatibility and an Extreme Bottleneck." *Proceedings of the National Academy of Sciences* 106 (13): 5246–51. doi:10.1073/pnas.0808012106.

Gutenkunst, Ryan N., Ryan D. Hernandez, Scott H. Williamson, and Carlos D. Bustamante. 2009. "Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data." *PLoS Genetics* 5 (10). doi:10.1371/journal.pgen.1000695.

Hall, Megan C, and John H Willis. 2005. "Transmission Ratio Distortion in Intraspecific Hybrids of Mimulus Guttatus: Implications for Genomic Divergence." *Genetics* 170 (1): 375–86. doi:10.1534/genetics.104.038653.

Hansen, Thomas E. 2013. "The Evolution of Genetic Architecture." *Annual Review of Ecology, Evolution, and Sytematics* 37 (2006): 123–57. doi:10.2307/annurev.ecolsys.37.091305.30000007.

Harrell Jr, Frank E, with contributions from Charles Dupont, and many others. 2016. "Hmisc: Harrell Miscellaneous." https://cran.r-project.org/package=Hmisc.

Haudry, A, A Cenci, C Guilhaumon, E Paux, S Poirier, S Santoni, J David, and S Glémin. 2008. "Mating System and Recombination Affect Molecular Evolution in Four Triticeae Species." *Genetics Research* 90 (May): 97–109. doi:10.1017/S0016672307009032.

Hazzouri, Khaled M., Juan S. Escobar, Rob W. Ness, L. Killian Newman, April M. Randle, Susan Kalisz, and Stephen I. Wright. 2013. "Comparative Population Genomics in Collinsia Sister Species Reveals Evidence for Reduced Effective Population Size, Relaxed Selection, and Evolution of Biased Gene Conversion with an Ongoing Mating System Shift." *Evolution* 67 (5): no-no. doi:10.1111/evo.12027.

Hirakawa, H., Y. Okada, H. Tabuchi, K. Shirasawa, a. Watanabe, H. Tsuruoka, C. Minami, et al. 2015. "Survey of Genome Sequences in a Wild Sweet Potato, Ipomoea Trifida (H. B. K.) G. Don." *DNA Research*, 1–9. doi:10.1093/dnares/dsv002.

Holland, James B, Wyman E Nyquist, and Cuauhtemoc T Cervantes-Martinez. 2003. "Estimating and Interpreting Heritability for Plant Breeding.Pdf." *Plant Breed. Rev.* doi:10.1002/9780470650202.ch2.

Institute, DoE Joint Genome. 2017. "Mimulus Genome Project." *Institute, DoE Joint Genome*. http://genome.jgi.doe.gov/mimulus/mimulus.home.html.

Invarsson, Pär K. 2002. "A Metapopulation Perspective on Genetic Diversity and Differentiation in Partially Self- Fertilizing Plants." *Evolution* 56 (12): 2368–73.

Iverson, Kenneth E. 1962. "A Programming Language." Proceedings of the May 1-3,

1962, Spring Joint Computer Conference. ACM, 1962.

Ivey, Christopher T., Leah S. Dudley, Alisa A. Hove, Simon K. Emms, and Susan J. Mazer. 2016. "Outcrossing and Photosynthetic Rates Vary Independently within Two *Clarkia* Species: Implications for the Joint Evolution of Drought Escape Physiology and Mating System." *Annals of Botany*, mcw134. doi:10.1093/aob/mcw134.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. "NCBI BLAST: A Better Web Interface." *Nucleic Acids Research* 36 (Web Server issue): 5–9. doi:10.1093/nar/gkn201.

Koch, M, B Haubold, and T Mitchell-Olds. 2000. "Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase in Arabidopsis, Arabis, and Related Genera (Brassicaceae)." *Molecular Biology and Evolution* 17: 1483–98.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, and Adam M. Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation." *Genome Research* 27: 1–15. doi:10.1101/071282.

