UNDERSTANDING NATURAL EXPRESSION OF CYTOPLASMIC MALE STERILITY IN FLOWERING PLANTS USING A WILDFLOWER *LOBELIA SIPHILITICA* L. (CAMPANULACEAE)

A dissertation submitted

to Kent State University in partial

fulfillment of the requirements for the

degree of Doctor of Philosophy

by

Binaya Adhikari

August 2018

Dissertation written by

Binaya Adhikari

B. Sc., Tribhuvan University Nepal, 2000

M. Sc., Tribhuvan University Nepal, 2002

M. S., Mississippi State University, 2012

Ph. D., Kent State University, 2018

Approved by

Dr. Andrea L. Case, Chair, Doctoral Dissertation Committee

Dr. Oscar J. Rocha, Members, Doctoral Dissertation Committee

Dr. Christina M. Caruso

Dr. Srinivasan Vijayaraghavan

Dr. Soumitra Basu

Dr. John Portman

Accepted by

Dr. Laura G. Leff, Chair, Department of Biological Sciences

Dr. James L. Blank, Dean, College of Arts and Sciences

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES AND TABLES

LIST OF ABBREVIATIONS

ATP	Adenosine triphosphate
CMS	Cytoplasmic Male Sterility
dCTP	Deoxycytidine triphosphate
F_{ST}	Fixation Index
Indels	Insertions-Deletions
kb	Kilobase (1000 base pairs)
LD	Linkage Disequilibrium
MP	Maximum Parsimony
mRNA	Messenger Ribonucleic acid
N-P-K	Nitrogen-Phosphorus-Potassium
NADH	Nicotinamide adenine dinucleotide
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
SNP	Single Nucleotide Polymorphism
SSC	Saline Sodium Citrate
SSS	Substoichiometric Shifting
Ta	Annealing Temperature
uL	microliter (10^{-3} mL)
α-^{32}P-dCTP	Alpha phosphate group of dCTP labeled with phosphorus isotope 32

DEDICATION

To My Parents:

Mr. Bijaya Prasad Adhikari

Mrs. Sabitri Adhikari

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Andrea L. Case for providing me the opportunity to work in her lab, and for all the resources and guidance she provided me throughout the entire process of this research and writing. I would like to thank members of my dissertation committee Dr. Christina Caruso, Dr. Oscar Rocha, Dr. Srinivasan Vijayaraghavan, Dr. Soumitra Basu, and Dr. John Portman for their valuable comments on my dissertation. My sincere gratitude goes to Dr. Vijayaraghavan for kindly allowing me use his lab spaces and supplies, when I ran out of them, for Northern study, and Dr. Basu for letting me use phosphor imager.

I would like to thank Greg Wright and David Gilson at The Holden Arboretum, Kirtland Ohio, and Chris Rizzo, Melissa Davis, and staff members at the Herrick Conservatory, Kent State University for their support in growing my research plants. I am grateful to Hannah Madson for letting me use her plastid sequences and DNA samples. I am thankful to Aleecia Willis, Allison Johnson, Erica Robinson, Nicole Moscollic, and Victoria Willis for their help in growing and caring of plants and entering data. My thanks also go to Dr. Ayan Mukherjee, Rahul Bhattacharjee, and Suranjana Goswami for sharing their experience with Northerns. I would like to thank my lab mates Tony Miller and Thomas Mitchell for providing me feedback.

I would like to express my sincere gratitude to the Department of Biological Sciences, Kent State University for providing me facilities, a teaching assistantship as well as research and travel grants. I would like to thank Graduate Student Senate at Kent State University for providing me research and travel grants. I would like to thank tutors at library writing commons for their help.

I would like to express my deep love and thanks to my wife Bipana Pandit Adhikari for her continuous support during the entire period of my graduate studies. She has devoted so many

years for my study—praying for my success, boosting my morale, taking care of the family, and freeing me for my study. Thank you my lovely daughter Nayana Adhikari and son Nischaya Raj Adhikari for replenishing my energy when I am at home after long work-hours. You guys have compromised your fun times, including weekends, for your daddy's study!

Thank you, my parents Bijaya Prasad and Sabitri Adhikari, brothers Bipin and Bikal Adhikari and Shyam Raj Pandit, and my sisters Kalpana and Ambika Adhikari and Sapana Pandit for your blessings. Thank you, my parents-in-law Krishna Prasad and Radha Pandit, brothers Sheshmani Bhattarai, Murali Dhakal, Bharat Paudel, and Jiban Pandit, and sisters Kalpana Bhattarai, Sapana Dhakal, Tapana Paudel, and Menuka Pandit for your best wishes. Thanks to all the kids of our family for being my source of energy.

Binaya Adhikari

Sunday, July 23, 2018

Kent Ohio, USA

CHAPTER I. INTRODUCTION

Cytonuclear incompatibility is an important form of intraspecific incompatibility in eukaryotic organisms (Greiner *et al.* 2011; Burton *et al.* 2013; Crespi & Nosil 2013; Hill 2016). Genes encoded by mitochondrial and plastid genomes support essential metabolic functions, including cellular respiration and photosynthesis, respectively. Cytoplasmic gene products must often interact with nuclear products in order to function properly. Thus, they typically co-evolve to maintain compatibility. Cytonuclear interactions can be incompatible if one partner acquires a mutation without a compensatory change in the other, resulting in genetic mismatching with various deleterious consequences (e.g. Blier *et al.* 2001; Levin 2003; Paliwal *et al.* 2014). An interesting and well-studied example of cytonuclear incompatibilities in eukaryotes is cytoplasmic male sterility (CMS) found in flowering plants (Kaul 1988; Pelletier & Budar 2007).

Cytonuclear male sterility (CMS) in flowering plants

CMS genes are chimeric genes formed by physical rearrangements in mitochondrial genomes (Schnable & Wise 1998). Physical rearrangement is favored by the structural fluidity and highly repetitive content of flowering plant mitochondrial genomes, and rearrangement facilitates the formation of novel chimeric open reading frames or ORFs (Palmer & Herbon 1989; Kubo & Newton 2008; Darracq *et al.* 2010). Novel ORFs can gain function, usually with deleterious effects (Wolstenholme & Fauron 1995). Male sterility in angiosperms—a lack of

functional pollen—is one of the best-studied examples of deleterious effects of mitochondrial chimeric ORFs (Schnable & Wise 1998; Budar *et al.* 2003; Hanson & Bentolila 2004). The expression of male sterility depends on the cytonuclear interactions because CMS genes are usually accompanied by nuclear 'restorer of fertility' genes, which can specifically counter the action of a specific CMS gene or 'CMS type' (Hanson & Bentolila 2004).

CMS genes are widespread among flowering plants (Laser & Lerston 1972; Carlsson *et al.* 2008; Gobron *et al.* 2013), including crop plants (see reviews by Hanson & Bentolila 2004, Chen & Liu 2014). However, because compatible nuclear restorers suppress the action CMS genes in all flowering plant species, the male-sterile phenotype is not commonly expressed in nature. This is because plants carrying both a CMS gene and a compatible nuclear restorer are hermaphroditic, with normal, fertile pollen in addition to fertile ovules (Delannay *et al.* 1981; Richards 1997; Geber *et al.* 1999). High frequencies of both CMS and nuclear restorer alleles account for the fact that hermaphroditism is the predominant sexual condition in flowering plants (Bateman *et al*. 2011). Natural CMS–restorer incompatibility that results in a male sterile (or female) phenotype occurs rarely, in less than 1% of wild flowering plant species (Godin & Demyanova 2013). Because pollen is required for sexual reproduction, male sterile (female) and fertile (hermaphroditic) plants co-occur within natural populations of these species, forming a dimorphic sexual condition called 'gynodioecy' (Darwin 1987). Despite being a rare sexual condition in terms of the total number of flowering plants species, gynodioecy is found in about 21% of the total flowering plant families and has evolved numerous times independently in different flowering plant lineages (Caruso *et al.* 2016).

Significance of CMS–restorer systems in understanding cytonuclear incompatibility

Gynodioecy resulting from CMS–restorer incompatibility has received a significant

attention of biologists from three main perspectives. First, the evolution and interaction between mitochondrial CMS genes and nuclear restorers is considered as a classic example of cytonuclear 'genetic conflict' (Cosmides & Tooby 1981; Werren & Beukeboom 1998; Fishman & Willis 2006; Chase 2007; Chen & Liu 2014) and co-evolution (Frank 1989) in eukaryotes. The cause of cytonuclear conflict lies in differential patterns of inheritance between nuclear and cytoplasmic genomes: nuclear genomes are biparentally inherited while cytoplasmic genomes are uniparentally—largely maternally—inherited in flowering plants (Birky 2001). Thus, hermaphrodites gain their fitness through both seeds and pollen while females gain their fitness only through seeds. Consequently, CMS genes can spread through populations if females produce slightly more and/or better quality seeds (than hermaphrodites) by diverting resources otherwise used for pollen production to seed production or 'reproductive compensation' (Darwin 1877; Gouyon *et al.* 1991; Bailey *et al.* 2003; Shykoff *et al.* 2003; Dufay *et al.* 2007). Because CMS genes can spread despite their detrimental effect to the fitness (*via* pollen) of individuals carrying them, they are often referred to as 'selfish' genetic elements (Werren *et al.* 1988). The 'selfish' spread of a CMS gene is countered by the evolution of a compatible nuclear restorer allele that suppresses the action of CMS genes (Case *et al*. 2016), thus creating a 'conflict' between cytoplasmic (CMS) and nuclear (restorers) genes. The evolution of a compatible nuclear restorer in response to the invasion of a novel CMS type is often referred to as cytonuclear co-evolution and is compared to the evolutionary arms race between parasites and resistant genes in their hosts (Daugherty & Malik 2012).