Leinonen, Tuomas, R J Scott McCairns, Robert B O'Hara, and Juha Merilä. 2013. "Q(ST)-F(ST) Comparisons: Evolutionary and Ecological Insights from Genomic Heterogeneity." *Nature Reviews. Genetics* 14 (March). Nature Publishing Group: 179–90. doi:10.1038/nrg3395.

Lepais, Olivier, and Jason Weir. 2016. "SimRAD: Simulations to Predict the Number of RAD and GBS Loci." https://cran.r-project.org/package=SimRAD.

Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." Genomics 0 (0): 3. http://arxiv.org/abs/1303.3997.

Lin, JZ, and K Ritland. 1997. "Quantitative Trait Loci Differentiating the Outbreeding Mimulus Guttatus from the Inbreeding M. Platycalyx." *Genetics* 146: 1115–21. http://www.genetics.org/content/146/3/1115.short.

Marais, G, B Charlesworth, and S I Wright. 2004. "Recombination and Base Composition: The Case of the Highly Self-Fertilizing Plant Arabidopsis Thaliana." *Genome Biology* 5 (7): R45. doi:10.1186/gb-2004-5-7-r45.

McCall, Andrew C., and Rebecca E. Irwin. 2006. "Florivory: The Intersection of Pollination and Herbivory." *Ecology Letters* 9 (12): 1351–65. doi:10.1111/j.1461-

0248.2006.00975.x.

McDonald, J Andrew, Debra R Hansen, Joshua R Mcdill, and Beryl B Simpson. 2011. "A Phylogenetic Assessment of Breeding Systems and Floral Morphology of North American Ipomoea (Convolvulaceae)." *Journal of the Botanical Research Institute of Texas* 5 (1): 159–77.

McVean, Gilean A T, and Brian Charlesworth. 2000. "The Effects of Hill-Robertson Interference between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation." *Genetics* 155 (2): 929–44.

Moeller, David a, and Monica a Geber. 2005. "Ecological Context of the Evolution of Self-Pollination in Clarkia Xantiana: Population Size, Plant Communities, and Reproductive Assurance." *Evolution; International Journal of Organic Evolution* 59 (4): 786–99. doi:10.1554/04-656.

Motten, A. F., and J. C. Stone. 2000. "Heritability of Stigma Position and the Effect of Stigma-Anther Separation on Outcrossing in a Predominantly Self-Fertilizing Weed, Datura Stramonium (Solanaceae)." *American Journal of Botany* 87 (3): 339–47. doi:10.2307/2656629.

Moyle, Leonie C., and Elaine B. Graham. 2006. "Genome-Wide Associations between Hybrid Sterility QTL and Marker Transmission Ratio Distortion." *Molecular Biology and Evolution* 23 (5): 973–80. doi:10.1093/molbev/msj112.

Muir, Christopher D, James B Pease, and Leonie C Moyle. 2014. "Quantitative Genetic Analysis Indicates Natural Selection on Leaf Phenotypes across Wild Tomato Species (Solanum Sect. Lycopersicon; Solanaceae)." *Genetics* 205 (4).

Ness, Rob W., Stephen I. Wright, and Spencer C Barrett. 2010. "Mating-System Variation, Demographic History and Patterns of Nucleotide Diversity in the Tristylous Plant Eichhornia Paniculata." *Genetics* 184 (2): 381–92. doi:10.1534/genetics.109.110130.

Ness, Rob W, Mathieu Siol, and Spencer C Barrett. 2012. "Genomic Consequences of Transitions from Cross- to Self-Fertilization on the Efficacy of Selection in Three Independently Derived Selfing Plants." *BMC Genomics* 13: 611. doi:10.1186/1471-2164-13-611.

Nicotra, Adrienne B., Andrea Leigh, C. Kevin Boyce, Cynthia S. Jones, Karl J. Niklas, Dana L. Royer, and Hirokazu Tsukaya. 2011. "The Evolution and Functional Significance of Leaf Shape in the Angiosperms." *Functional Plant Biology* 38 (7): 535–

52. doi:10.1071/FP11057.

Nielsen, Rasmus. 2005. "Molecular Signatures of Natural Selection." *Annual Review of Genetics* 39 (January): 197–218. doi:10.1146/annurev.genet.39.073003.112420.

Nordborg, Magnus. 2000. "Linkage Disequilibrium, Gene Trees and Selfing: An Ancestral Recombination Graph with Partial Self-Fertilization." *Genetics* 154 (2): 923–29.

Nordborg, Magnus, and Peter Donelly. 1997. "The Coalescent Process with Selfing." *Genetics* 146: 1185–95.

Ornduff, Robert. 1969. "Reproductive Biology in Relation to Systematics." *Taxon* 18 (2): 121–33.

Orr, HA. 1998. "Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data." *Genetics* 149: 2099–2104. http://www.genetics.org/content/149/4/2099.short.

Pannell JR, and B Charlesworth. 2000. "Effects of Metapopulation Processes on Measures of Genetic Diversity." *Philosophical Transactions of the Royal Society B-Biological Sciences* 355 (1404): 1851–64. doi:10.1098/rstb.2000.0740.

Parchman, Tom, Zach Gompert, and Alex Buerkle. 2011. "Amplified Restriction Fragments for Genomic Enrichment Oligonucleotide Sequences," no. August: 1–9.

Pettengill, James B., and David A. Moeller. 2012. "Tempo and Mode of Mating System Evolution between Incipient Clarkia Species." *Evolution* 66 (4): 1210–25. doi:10.1111/j.1558-5646.2011.01521.x.

Pollak, E. 1987. "On the Theory of Partially Inbreeding Finite Populations. I. Partial Selfing." *Genetics* 117 (2): 353–60.

Pollak, E., and M. Sabran. 1992. "On the Theory of Partially Inbreeding Finite Populations. III. Fixation Probabilities under Partial Selfing When Heterozygotes Are Intermediate in Viability." *Genetics* 131 (4): 979–85.

Qiu, Suo, Kai Zeng, Tanja Slotte, Stephen Wright, and Deborah Charlesworth. 2011. "Reduced Efficacy of Natural Selection on Codon Usage Bias in Selfing Arabidopsis and Capsella Species." *Genome Biology and Evolution* 3 (1): 868–80. doi:10.1093/gbe/evr085.

Rice, William E. R. 1989. "Analyzing Tables of Statistical Tests." *Evolution*. doi:10.2307/2409177.

Roels, Sarah a Bodbyl, and John K Kelly. 2011. "Rapid Evolution Caused by Pollinator Loss in Mimulus Guttatus." *Evolution; International Journal of Organic Evolution* 65 (9): 2541–52. doi:10.1111/j.1558-5646.2011.01326.x.

Roselius, Kerstin, Wolfgang Stephan, and Thomas Städler. 2005. "The Relationship of Nucleotide Polymorphism, Recombination Rate and Selection in Wild Tomato Species." *Genetics* 171 (2): 753–63. doi:10.1534/genetics.105.043877.

Sargent, Risa D, Carol Goodwillie, Susan Kalisz, and Richard H Ree. 2007. "Phylogenetic Evidence for a Flower Size and Number Trade-Of." *American Journal of Botany* 94 (12): 2059–62. doi:10.3732/ajb.94.12.2059.

SAS Instititute. 2017. "JMP Pro 13." Cary, NC.

Schafleitner, Roland, Luz R Tincopa, Omar Palomino, Genoveva Rossel, Ronald F Robles, Rocio Alagon, Carlos Rivera, et al. 2010. "A Sweet Potato Gene Index Established by de Novo Assembly of Pyrosequencing and Sanger Sequences and Mining for Gene-Based Microsatellite Markers." *BMC Genomics* 11 (1). BioMed Central Ltd: 604. doi:10.1186/1471-2164-11-604.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. "Fiji: An Open-Source Platform for Biological Image Analysis." *Nature Methods* 9 (7): 185. doi:10.1038/nmeth.2019.

Schueller, Sheila K. 2004. "Self-Pollination in Island and Mainland Populations of the Introduced Hummingbird-Pollinated Plant, Nicotiana Glauca (Solanaceae)." *American Journal of Botany* 91 (5): 672–81. doi:10.3732/ajb.91.5.672.