Second, understanding how CMS–restorer incompatibility (or gynodioecy) arises and is maintained is important for understanding the evolution of diverse sexual systems in flowering plants (reviewed in Bawa & Beach 1981). Gynodioecy is considered by many researchers to be

an intermediate step in the evolution of separate sexes (dioecy) from hermaphroditism (Darwin 1877; Lloyd 1974; Charlesworth & Charlesworth 1978; Barrett 2002). In this pathway, female plants arise in an original hermaphroditic population, thus forming gynodioecy. Natural selection then favors the loss of female function in hermaphrodites to increase male reproductive success, ultimately forming male plants by genetic modifications (Charlesworth 1999; Barrett 2002; Barrett & Hough 2013).

Third, CMS–restorer incompatibility has been exploited as a useful tool in commercial crop improvement programs in several agricultural and horticultural plants (reviewed by Havey 2004; Chen & Liu 2014). Because CMS genes decoupled from compatible restorers can turn hermaphrodites into females, they serve as efficient and cost-effective measures of producing large quantities of hybrid seed (Tester & Langridge 2010). Hybrids are often desirable for agriculture because they combine traits of interest and often express increased vigor. Hybrid crops have contributed significantly in the increase of the world crop production (Tester & Langridge 2010), as more than half of the major crops (e.g., rice, wheat, maize, sorghum, rapeseed, sunflower, etc.) are hybrid varieties (Li *et al.* 2007). The most efficient way to generate hybrid seeds, particularly in wind-pollinated crops, is to plant males of one variety into fields with females of another. In this situation, all seed produced by females would be hybrid seed. Female parents required for hybridizations can be prepared by expressing appropriate CMS genes that make the plants male sterile without costly and tedious hand, mechanical, or chemical emasculation (the removal of functional pollen/anthers). If efficient crop production requires hybrids to produce pollen, male fertility can be restored in new hybrids by breeding in compatible nuclear restorers to counteract the effects of the CMS genes inherited from their female parents.

Current knowledge about CMS genes and nuclear restorers

Because of their commercial applications, CMS genes and their restorers have been studied extensively in crop plants. These studies have provided detailed understanding of genetic and molecular mechanisms of stamen and pollen development (see reviews in Kaul 1988; Schenable & Wise 1998; Hanson & Bentolila 2004; Chase 2007; Guo & Liu 2012). Distant intra- or inter-specific genetic crosses or somatic hybridizations (i.e., protoplast fusion) are required to decouple a CMS cytoplasm from its compatible nuclear restorer and uncover CMS genes in fixed hermaphroditic species. Once CMS lines are identified, candidate CMS genes are located and confirmed by comparing transcriptomes or proteomes derived from CMS cytoplasms in male-fertile (hermaphrodite) and male-sterile (female) individuals, with and without compatible restorer(s) respectively (e.g., Liu *et al*. 2007; Luo *et al*. 2013).

CMS genes are chimeric open reading frames (ORFs) containing partial sequences of one or more mitochondrial essential genes of electron transport chain, such as subunits of *atp*, *cox*, and *cob,* and additional mitochondrial sequences of unknown function (reviewed by Hanson & Bentolila 2004; Chen & Liu 2014). More than 28 CMS genes from 13 crops have so far been identified and characterized, collectively associated with at least 10 essential mitochondrial genes. In addition to known mitochondrial sequences, several CMS genes have been found to contain sequences of unknown origin (Hanson & Bentolila 2004; Chen & Liu 2014). Most CMS genes encode transmembrane proteins that are predicted to interfere with the respiratory pathway in either the gametophytic (microspores or pollen) or sporophytic (tapetal or meiocytes) anther tissues, thus causing damage or death of these tissues leading to pollen abortion or sterility (Guo & Liu 2012).

Multiple hypotheses have been proposed to explain why CMS genes only affect anther

tissues and not other parts of the plant. The first proposes that CMS transcripts and proteins only accumulate in anther tissues, although empirical evidence for such models is still lacking (Chen & Liu 2014). A second hypothesis proposes that CMS transcripts and proteins are expressed constitutively, but only anther tissues are affected because normal anther development demands much higher energy than what vegetative organs require (Warmke & Lee 1978). A third hypothesis proposes that CMS transcripts are expressed constitutively, but protein mainly accumulate in anther tissues (Abad *et al.* 1995). Empirical studies to test these models are scant.

Physiological mechanisms explaining how CMS gene products actually cause male sterility are also rarely tested. One potential explanation is that CMS proteins directly kill anther cells via mitochondrial dysfunction, leading to pollen abortion ("cytotoxicity" model, Levings III 1993). Another possibility is that CMS proteins cause mitochondrial deficiencies in anther tissues when there is tremendous energy demand (e.g., "energy deficiency" model, e.g., Lee & Warmke 1979; but see Touzet & Meyer 2014). A third potential mechanism proposes that CMS proteins cause premature or delayed death of tepetal cells (the innermost cell layer(s) of anthers). That is, while normal pollen development requires timely degeneration of tapetum, degeneration at a wrong time disrupts the process (e.g., "aberrant programmed cell death" model, Kawanabe *et al*. 2006; Ji *et al*. 2013).

Crop research has also revealed some information about the genetics and molecular mechanism of action of nuclear restorers. Restorers usually encode pentatricopeptide repeat (PPR) proteins, which are members of a large family of RNA-binding proteins in eukaryotes that regulate multiple post-transcriptional processes in organelles (Hanson & Bentolila 2004). Restorer gene products target mitochondrial transcriptomes and suppress the expression of CMS genes (Bentolila *et al.* 2002). At the genomic level, nuclear restorers could cause

substoichiometric shifting (SSS), a change in the stoichiometry of mitochondrial subgenomic molecule potentially containing a CMS gene to onset or end male sterility (e.g., Janska *et al*. 1998; Mackenzie & Chase 1990). At the post-transcriptional level, restorer gene products process CMS-associated transcripts by editing, polyadenylation, cleavage, or degradation, making CMS transcripts non-effective (Menassa *et al*. 1999; Moneger *et al*. 1994; Pring *et al*. 1998). Finally, fertility restoration in some plants may be controlled by translational, post-translational, or metabolic modifications (Chen & Liu 2014).

Models for the evolution and maintenance of gynodioecy in flowering plants

Despite recent progress in our understanding of the structure and molecular action of CMS genes and nuclear restorers via research in crop plants, evolutionary mechanisms that cause stable CMS–restorer incompatibility (or gynodioecy) in flowering plants are still not well understood. Genetic models propose that polymorphisms of CMS genes and nuclear restorers within populations are required for gynodioecy to persist. Predictions about the initial steps of CMS evolution are common to all genetic models. Specifically, a new CMS gene invades and spreads 'selfishly' through a population because it is maternally inherited and feminizing (Burt & Trivers 2006). Females, being obligatory outcrossers, could gain further fitness benefit by avoiding inbreeding depression that hermaphrodites may suffer when they inbreed (Thompson & Tarayre 2000; Delph 2004; Chang 2007). However, females suffer from pollen limitation when CMS genes (and thus females) become too frequent (McCauley & Taylor 1997; Frank & Barr 2001; Alonso 2005; Zhang *et al.* 2008). Thus, the high female frequency that results from the invasion and spread of a novel CMS gene should create strong selection for a compatible nuclear restorer, which is then expected to spread through the population, restoring male fertility. At this point, models of gynodioecy differ in their assumption of how CMS polymorphism is maintained

over the longer term.

According to a first type of models, nuclear restorers are expected to sweep to fixation where a target CMS gene is present because of the fitness benefits of producing pollen. If so, all plants in the population should eventually be restored hermaphrodites that carry a common CMS type (e.g., Case *et al.* 2016). Under these models, gynodioecy will only be observed when another CMS type invades, and only until another restorer sweep occurs and restores monomorphic hermaphroditism. Over time, multiple CMS types could periodically invade a population and replace existing CMS types in an "epidemic" fashion (Frank 1989; Couvet *et al.* 1998). But at any point in time, there should be few CMS types maintained within populations and gynodioecy should be transient (McCauley & Bailey 2009). According to a second type of models, a newly selected restorer would not necessarily spread to fixation. Instead, its spread would slow once it reached a reasonably high frequency (Charlesworth 1981; Gouyon *et al.* 1991; Bailey *et al.* 2003). This is expected to occur whenever hermaphrodites carrying excess or 'silent' restorers (that is, restorers not paired with their target CMS type) suffer a fitness cost (reviewed in Charlesworth 1981; Delph *et al.* 2007). In this "cost of restoration" scenario, restorers experience a type of balancing selection called negative frequency-dependent selection, and cycle through higher and lower-frequencies without going to fixation or being eliminated from the population (reviewed in Delph & Kelly 2014). Consequently, multiple CMS types and restorers can be maintained within a population at intermediate frequency over a long period of time. Finally, according to a recent 'mixed' model, a fertile (non-CMS) cytotype could be selectively maintained in a population for a long period of time while multiple CMS types competitively replace one another by partial selective sweeps (McCauley & Bailey 2009).

Distinguishing among alternative models of CMS–restorer polymorphism is difficult

because multiple CMS types could co-occur within a population at a given point in time under any of these models. One could, however, potentially distinguish among these models based on the time of maintenance of individual CMS types within a population. Balancing selection associated with a cost of restoration predicts long-term maintenance of individual CMS types. Thus, CMS types are expected to be relatively older than the CMS types in populations experiencing frequent replacement (epidemic dynamics). CMS types are also expected to be younger if they are formed *de novo,* by mitochondrial recombination (Hanson & Bentolila 2004).