Sicard, Adrien, and Michael Lenhard. 2011. "The Selfing Syndrome: A Model for Studying the Genetic and Evolutionary Basis of Morphological Adaptation in Plants." *Annals of Botany* 107 (9): 1433–43. doi:10.1093/aob/mcr023.

Slotte, Tanja, Khaled M Hazzouri, J Arvid Ågren, Daniel Koenig, Florian Maumus, Ya-Long Guo, Kim Steige, et al. 2013. "The Capsella Rubella Genome and the Genomic Consequences of Rapid Mating System Evolution." *Nature Genetics* 45 (7). Nature Publishing Group: 831–35. doi:10.1038/ng.2669.

Slotte, Tanja, KM Hazzouri, and David Stern. 2012. "Genetic Architecture and Adaptive

Significance of the Selfing Syndrome in Capsella." *Evolution* 66 (5): 1360–74. doi:10.5061/dryad.f0gg64v5.

Smith, Stacey D. 2015. "Pleiotropy and the Evolution of Floral Integration." *New Phytologist* 209 (1): 80–85.

Snell, Rebecca, and Lonnie W Aarssen. 2005. "Life History Traits in Selfing versus Outcrossing Annuals: Exploring the 'Time-Limitation' Hypothesis for the Fitness Benefit of Self-Pollination." *BMC Ecology* 5: 2. doi:10.1186/1472-6785-5-2.

Stebbins, GL. 1957. "Self Fertilization and Population Variability in the Higher Plants." *American Naturalist* 91 (861): 337–54. http://www.jstor.org/stable/10.2307/2458946.

Stern, David L. 2013. "The Genetic Causes of Convergent Evolution." *Nature Reviews Genetics* 14 (11). Nature Publishing Group: 751–64. doi:10.1038/nrg3483.

"Sweetpotato Genomics Resource." 2017. *Michigan State University*. http://sweetpotato.plantbiology.msu.edu/.

Sweigart, Andrea L, and John H Willis. 2003. "Patterns of Nucleotide Diversity in Two Species of Mimulus Are Affected by Mating System and Asymmetric Introgression." *Evolution; International Journal of Organic Evolution* 57 (11): 2490–2506. doi:10.1554/02-726.

Takebayashi, N, D E Wolf, and L F Delph. 2006. "Effect of Variation in Herkogamy on Outcrossing within a Population of *Gilia Achilleifolia*." *Heredity* 96 (2): 159–65. doi:10.1038/sj.hdy.6800780.

Team, R Core. 2016. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org/.

USDA, and NRCS. 2017. "The PLANTS Database." *National Plant Data Team, Greensboro NC USA.*

Van der Auwera, Geraldine A., Mauricio O. Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. "From fastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." In *Current Protocols in Bioinformatics*, 1–33. doi:10.1002/0471250953.bi1110s43.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. "Pilon: An Integrated Tool

for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." *PLoS ONE* 9 (11). doi:10.1371/journal.pone.0112963.

Wessinger, CA, Lena C Hileman, and Mark D. Rausher. 2014. "Identification of Major Quantitative Trait Loci Underlying Floral Pollination Syndrome Divergence in Penstemon." *Proceedings of the Royal Society B.* 369 (1648): 20130349. http://rstb.royalsocietypublishing.org/content/369/1648/20130349.short.

Wilson, Peter J., Ken Thompson, and John G. Hodgson. 1999. "Specific Leaf Area and Leaf Dry Matter Content as Alternative Predictors of Plant Strategies." *New Phytologist* 143 (1): 155–62. doi:10.1046/j.1469-8137.1999.00427.x.

Wright, Stephen I, Beatrice Lauga, and Deborah Charlesworth. 2002. "Rates and Patterns of Molecular Evolution in Inbred and Outbred Arabidopsis." *Molecular Biology and Evolution* 19 (9): 1407–20. doi:10.1093/oxfordjournals.molbev.a004204.

Wu, Yonghui, Prasanna R. Bhat, Timothy J. Close, and Stefano Lonardi. 2008. "Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph." *PLoS Genetics* 4 (10). doi:10.1371/journal.pgen.1000212.

Xu, Shizhong. 2003. "Theoretical Basis of the Beavis Effect." *Genetics* 165 (4): 2259–68.

Yang, Sihai, Long Wang, Ju Huang, Xiaohui Zhang, Yang Yuan, Jian-Qun Chen, Laurence D Hurst, and Dacheng Tian. 2015. "Parent–progeny Sequencing Indicates Higher Mutation Rates in Heterozygotes." *Nature* 523: 463–67. doi:10.1038/nature14649.

Zeng, Kai, and Brian Charlesworth. 2009. "Estimating Selection Intensity on Synonymous Codon Usage in a Nonequilibrium Population." *Genetics* 183 (2): 651–62. doi:10.1534/genetics.109.101782.

Zhang, Bao, Chao Liu, Yaqin Wang, Xuan Yao, Fang Wang, Jiangsheng Wu, and Graham J King. 2015. "Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 Gene Converts Flower Colour from White to Yellow in Brassica Species." *New Phytologist* 206: 1513–26.

Biography

Joanna Lucy Rifkin was born in Malden, MA in 1987 and grew up in Cambridge, MA. She received a BA from Amherst College (Amherst, MA) in Biology and Classical Studies, graduating cum laude and with distinction in 2009. She entered the University Program in Genetics and Genomics at Duke University in 2011 and expects to complete her PhD in June 2017 and will begin postdoctoral research at the University of Toronto in July 2017.

Publications:

Rifkin, J.L., Castillo, A., Liao, I., and M.D. Rausher. (2017). Evolution of Reduced Genetic Variation and Relaxed Selection in the highly selfing species *Ipomoea lacunosa*: a Genomic Perspective. In review.

Rifkin, J. L., Nunn, C. L., and L.Z. Garamszegi. (2012). Do animals living in larger groups experience greater parasitism? A meta-analysis. The American Naturalist, 180(1), 70–82. doi:10.1086/666081

Temeles, E. J., Miller, J. S., & **J. L. Rifkin**. (2010). Evolution of sexual dimorphism in bill size and shape of hermit hummingbirds (Phaethornithinae): a role for ecological causation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 365(1543), 1053–63. doi:10.1098/rstb.2009.0284

Grants, awards and fellowships:

2016 Bass Instructor of Record Fellowship. The Graduate School, Duke University, Durham, NC, USA. $11,235

2015 Preparing Future Faculty Fellowship Program. The Graduate School, Duke University, Durham, NC, USA. $500

2015 Grant-in-Aid of Research. Department of Biology, Duke University, Durham, NC, USA. $1000

2015 Student Conference Travel Award. University Program in Genetics and Genomics, Duke University, Durham, NC, USA. $400

2015 Student Conference Travel Award. The Graduate School, Duke University, Durham, NC, USA. $700

2015 Student Conference Travel Award. Society for the Study of Evolution, St. Louis, MO, USA. $500

2015 Doctoral Dissertation Improvement Grant. "How is the Selfing Syndrome Assembled? A Common Garden Fitness Study of its Component Parts." National Science Foundation, Arlington, VA, USA. $20,565

2013 Graduate Research Fellowship Program. National Science Foundation, Arlington, VA, USA. $90,000

2013 Sigma-Xi grant-in-aid of research program. "The Role of Standing Genetic Variation in Selfing Rate." Sigma Xi, The Scientific Research Society Research Triangle Park NC, USA. $500

2011 James B. Duke Fellowship, Duke University. $20,000

2011 Chancellor's Scholarship, Duke University. $5,000

Professional Memberships

Society for the Study of Evolution

American Society of Naturalists

Department Service

Co-chair, Duke University Biology Department Population Biology Seminar 2014-2016

Member and co-chair, Duke University Program in Genetics and Genomics Retreat Committee, Biology Department Population Biology Seminar 2011-2013