Inferring relative times of maintenance of CMS types is challenging because the genes themselves have not been identified in species with natural CMS–restorer incompatibility. Also, because CMS genes are chimeric, containing sequences of multiple known or unknown sources, inferring relative times of origin using a 'molecular clock' is challenging (Zuckerkandl & Pauling 1965; Bromham & Penny 2003; Kumar 2005). Therefore, an alternative strategy to infer relative time of occurrence of a CMS type is to study their geographic distribution. That is, CMS types that are widespread are expected to be relatively older then those that are restricted in distribution unless gene flow is limited.

Lobelia siphilitica L. (Campanulaceae)

The goal of my doctoral research was to evaluate the mechanisms maintaining natural gynodioecy in flowering plants using a wildflower *Lobelia siphilitica* L. (Campanulaceae) as a model (Figure 1.1). *Lobelia siphilitica* is a wildflower native to eastern North America and mainly grows in wet habitats (Figure 1.1.A). Flowers are ca. 3 cm long; blue to purple (Figure 1.1.B). Anthers and filaments fuse to form a tube that covers pistil in perfect flowers at anthesis.

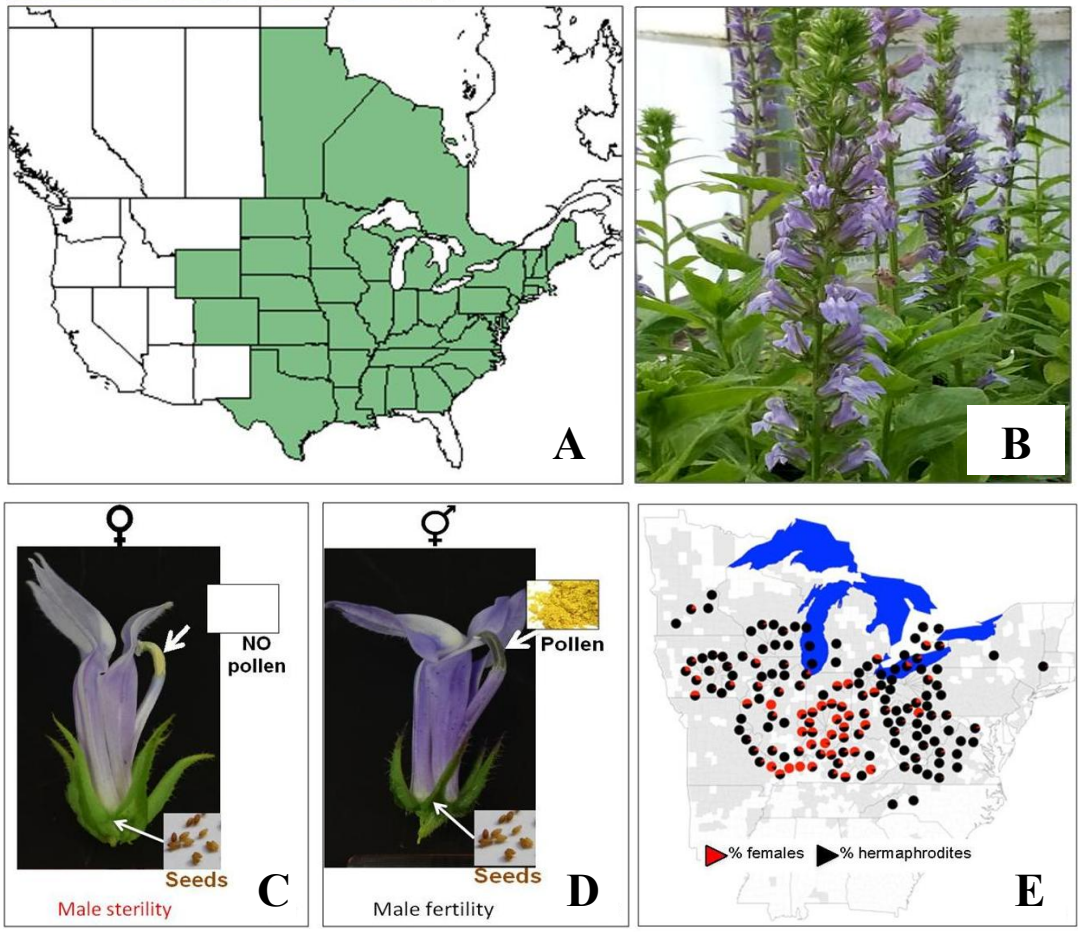


Figure 1.1. *Lobelia siphilitica* L.: (A) state-level geographic distribution of the species in eastern North America (USDA, NRCS 2018); (B) an inflorescence; (C) a male-sterile flower on a female plant; (D) a male-fertile flower on a hermaphrodite; and (E) county-level distribution (gray shading) with population sex ratios (pie charts) based on field surveys (Caruso & Case 2007; Madson 2012; Case AL & Caruso CM, unpublished data).

Anthers in pistillate (female) flowers are thin and whitish with no pollen (Figure 1.1.B) while those in perfect (hermaphroditic) flowers are dark-purple with viable pollen (Figure 1.1.C). The species is self-compatible (Johnston 1992). Plants flower in wild populations from early August to early October (Caruso CM & Case AL, personal communication) (Figure 1.1.D).

For multiple reasons, *L. siphilitica* is a suitable model for understanding the maintenance of gynodioecy in flowering plants. The species is relatively easy to work with because their large flowers and inflorescences facilitate hand-pollination and sex identification. Each fruit typically produces hundreds of seeds such that sufficient progeny can be produced for the estimation of offspring sex ratios. In addition, perfect flowers are protandrous, with completely separate male and female phases, reducing self-pollination. Moreover, the plants can be grown easily in greenhouse conditions. Furthermore, the plants undergo a dormant stage during winter and can grow clonally in the following summer (Beaudoin Yetter 1989), providing opportunities for follow-up studies using specific genotypes. Most importantly for the study of gynodioecy, the species has highly variable proportions of females in natural populations, ranging from 0–100%, making it an ideal system to understand the causes of such variation (Figure 1.1.E). Finally, numerous previous studies in *L. siphilitica* provide useful background information pertaining natural gynodioecy in this species (e.g., Dudle *et al.* 2001; Bailey 2002; Caruso & Case 2007; Case & Caruso 2010; Hovatter *et al.* 2011, 2013; Madson 2012; Caruso & Case 2013; Delph & Montgomery 2014; Caruso *et al.* 2015; Rivkin *et al.* 2015; Bailey *et al.* 2017; Eisen *et al.* 2017). This makes deeper investigation of mechanisms of gynodioecy easier in this system.

Scope and organization of this dissertation

In this dissertation, I first tested whether long-term maintenance (balancing selection), frequent replacement (epidemic dynamics), or a scenario similar to the 'mixed' model may result

in CMS polymorphism in *L. siphilitica* by using plastid and mitochondrial genetic markers as well as crossing experiments. In addition, I evaluated the patterns of evolution and diversity of mitochondrial genomes (which carry CMS genes) with respect to plastid genomes (which do not carry CMS but are typically co-inherited with mitochondria). Moreover, I tested potential causes of highly variable frequencies of females in natural populations of this species. Finally, I aimed to directly identify and characterize CMS genes of this species at the molecular level.

Chapter II: In Chapter II, I used patterns of diversity and distribution of cytoplasmic markers to evaluate alternate mechanisms maintaining gynodioecy in *L. siphilitica*. In the absence of information about actual CMS types, cytoplasmic markers are expected to serve as close proxies of CMS types (e.g., Ingvarsson & Taylor 2002; Stadler & Delph 2002; Houliston & Olson 2006; Touzet & Delph 2009; Lahiani *et al.* 2013; Delph & Montgomery 2014) assuming that cytoplasmic genomes are usually co-inherited (uniparentally) as coherent non-recombining units (e.g., Palmer *et al.* 2000). This chapter had three specific aims.

First, I examined the patterns of distribution of mitotypes to understand relative ages or times of occurrences of CMS types in this species. I predicted that older mitotypes (thus CMS types) should have a wide geographic distribution because they have sufficient time to migrate via dispersal, unless gene flow is particularly limited in parts of the species range. By contrast, newer mitotypes should always have a restricted distribution. A population could have newer CMS types if they were formed *de novo* by mitochondrial recombination (Hanson & Bentolila 2004).

The second specific aim of Chapter II was to test whether populations of *L. siphilitica,* especially those with high female frequencies, had higher mitotype diversities as expected if such populations had higher numbers of unique CMS types (e.g., Delph & Kelly 2014). Alternatively,

frequent formation of novel CMS types (e.g., Hanson & Bentolila 2004) could result in high female frequencies because compatible restorers for new CMS types could be rare or absent.

Finally, I compared the patterns of genetic variations between two cytoplasmic genomes—plastid and mitochondrial—in *L. siphilitica.* Two cytoplasmic genomes in flowering plants are typically maternally co-inherited and clonally replicated (e.g., Palmer 1987; Milligan 1992; Rebound & Zeyl 1994; Dumolin-Lapegue *et al*. 1998; Birky 2001). Thus, they are expected to show similar patterns of diversity or strong linkage disequilibrium (LD) (Maynard Smith & Haigh 1974; Schnabel & Asmussen 1989). However, in certain species, such as gynodioecious flowering plants, the rules of strict maternal co-inheritance and clonal replication could be violated, potentially leading to discordant patterns of diversity or weak LD (reviewed by McCauley 2013). I tested this prediction in *L. siphilitica*.

Chapter III: In Chapter III, I used a large crossing experiment to assess the diversity and distribution of CMS types within and among populations of *L. siphilitica* and used this information to understand potential causes of high female frequencies in natural populations of this species. In crossing experiments, predictions about how many and which CMS types (and/or nuclear restorers) are present in parental plants are made based on progeny sex ratios (e.g., Koelewijn & van Damme 1995; de Haan *et al.* 1997b; Charlesworth & Laporte 1998; Dudle *et al.* 2001; Bailey 2002; van Damme *et al.* 2004; Damme *et al.* 2004; Dufay *et al.* 2009; Garraud *et al.* 2011). This chapter had four specific aims.

First, I assessed the diversity of CMS genes within and across populations of *L. siphilitica*. Theoretical models of gynodioecy assume fewer CMS types at the population level—two CMS types (Gouyon *et al.* 1991; Bailey *et al.* 2003) or one CMS type and a fertile cytotype (Jacobs & Wade 2003; Dufay *et al.* 2007) for computational efficiency. However, previous

crossing experiments in gynodioecious species have found up to four CMS types within species (de Haan *et al.* 1997b; van Damme *et al.* 2004) and up to three at the population level (Damme *et al.* 2004; Dufay *et al.* 2009). In *L. siphilitica,* three CMS types were found at the species level and two within a high female population (Dudle *et al.* 2001; Bailey 2002). In Chapter II, I found very high mitotype diversity within *L. siphilitica* (39 mitotypes) and its populations (up to six mitotypes). Thus, I predicted that diversity of CMS genes could also be higher in this species.

Second, I tested whether the mitotypes obtained in Chapter II correspond to CMS types identified by crossing experiments. Although I did not expect each individual mitotype to correspond to each specific CMS type, I predicted that a mitotype or a particular group of mitotypes would correspond to a unique CMS type. Understanding such correspondence would help validate conclusions drawn from the mitotype data (Chapter II) and increase the utility of future marker-based studies.

Third, using the same predictions as in Chapter II, I used the distribution of CMS types inferred from crossing studies to assess whether long-term maintenance (balancing selection), frequent replacement (epidemic dynamics) and/or frequent formation of novel CMS types has maintained CMS polymorphism in this species. I used the patterns of restoration to infer the distribution of a CMS type. Because a restorer is expected to evolve in response to a CMS type (cytonuclear co-evolution), I assumed that if pollen from a population could restore a CMS type from a different population, the siring (pollen) population is likely to have the CMS type.

Finally, because high female populations of *L. siphilitica* also contained higher mitotype diversities (Chapter II; Delph & Montgomery 2014), I tested whether such populations also have higher CMS diversity (Delph & Kelly 2014). If there is a high CMS diversity, compatible restorers for each CMS type may not be present in the population or could be rare, resulting in

high female frequency. High female frequencies could also be caused by complex restoration genetics, involving multiple restorer loci and their interactions (e.g., 'quantitative restoration,' Ehlers *et al.* 2005; 'threshold restoration,' Bailey & Delph 2007). Complex restoration genetics has recently been reported by multiple crossing studies (e.g., Koelewijn & van Damme 1995; Charlesworth & Laporte 1998; Garraud *et al.* 2011).

Chapter IV: In the first two projects (Chapters II and III), I asked several important questions concerning the genetic mechanisms of natural gynodioecy in *L. siphilitica* and effects of this condition on the patterns of cytoplasmic genetic diversity. In the third project (Chapter IV), I searched for actual CMS genes in *L. siphilitica.* The major hurdles in the discovery of CMS genes include their complex structure, uncertain position in the mitochondrial genome, and multiple possible ways by which the nuclear restorers suppress CMS expression (Chen & Liu 2014).

I used the basic features shared by CMS genes in crop plants (reviewed by Hanson & Bentolila 2004; Chen & Liu 2014) to search for CMS-specific mitochondrial transcripts in *L. siphilitica* using Northern hybridization assays (following Case & Willis 2008). This method assumes that CMS genes are co-transcribed with one of the essential mitochondrial genes, usually one of the subunits of genes involved in respiratory pathways (Hanson & Bentolila 2004). Thus, potential CMS-associated transcripts could be identified by comparing mitochondrial transcripts found in females—that carry expressed CMS genes—but not in their hermaphrodite full sibs—that should carry the same CMS gene but do not express it (Case & Willis 2008). I used maternal families that were inferred to have different unique CMS types from the crossing experiment (Chapter III) for this study.

Overall, my doctoral research used multiple approaches to understand mechanisms of

CMS–restorer incompatibility (or natural gynodioecy) in flowering plants using *Lobelia siphilitica* as a model. Although actual CMS genes have not been identified or characterized at molecular level in this species, cytoplasmic markers and crossing experiments allowed me to document CMS polymorphism within and across populations and to compare against expected patterns under various theoretical models: (1) long-term maintenance of CMS types (consistent with balancing selection), (2) frequent replacement of CMS types (consistent with epidemic dynamics), and (3) frequent formation of novel CMS types. In addition, I used patterns of genetic diversity in plastid and mitochondrial markers to estimate the level of association (LD) between two cytoplasmic genomes in this species and to understand potential mechanisms that could determine the level of cytoplasmic LD. Finally, I attempted to directly identify and characterize actual CMS genes in *L. siphilitica* by analyzing mitochondrial transcripts of male sterile (female) and fertile (hermaphrodite) plants. My study represents one of the most robust studies to date using cytoplasmic markers and crossing experiment to evaluate mechanisms that maintain gynodioecy in flowering plants.

REFERENCES

Abad AR, Mehrtens BJ, Mackenzie SA (1995) Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell* 7:271–85

Alonso C, Herrera CM (2008) Site-specific features affect pollination success of a gynodioecious understory shrub in a gender-specific mode. *Ecoscience* 14: 358–365

Bailey MF (2002) A cost of restoration of male fertility in a gynodioecious species, *Lobelia*

siphilitica. *Evolution* 56: 2178–2186

Bailey MF, Case AL, Caruso CM (2017) Physiological effects of temperature do not explain prevalence of females in populations of gynodioecious *Lobelia siphilitica* growing in warmer climates. *American Journal of Botany* 104 (3): 411–418

Bailey MF, Delph LF, Lively CM (2003) Modeling gynodioecy: novel scenarios for maintaining polymorphism. *American Naturalist* 161: 762–776

Barrett SCH, Hough J (2013) Sexual dimorphism in flowering plants. *Journal of Experimental Botany* 64(1): 67–82

Bateman RM, Hilton J, Rudall P (2011) Spatial Separation and Developmental Divergence of Male and Female Reproductive Units in Gymnosperms, and Their Relevance to the Origin of the Angiosperm Flower Pg. 8–48 *In:* Wanntorp L, DeCraene LPR (eds.) *Flowers on the Tree of Life*. Cambridge University Press, UK

Bawa K, Beach J (1981) Evolution of Sexual Systems in Flowering Plants. *Annals of the Missouri Botanical Garden* 68(2): 254–274

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 99(16): 10887–92

Birky CW Jr. (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35: 125–148

Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: Evidence for intergenomic co-adaptation. *Trends Genet* 17(7): 400–406

Bromham L, Penny D (2003) The modern molecular clock. *Nat Rev Genet* 4: 216–224

Budar F, Roux F (2011) The role of organelle genomes in plant adaptation. *Plant Signaling & Behavior* 6(5): 635–639

Burt A, Trivers R (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, MA

Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics* 44: 281–302

Carlsson J, Leino M, Sohlberg J, Sundström JF, Glimelius K (2008) Mitochondrial regulation of flower development. *Mitochondrion* 8(1): 74–86

Caruso CM, Benscoter AM, Gale NV, Seifert EK, Mills ER, Case AL (2015) Effects of crossing distance on performance of the native wildflower *Lobelia siphilitica*: Implications for ecological restoration. *Journal of the Torrey Botanical Society* 142 (2): 140–151

Caruso CM, Case AL (2007) Sex ratio variation in gynodioecious *Lobelia siphilitica*: effects of population size and geographic location. *J Evol Biol* 20: 1396–1405

Caruso CM, Eisen K, Case AL (2016) An angiosperm-wide analysis of the correlates of gynodioecy. *International Journal of Plant Sciences* 177: 115–121

Case AL, Caruso CM (2010) A novel approach to estimating the cost of male fertility restoration in gynodioecious plants. *New Phytologist* 186: 549–557

Case AL, Caruso CM (2013) Testing models of sex-ratio evolution in a gynodioecious plant: female frequency co-varies with the cost of male fertility restoration. *Evolution* 67: 561–566

Case AL, Finseth FR, Barr CM, Fishman L (2016) Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proc R Soc B* 283: 20161493

Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* 62: 1026–1039

Chang SM (2007) Gender-specific inbreeding depression in a gynodioecious plant, *Geranium maculatum* (Geraniaceae). *American Journal of Botany* 94: 1193–1204

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46: 27–39

Charlesworth D (1999) Theories on the evolution of dioecy, Pg. 33–60 *In*: Geber MA, Dawson TE, Delph LF (eds.). *Gender and sexual dimorphism in flowering plants*. Berlin: Springer-Verlag

Charlesworth D, Charlesworth B (1978) A model for the evolution of dioecy and gynodioecy. *Am Nat* 112(4): 975–997

Charlesworth D, Laporte V (1998) The Male-Sterility Polymorphism of *Silene vulgaris*: Analysis of Genetic Data From Two Populations and Comparison With *Thymus vulgaris*. *Genetics* 150(3): 1267–1282

Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet* 23: 81–90

Chen L, Liu Y-G (2014) Male Sterility and Fertility Restoration in Crops. *Annual Review of Plant Biology* 65(1): 579–606

Cosmides LM, Tooby J (1981) Cytoplasmic inheritance and intragenomic conflict. *J Theor Biol.* 89: 83–129

Couvet D, Ronce O, Gliddon C (1998) Maintenance of nucleo-cytoplasmic polymorphism in a metapopulation: the case of gynodioecy. *American Naturalist* 152: 59–70

Crespi B, Nosil P (2013) Conflictual speciation: Species formation via genomic conflict. *Trends in Ecology & Evolution* 28: 48–57

Darracq A, Varre JS, Touzet P (2010) A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics* 11: 233

Darwin C (1877) *The different forms of flowers on plants of the same species*. J Murray, London

Daugherty M D, Malik H S (2012) Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46: 677–700

Delannay X, Gouyon PH, Valdeyron G (1981) Mathematical study of the evolution of gynodioecy with cytoplasmic inheritance under the effect of a nuclear restorer gene. *Genetics* 99 (1): 169–181

Delph LF (2004) Testing for sex differences in biparental inbreeding and its consequences in a gynodioecious species. *American Journal of Botany* 91: 45–51

Delph LF, Kelly JK (2014) On the importance of balancing selection in plants. *New Phytologist* 201: 45–56

Delph LF, Montgomery BR (2014) The evolutionary dynamics of gynodioecy in *Lobelia*. *International Journal of Plant Sciences* 175: 383–391

Delph LF, Touzet P, Bailey MF (2007) Merging theory and mechanism in studies of gynodioecy. *Trends Ecol Evol* 22(1): 17–24

Dudle DA, Mutikainen P, Delph LF (2001) Genetics of sex determination in the gynodioecious species *Lobelia siphilitica*: Evidence from two populations. *Heredity (Edinb)* 86: 265–276

Dufay M, Cuguen J, Arnaud JF, Touzet P (2009) Sex ratio variation among gynodioecious populations of sea beet: can it be explained by negative frequency‐dependent selection? *Evolution* 63: 1483–1497

Dufaÿ M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99: 349–356

Ehlers BK, Maurice S, Bataillon T (2005) Sex inheritance in gynodioecious species: a polygenic view. *Proceedings of the Royal Society B: Biological Sciences* 272(1574): 1795–1802

Eisen K, Case AL, Caruso CM (2017) Variation in Pollen-Dispensing Schedules of *Lobelia siphilitica*. *International Journal of Plant Sciences* 178 (1): 79–84

Fishman L and Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* 60: 1372–1381

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist* 133: 345–376

Garraud C, Brachi B, Dufaÿ M, Touzet P, Shykoff J (2011) Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* 106: 757–764

Geber MA, Dawson TE, Delph LF (eds.) (1999). *Gender and sexual dimorphism in flowering plants*. Springer-Verlag, New York

Gobron N, Waszczak C, Simon M, Hiard S, Boivin S, Charif D, *et al.* (2013) A Cryptic Cytoplasmic Male Sterility Unveils a Possible Gynodioecious Past for *Arabidopsis thaliana*. *PLoS ONE* 8(4): e62450

Godin VM, EI Demyanova (2013) On the distribution of gynodioecy in the flowering plants.

Bot J 93: 1465–1487

Gouyon PH, Vichot F, Van Damme JMM (1991) Nuclear-cytoplasmic male sterility: single point equilibria versus limit cycles. *The American Naturalist* 137: 498–514

Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. *Molecular Ecology* 20: 671–691

Guo JX, Liu YG (2012) Molecular control of male reproductive development and pollen fertility in rice. *J. Integr. Plant Biol.* 54: 967–78

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16 Suppl 1: S154–69

Havey MJ (2004) The Use of Cytoplasmic Male Sterility for Hybrid Seed Production. *In*: Daniell H, Chase C (eds.) *Molecular Biology and Biotechnology of Plant Organelles*. Springer, Dordrecht

Hill GE (2016) Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* 6: 5831–5842

Houliston GJ, Olson MS (2006) Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics* 174: 1983–1994

Hovatter SR, Blackwood CB, Case AL (2013) Conspecific plant-soil feedback scales with population size in *Lobelia siphilitica* (Lobeliaceae). *Oecologia* 173(4): 1295–307

Hovatter SR, Dejelo C, Case AL, Blackwood CB (2011) Metacommunity organization of soil microorganisms depends on habitat type defined by presence of *Lobelia siphilitica* plants. *Ecology* 92: 57–65

Ingvarsson PK, Taylor DR (2002) Genealogical evidence for epidemics of selfish genes. *Proceedings of the National Academy of Sciences, USA* 99: 11265–11269

Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA (1998) Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* 10: 1163–80

Ji C, Li H, Chen L, Xie M, Wang F, *et al.* (2013) A novel rice *bHLH* transcription factor, *DTD*, acts coordinately with *TDR* in controlling tapetum function and pollen development. *Mol. Plant* 6: 1715–18

Kaul MLH (1988) *Male Sterility in Higher Plants: Monographs on Theoretical and Applied Genetics 10*. Berlin: Springer Verlag

Kawanabe T, Ariizumi T, Kawai-Yamada M, Uchimiya H, Toriyama K (2006) Abolition of the tapetum suicide program ruins microsporogenesis. *Plant Cell Physiol.* 47: 784–87

Koelewijn HP and Van Damme JM (1995) Genetics of male sterility in gynodioecious *Plantago coronopus*. I. Cytoplasmic variation. *GENETICS* 139(4): 1749–1758

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8: 5–14

Kumar S (2005) Molecular clocks: Four decades of evolution. *Nat Rev Genet* 6: 654–662.

Lahiani E, Dufaÿ M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P (2013) Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity (Edinb)* 111: 157–64

Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male-sterile angiosperms. *Bot Rev* 38: 425–454

Lee SJ, Warmke HE (1979) Organelle size and number in fertile and T-cytoplasmic male-sterile corn. *Am. J. Bot.* 66: 141–48

Levin DA (2003) The cytoplasmic factor in plant speciation. *Syst Bot* 28: 5–11

Levings III CS (1993) Thoughts on cytoplasmic male sterility in cms-T maize. *Plant Cell* 5: 1285

Li S, Yang D, Zhu Y (2007) Characterization and use of male sterility in hybrid rice breeding. *J. Integr. Plant Biol.* 49: 791–804

Liu ZL, XuH, GuoJX, Liu YG (2007) Structural and expressional variations of the mitochondrial genome conferring the wild abortive type of cytoplasmic male sterility in rice. *J. Integr. Plant Biol.* 49: 908–14

Lloyd DG (1974) Theoretical sex ratios of dioecious and gynodioecious angiosperms. *Heredity* 32: 11–34

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1506): 3009–3021

Luo D, Xu H, Liu Z, Guo J, Li H, *et al*. (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45: 573–77

Mackenzie SA, Chase CD (1990) Fertility restoration is associated with loss of a portion of the mitochondrial genome in cytoplasmic male-sterile common bean. *Plant Cell* 2: 905–12

Madson HJ (2012) The effects of stochastic processes on sex-ratio variation in gynodioecious *Lobelia siphilitica* L. (Lobeliaceae). MS Thesis submitted to Kent State University, Kent OH, USA

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* 104(4): 611–620

Menassa R, L Homme Y, Brown GG (1999) Post-transcriptional and developmental regulation of a CMS-associated mitochondrial gene region by a nuclear restorer gene. *Plant J.* 17: 491–99

Moneger F, Smart CJ, Leaver CJ (1994) Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. *EMBO J.* 13: 8–17

Paliwal S, Fiumera AC, Fiumera HL (2014) Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* 198(3): 1251–1265

Palmer JD, Herbon LA (1989) Plant Mitochondrial-DNA Evolves Rapidly In Structure, But Slowly In Sequence. *Journal of Molecular Evolution* 28: 87–97

Pelletier G, Budar F (2007) The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. *Curr Opin Biotechnol* 18(2): 121–125

Pring DR, Chen W, Tang HV, Howad W, Kempken F (1998) Interaction ofmitochondrial RNA editing and nucleolytic processing in the restoration of male fertility in sorghum. *Curr. Genet.* 33: 429–36

Ramsey J, Bradshaw HD, Jr, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57: 1520–1534

Richards AJ (1997) *Plant Breeding Systems*. Chapman and Hall: London

Rivkin LR, Case AL, Caruso CM (2015) Frequency-dependent fitness in gynodioecious *Lobelia siphilitica*. *Evolution* 69(5): 1232–1243

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N (2016) Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biology* 14(12): e2000234

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci* 3: 175–180

Shykoff JA, Kolokotronis SO, Collin CL, Lopez-Villavicencio M (2003) Effects of male sterility on reproductive traits in gynodioecious plants: a meta-analysis. *Oecologia* 135: 1–9

Skrede I, Brochmann C, Borgen L, Rieseberg LH (2008) Genetics of intrinsic postzygotic isolation in a circumpolar plant species, *Draba nivalis* (Brassicaceae). *Evolution* 62: 1840–1851

Stadler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences, USA* 99: 11730–11735

Sweigart AL, Willis JH (2012) Molecular evolution and genetics of postzygotic reproductive isolation in plants. *F1000 Biology Reports* 4: 23

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. *Science* 327: 818–22

Thompson JD, Tarayre M (2000) Exploring the genetic basis and proximate causes of female fertility advantage in gynodioecious *Thymus vulgaris*. *Evolution* 54: 1510–1520

Touzet P, Delph LF (2009) The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics* 181: 631–644

Touzet P, Meyer EH (2014) Cytoplasmic male sterility and mitochondrial metabolism in plants. *Mitochondrion* 19: 166–71

USDA, NRCS (2018) The PLANTS Database (http://plants.usda.gov, 26 July 2018). National Plant Data Team, Greensboro, NC 27401-4901 USA

van Damme JMM, Hundscheid MPJ, Ivanovic S, Koelewijn HP (2004) Multiple CMS–restorer gene polymorphism in gynodioecious *Plantago coronopus*. *Heredity* 93: 175–181

Warmke HE, Lee SL (1978) Pollen abortion in T cytoplasmic male-sterile corn (*Zea mays*): a suggested mechanism. *Science* 200: 561–63

Werren JH (1998) Wolbachia and speciation Pg. 245-260 *In*: *Endless Forms: Species and Speciation*. Oxford Univ. Press; New York

Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10863–10870

Werren JH, Nur U, Wu CI (1988) Selfish genetic elements. *Trends Ecol Evol.* 3(11): 297–302

Wolstenholme DR, and Fauron CM-R (1995) Mitochondrial genome organization Pg. 1–60 *In*: CS LevingsandI Vasil K (eds.) *The molecular biology of plant mitochondria*. Kluwer Academic Publishers, Boston, MA

Zhang YW, Wang Y, Yu Q, Zhao JM (2008) Sex expression, female frequency, and reproductive output in a gynodioecious clonal herb, *Glechoma longituba* (Lamiaceae). *Plant Ecol* 199: 255–264

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins Pg. 97–166. *In*: Bryson V, Vogel HJ (eds.) *Evolving Genes and Proteins*. Academic Press

CHAPTER II. CYTOPLASMIC DISCORDANCE IS ASSOCIATED WITH SEX-RATIO VARIATION IN GYNODIOECIOUS *LOBELIA SIPHILITICA* L. (CAMPANULACEAE)

ABSTRACT

Plastid and mitochondrial genomes in angiosperms are expected to show strong linkage disequilibrium (LD) because they are typically maternally co-inherited and clonally replicated. However, cytoplasmic LD could weaken in species in which one or both of the cytogenomes undergo occasional biparental inheritance and/or recombine frequently, such as gynodioecious flowering plants. In these species, cytoplasmic (mitochondrial) male sterility (CMS) genes disrupt pollen production, making some individuals female. Theoretical models propose that stable gynodioecy results from the polymorphism of CMS genes within a population. The co-occurring mitotypes (associated with multiple CMS types) may increase the chances of heteroplasmy—the co-occurrence of multiple mitotypes within an individual—when there is biparental inheritance. This provides opportunities for homologous recombination producing novel mitotypes independent of plastid haplotypes. This may weaken LD between cytoplasmic genomes. Thus, populations with more females, which are expected to have higher CMS diversity (and more mitotypes), may contribute to the breakdown of cytoplasmic LD. I used plastid and mitochondrial marker sequences from 266 individuals of gynodioecious *Lobelia siphilitica* representing 61 populations varying widely in female frequencies to estimate

cytoplasmic LD and to evaluate potential mechanisms affecting cytoplasmic LD. I found evidence of weak cytoplasmic LD overall, high mitochondrial recombination, and positive correlations between mitotype diversity, mitotype rarity, and female frequency across populations. No such correlations were found with plastid marker. My data suggest that weak cytoplasmic LD results from processes that likely create and maintain multiple CMS types, especially in high-female populations.

Keywords: cytoplasmic LD, CMS, *Lobelia siphilitica,* gynodioecy

INTRODUCTION

Plastid and mitochondrial genetic variation in angiosperms is expected to show strong linkage disequilibrium, LD (Maynard Smith & Haigh 1974; Schnabel & Asmussen 1989) because cytogenomes are typically maternally co-inherited and clonally replicated (Palmer 1987; Milligan 1992; Rebound & Zeyl 1994; Dumolin-Lapegue *et al*. 1998; Birky 2001). Only a few studies have evaluated the level of plastid–mitochondrial LD in angiosperms, and those studies report contrasting patterns among species (strong LD: *Beta vulgaris* ssp. *maritima,* Desplanque *et al.* 2000; *Silene vulgaris,* Olson & McCauley 2000; weaker LD: *Silene vulgaris,* Storchova & Olson 2004; Houliston & Olson 2006; *Quercus* spp., Dumolin-Lapegue *et al.* 1998; *S. nutans*, Lahiani *et al.* 2013; *Triticum* spp., Tsujimura *et al.* 2013; *Fragaria* spp., Govindarajulu *et al.* 2015), and even for the same species (e.g., *S. vulgaris*). Inconsistent patterns of cytoplasmic LD found by these studies raise interesting questions regarding whether or not the processes weakening cytoplasmic LD vary among species or among populations within species.

Although direct tests of plastid–mitochondrial LD are limited, there is burgeoning evidence that angiosperm cytogenomes can violate the rules of maternal co-inheritance and clonal replication (Corriveau & Coleman 1988; Smith 1989; Barr *et al.* 2005; Greiner *et al.* 2014; Breton & Stewart 2015). Biparental inheritance (or 'paternal leakage') could potentially weaken plastid–mitochondrial LD if it leads to heteroplasmy, the co-occurrence of different plastid and/or mitochondrial variants within an individual (McCauley *et al.* 2005; Welch *et al.* 2006; Pearl *et al.* 2009; Bragin *et al.* 2012; Levsen *et al.* 2016). First, a 'drift-like' process could result in a loss of one of the cytoplasmic variants during the growth and development of a plant (i.e. 'vegetative sorting,' Mogensen 1996; Birky 2001). Sorting could restore homoplasmy but with novel plastid–mitochondrial combinations within individuals. Second, homologous (or sexual) recombination is known to occur in plant mitochondrial genomes under heteroplasmic conditions, resulting in novel, recombinant mitochondrial haplotypes independent of plastid genomes (Stadler & Delph 2002; Barr *et al.* 2005; McCauley *et al.* 2005; McCauley & Ellis 2008). Plastid recombination is typically not expected even under heteroplasmy because plastid fusion is rare (Sears 1980; Birky 1995; Nagata 2010).

Gynodioecious angiosperms are interesting systems to understand cause and consequences of the breakdown of cytoplasmic LD. In these species, cytoplasmic male sterility (CMS) genes located in mitochondrial genomes disrupt pollen production in some individuals resulting in functionally female phenotype. Consequently, male sterile (females) and fertile (hermaphrodites) individuals co-occur within natural populations (Frank 1989; Sakai & Weller 1999; Shykoff *et al.* 2003). Species with gynodioecy may show weaker cytoplasmic LD (compared to non-gynodioecious species) because several studies have reported non-maternal inheritance, heteroplasmy, and/or recombination of mitochondrial genomes in these species

(Bragin *et al.* 2012; reviewed by McCauley 2013; Levsen *et al.* 2016). However, mechanisms that create the above conditions have not been understood.

I hypothesized that weak cytoplasmic LD in gynodioecy is associated with high diversity of CMS genes in these species as suggested by theoretical models (Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991; Couvet *et al.* 1998; Dufay *et al.* 2007; McCauley & Bailey 2009). Although actual CMS genes have not been identified or characterized at molecular level in any natural gynodioecious species, higher mitotype diversity (than non-gynodioecious species), presumably associated with higher CMS gene diversity, have often been reported (Touzet & Delph 2009; Lahiani *et al.* 2013; Delph & Montgomery 2014). The co-occurring mitotypes statistically increase the chances of heteroplasmy in the event of paternal leakage, subsequently increasing the opportunities for vegetative sorting, and/or mitochondrial homologous recombination—both processes contributing to the breakdown of cytoplasmic LD.

I also hypothesized that overall cytoplasmic LD is better captured if populations with higher female frequencies are adequately sampled. This is because high-female populations are expected to have higher CMS diversity and also higher mitotype diversity (Delph & Kelly 2014; Delph & Montgomery 2014). Thus, the processes weakening cytoplasmic LD (heteroplasmy, sorting, and recombination) are expected to occur more frequently in high-female populations. This could explain why there are inconsistent reports of both weak (Lahiani *et al.* 2013; Storchova & Olson 2004; Houliston & Olson 2006) and strong (Desplanque *et al.* 2000; Olson & McCauley 2000) LD in gynodioecious species, including contrasting reports within the single species (*S. vulgaris*).

In this study, I estimated the level of overall plastid–mitochondrial LD in a gynodioecious wildflower *Lobelia siphilitica* L. (Campanulaceae) and evaluated potential mechanisms affecting cytoplasmic LD, such as mitochondrial recombination, and assessed cytoplasmic genetic diversity within and across the populations. In order to better estimate cytoplasmic LD and cytoplasmic genetic diversity, I sequenced a hypervariable plastid marker and three mitochondrial gene-based markers—an intron (*nad7ab*), a coding sequence (*atp6*), and a pseudogene (*Ψrps12*)—for 266 individuals from 61 geographically widespread populations representing full spectrum of female frequencies (0–100%) to ask the following specific questions.

First, I asked whether plastid and mitochondrial genomes in *L. siphilitica* have weak LD as expected if the two cytogenomes were not strictly maternally co-inherited and/or clonally replicated in this species. Second, I asked if there is evidence of mitochondrial recombination in *L. siphilitica* as expected if populations of this species had multiple mitotypes (corresponding to multiple CMS types) increasing the chances of heteroplasmy. Third, I asked whether mitochondrial and/or plastid genetic diversity was positively correlated with female frequency as expected if high-female populations had higher numbers of unique CMS types (Delph & Kelly 2014). If plastid and mitochondrial genomes in *L. siphilitica* have strong LD, I expected to find similar patterns of genetic diversity with both cytoplasmic markers. Finally, I used the patterns of diversity and distribution of mitotypes within and across populations to understand potential mechanism maintaining gynodioecy in *L. siphilitica*.

Multiple possible mechanisms could cause high mitotype diversity in populations of a gynodioecious species. Therefore, I asked whether this is caused by long-term maintenance of multiple mitotypes as expected if CMS types are under balancing selection (Charlesworth 1981;

Gouyon *et al.* 1991; Dufay *et al.* 2007; Delph & Kelly 2014), frequent invasion and periodic replacement of mitotypes as expected if CMS types are under epidemic dynamics (Frank 1989; Couvet *et al.* 1998), or frequent formation of novel mitotypes as expected if novel CMS types are periodically formed by recombination activities (Hanson & Bentolila 2004; Knoop 2004; Kubo & Newton 2008).

My study was based on the assumption that marker-based mitotypes represent CMS types because of their co-occurrence within the same (mitochondrial) genome. Indeed, mitotypes jointly defined by *nad7ab* and *atp6* were found to correspond strongly to CMS types inferred from my recent crossing study, providing validation to my assumption (Chapter III). My study still had some limitations. First, I aimed to include more populations to cover a broader geographic range of the species as opposed to intensively sampling within populations. Thus, I did not have enough power to estimate cytoplasmic LD or to carry out formal tests of recombination at the population level. Instead, I estimated overall LD and recombination tests at the species level. Nevertheless, I used indirect methods to infer recombinant mitotypes from the pool of mitotypes found within species and used this information to compare potential rates of recombination among populations varying in frequencies of females (see Methods). Second, given the nature of mitochondrial markers used in my study (that is, an intron, a coding gene, and a pseudogene) I could not directly determine the ages of mitotypes. Instead, I inferred their times of maintenance based on their frequency and geographic distribution (see Methods).

MATERIALS AND METHODS

Sampling and marker sequencing

I sampled 266 individuals from 61 geographically widespread populations representing 12 US states and the Canadian province of Ontario, which comprised a full range of population sex ratios (0–100% female; Figure 2.1). This maximized my ability to compare cytoplasmic genetic diversity within and among populations to variation in population sex ratio. I selected study populations from a pool of 160 natural sites for which I had sex-ratio data (Caruso & Case 2007, 2013 and unpublished data). The mean sex ratio of populations used in this study was relatively similar (23.4%; *N*=61) to the mean sex ratio of the entire set (18.2%; *N*=160 populations), although I slightly over-sampled high-female populations relative to the full sample (Figure 2.1). I chose to maximize the number of populations sampled rather than individuals per population, assuming high population structure typical of cytogenomes (e.g., Tarayre & Thompson 1997; Petit *et al.* 2005; McCauley & Ellis 2008; Adhikari & Wallace 2014). Accordingly, I sampled 2–7 individuals (mean= 4) from each population (total *N*= 266 individuals sampled) and included plants of both sexes where present. The final dataset comprised 87 females, 177 hermaphrodites, and 2 plants of undetermined sex (Table 2.S1).

DNA samples used here represent a subset of samples previously studied for variation in plastid *psbK–rps16* intergenic region—henceforth 'plastid marker' (Madson 2012). This single plastid marker contains five variable features (Knox 2014): (1) 1 or 2 tandem copies of *trnQ*; (2) 0–2 copies of pseudo-*trnQ* (*Ψtrn*Q); (3) 1–3 copies of a 94-bp repeat; (4) 1–13 repeats of an imperfect and hypervariable 48- to 54-bp minisatellite with 98 unique motifs; and (5) a unique, chimeric open reading frame (*orf262*). A phylogenetic analysis based on informative sites,

excluding the minisatellite, resolved into 13 unique haplogroups (Madson 2012). I compared these plastid sequences with the mitochondrial sequences generated for the present study.

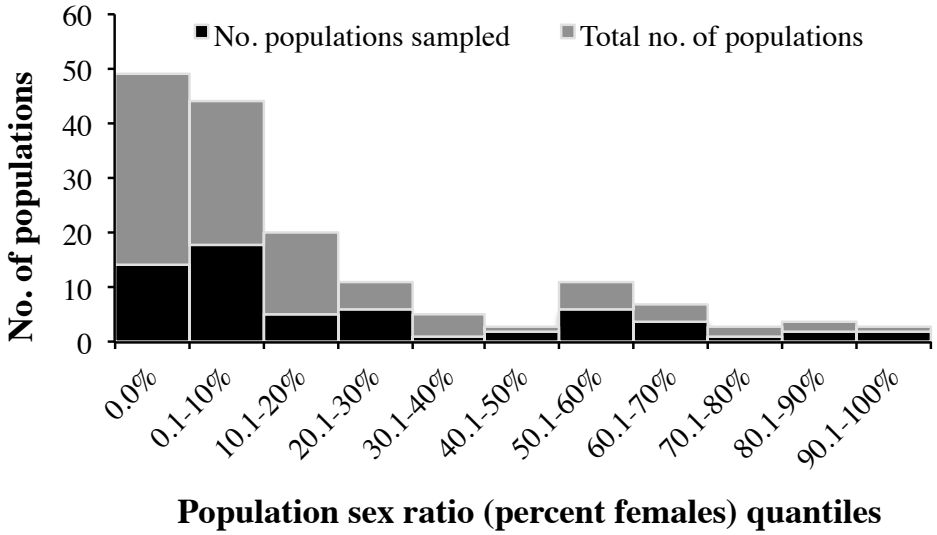


Figure 2.1. Distribution of sex ratios (percent female) of 160 natural populations of *L. siphilitica* (gray) and 61 populations used in this study (black) that came from 12 U. S. states and the Canadian province of Ontario. Detailed information on the study populations is provided in Table 2.S1.

I sequenced three mitochondrial genic regions: (1) *atp6* coding region; (2) pseudo-*rps12* (*Ψrps12)*; and (3) the first intron of *nad7* (*nad7ab*) (Table 2.S2). I used multiple mitochondrial markers for two reasons. First, given the typically slow point mutation rates of plant mitochondrial genomes, single markers were not expected to provide sufficient variation (Wolfe *et al.* 1987; Palmer & Herbon 1988). Second, only multiple markers would allow me to document unique multilocus mitotypes (e.g., Fragoso *et al.* 1989; Darracq *et al.* 2010) and test for intergenic recombination (Desplanque *et al.* 2000; McCauley & Ellis 2008).

All PCR reactions were carried out under standard reaction conditions: 25 μL in volume, containing 2.5 μL of template DNA and with a primer-specific annealing temperature (Table 2.S2). Purification and Sanger sequencing of amplicons was performed by Macrogen USA (Rockville, MD, USA) using ABI3730XL (Applied Biosystems, Foster City, CA, USA). I manually aligned sequences in Sequencher 5.2.3 (Gene Codes, Ann Arbor, MI, USA) and edited as needed.

Sequence data analysis

I used 13 unique haplogroup sequences excluding the minisatellite identified by Madson (2012), hereafter 'plastid haplotypes,' to estimate plastid genetic diversity metrics used in this study. I computed mitochondrial genetic diversity separately for individual markers and for concatenated mitochondrial sequences. Because plant mitochondrial genomes undergo inter- and intramolecular recombination, novel multilocus haplotypes are often formed (Stadler & Delph 2002; McCauley *et al.* 2005; McCauley & Ellis 2008). Concatenation allowed me to compare mitochondrial variation related to recombination with variation in plastid genomes. I used an arbitrary order of mitochondrial markers, *nad7ab–atp6–Ψrps12*, for concatenation because gene

order is unknown for any given mitotype in *L. siphilitica* and likely variable among mitotypes (Knoop 2004; Kubo & Newton 2008). However, using a different gene order did not alter the number of multilocus mitotypes, thus did not change my conclusions (data not shown).

I tallied numbers of segregating sites, the number of unique haplotypes, haplotype diversity (*Hd,* Nei & Tajima 1981), and nucleotide diversity (π, Nei 1987) among markers using DnaSP v5 (Librado & Rozas 2009). I estimated population genetic structure for haploid sequence data (F_{ST}, Hudson *et al.* 1992) using DnaSP v5. I also sequenced four other species of *Lobelia* (DNA samples courtesy of Knox EB) at mitochondrial markers and obtained sequences for the fifth species (*L. laxiflora*, Knox EB unpublished data) to be used as outgroups.

Test for plastid–mitochondrial LD

My first question was whether the two cytogenomes in *L. siphilitica* have weak LD as expected if the two cytogenomes were not strictly maternally co-inherited and/or clonally replicated in gynodioecious species. I estimated LD as standardized D', which measures the degree of deviation of pairs of alleles at two loci from the random association (Lewontin & Kojima 1960). D' ranges from -1 to 1, where -1= complete dissociation, 0= linkage equilibrium and 1= perfect association. I analyzed absolute values ($|D'|$) such that values closer to 1 would indicate stronger LD and the values closer to 0 would indicate weaker LD.

I used Arlequin v 3.5 (Excoffier & Lischer 2010) to calculate $|D'|$ between plastid haplotypes and mitotypes based on individual mitochondrial loci as well as 3-locus mitotypes. I averaged $|D'|$ to estimate overall concordance between each pair of plastid–mitochondrial markers (or 3-locus mitotypes). I used only common alleles/haplotypes to calculate average $|D'|$ because rare haplotypes artificially inflate D, yielding values of 1 or -1 because of their rarity,

rather than strong LD (McCauley & Ellis 2008). Defining 'rarity' of haplotypes was not straightforward. Therefore, I assumed that haplotypes found in fewer than 5 individuals *and* fewer than 3 populations were likely rare. Based on these criteria, most haplotypes were common; less than 7% of my samples had rare haplotypes at individual mitochondrial loci (*cf.* rarity of 3-locus mitotypes below). In order to test whether the level of plastid–mitochondrial LD varied among mitochondrial markers, I ran Kruskal–Wallis rank tests (Kruskal & Wallis 1952) to compare the values of $|D'|$ across three plastid–mitochondrial marker pairs.

In addition to direct LD tests of concordance between plastid and mitochondrial genomes, I used 4-gamete tests of recombination (implemented in DnaSP v5; Hudson & Kaplan 1985). The presence of all four possible allelic combinations at any two loci suggests recombination or recurrent mutation. The 4-gamete test also estimates a minimum number of recombination events (Rm) in the sequences. Note that evidence of plastid–mitochondrial recombination based on 4-gamete tests would indicate a lack of strong LD (i.e., discordance) between two cytogenomes rather than a physical exchange of haplotypes.

Tests for recombination within the mitochondrial genome

My second question was whether there is mitochondrial recombination in *L. siphilitica,* as expected if populations of this species contained multiple mitotypes (likely corresponding to multiple CMS types) thus, increasing the chances of heteroplasmy. To assess recombination, I used 4-gamete tests (1) between each pair of mitochondrial markers (intergenic recombination) and (2) among variable nucleotide sites within each mitochondrial marker (intragenic recombination). I also used three additional methods to infer mitochondrial recombination: (1) estimates of $|D'|$ within and among mitochondrial markers, (2) the architecture of mitotype

networks, and (3) comparisons of observed vs. expected number of multilocus mitotypes.

First, I determined $|D'|$ as described above, and tested whether LD varied within each marker (intragenic recombination) and among pairs of markers (intergenic recombination) using Kruskal–Wallis tests. In the absence of recombination, LD should be consistently high. Second, I examined haplotype networks for loops, which typically indicate recombination or recurrent mutation (Posada & Crandall 2001). I created Median-joining networks (Bandelt *et al.* 1999) for each individual marker and for the 3-locus mitotypes using Network v5 (fluxus-engineering.com). I used maximum parsimony post-processing to remove unnecessary median vectors from the networks (Polzin & Vahdati Daneshmand 2003). Finally, I compared the number of 3-locus mitotypes expected without recombination to the number observed in my sample. Likewise, I compared the expected number of plastid–mitochondrial combinations (cytotypes) with the observed number. I computed the expected numbers (assuming no recombination) as $\Sigma x_i - n + 1$, where n is the number of loci and x_i is the number of alleles per locus (McCauley 2013). Recombination and/or recurrent mutation would result in a greater-than-expected number of haplotypes/cytotypes.

Tests for association between population sex ratio and haplotype diversity

My third question was whether mitochondrial genetic diversity in *L. siphilitica* is positively correlated with female frequency, as expected if higher numbers corresponding CMS types are present in high- (than low-) female populations (Delph & Kelly 2014). I used three different genetic diversity measures: number of haplotypes (*H*), haplotype diversity (*Hd*), and nucleotide diversity (π) for each marker and the 3-locus mitotypes, and correlated them against mean population sex ratios averaged across 1–8 years (Caruso & Case 2007; Madson 2012; Case

AL & Caruso CM, unpublished data). I used Spearman's correlation because neither of the variables was normal ($P< 0.05$; Shapiro & Wilk 1965). I ran normality and correlation tests in R v3 (R core team).

Tests of models of gynodioecy in L. siphilitica

My final goal was to understand the model of gynodioecy in *L. siphilitica* based on the patterns of diversity and distribution of mitotypes within and among populations. Specifically, I asked whether the mitotypes in this species are maintained for a long time (balancing selection), they are frequently replaced (epidemic dynamics) and/or they are frequently formed by mitochondrial recombination. As any of these mechanisms could potentially result in multiple mitotypes within a population at a given time point, determining their times of maintenance was important to address this question.

Because mitotypes used in my study were defined by three unrelated mitochondrial geneic sequences—an intron (*nad7ab*), a coding sequence (*atp6*), and a pseudogene (*Ψrps12*), I could not directly estimate their ages. Therefore, I inferred relative times of occurrences of mitotypes based on their frequencies and distribution. I predicted that older mitotypes could migrate over time. Thus, I assumed that mitotypes present in at least five individuals and three populations are old. That is, newly formed mitotypes are unlikely to be widespread in a species like *L. siphilitica* with low rates of cytoplasmic gene flow (Madson 2012; this study). Instead, rare mitotypes are more likely to be recent in origin unless gene flow is limited in some part(s) of the species range. In addition, I used the following posthoc tests to understand the potential mechanisms by which high mitotype diversity was formed in *L. siphilitica*.

Inferring mechanism of origin of mitotypes

First, because I observed a predominance of rare 3-locus mitotypes, I asked how such a high number of rare mitotypes could have formed. Given the evidence of extensive mitochondrial recombination in my results, I had little doubt that many of my mitotypes formed locally by recombination. However, distinguishing recombinant mitotypes from among the pool of mitotypes collected from the wild population was not straightforward. Thus, I assumed that the rare mitotypes are derived (potentially by recombination) and common mitotypes could be parental. Common mitotypes are more likely to go through the process of paternal leakage and heteroplasmy before they could recombine without being lost via genetic drift and/or vegetative sorting. If recombination has created novel mitotypes locally, I predicted that populations should have unique assemblages of mitotypes. That is, populations varying in sex ratios not only vary in mitotype diversity but also in specific assemblages of mitotypes. To test this prediction, I ran Mantel correlations (Mantel 1967) comparing pairwise sex-ratio distances (computed in R v3) to pairwise genetic distances (Tamura & Nei 1993; computed in Arlequin v3.5). I used PASSaGE v2 for Mantel's tests with 10,000 permutations (Rosenberg & Anderson 2011).

Recombination and CMS type formation

CMS genes are chimeric genes formed by recombination/rearrangement activities in mitochondrial genomes (Hanson & Bentolila 2004; Knoop 2004; Kubo & Newton 2008). Thus, I asked whether the rare (potentially recombinant) mitotypes observed in my study (see Results) were formed in association with the formation of novel CMS types. If this is the case, I predicted that rare (and potentially recombinant) mitotypes would be more common in high-female populations and in female plants. In order to test this hypothesis, I tallied rare vs. common 3-

locus mitotypes for each sex-ratio quartile. I used a Chi-squared test of association to evaluate the null hypothesis that the distribution of rare mitotypes among populations was independent of the sex-ratio quartile. Similarly, I examined whether the rare mitotypes are more common in female plants.

RESULTS

Sequence dataset and characteristics

Plastid sequences of all 266 individuals resolved into 13 haplogroups, as found by Madson (2012). Although 17.3% of samples failed to sequence successfully for one or more of the mitochondrial markers, I were able to identify 39 three-locus mitotypes from a total sample of 225 individuals in 60 populations (Table 2.1). Based on SNPs (all loci) and indels (plastid and *nad7ab*), I observed 6–13 unique haplotypes at each marker and ~4-fold variation in haplotype and nucleotide diversity among markers (Table 2.1; Table 2.S3). I also found high cytoplasmic population-genetic structure in *L. siphilitica* (F_{ST}; Table 2.1). Some mutations are potentially homoplasious. The mutations in *nad7ab* (an insertion, two deletions, and a point mutation) were rare in *L. siphilitica* but present in distant *Lobelia* species (Table 2.S3), and a point mutation in *atp6* was inferred to be homoplasious based on the mitotype network (Figure 2.2A). Because homoplasy could affect my inference of recombination, I repeated the test for recombination both with and without the potentially homoplasious mutations.

Table 2.1. General characteristics of sequence data used in this study: Diversity metrics were computed for the entire sample and for individual sex (two samples with undetermined sexes were excluded). Values in parentheses include indels for *H* and *Hd*.

Marker, bp	N	No. of populations	*S*	Spar	No. of indels	*H*	*Hd*	π	*FST*
plastid, 1517	266	61	32	30	12	11 (13)	0.78 (0.82)	0.007	0.65
Female	88	40	27	23		7 (8)	0.79 (0.79)	0.007	
Hermaphrodite	176	59	30	30		10 (12)	0.77 (0.83)	0.007	
nad7ab, 872	265	61	8	8	6	8 (11)	0.34 (0.57)	0.0005	0.6
Female	88	39	8	6		8 (10)	0.51 (0.67)	0.0007	
Hermaphrodite	175	58	6	5		6 (9)	0.23 (0.51)	0.0003	
atp6, 706	241	60	11	4	0	11	0.72	0.002	0.45
Female	74	39	4	4		7	0.778	0.002	
Hermaphrodite	165	58	11	4		11	0.642	0.0018	
Ψrps12, 340	238	59	6	4	0	6	0.17	0.0007	0.51
Female	72	37	6	3		6	0.343	0.0015	
Hermaphrodite	164	58	4	1		3	0.094	0.0004	
3-locus mt, 1918	220	56	20	14	6	34 (39)*	0.76 (0.85)	0.001	0.65
Female	63	35	15	11		24 (28)	0.92 (0.94)	0.0013	
Hermaphrodite	155	57	16	9		18 (22)	0.66 (0.79)	0.0008	

N= number of individuals sequenced, *S*= number of segregating sites, *S*par= number of parsimony informative sites, *H*= number of haplotypes, *Hd*= haplotype diversity, π= nucleotide diversity, F_{ST} = population genetic structure. * Thirty-nine 3-locus mitotypes include 4 mitotypes (from 5 samples) that could be inferred despite missing sequences (total N= 225; populations= 60)

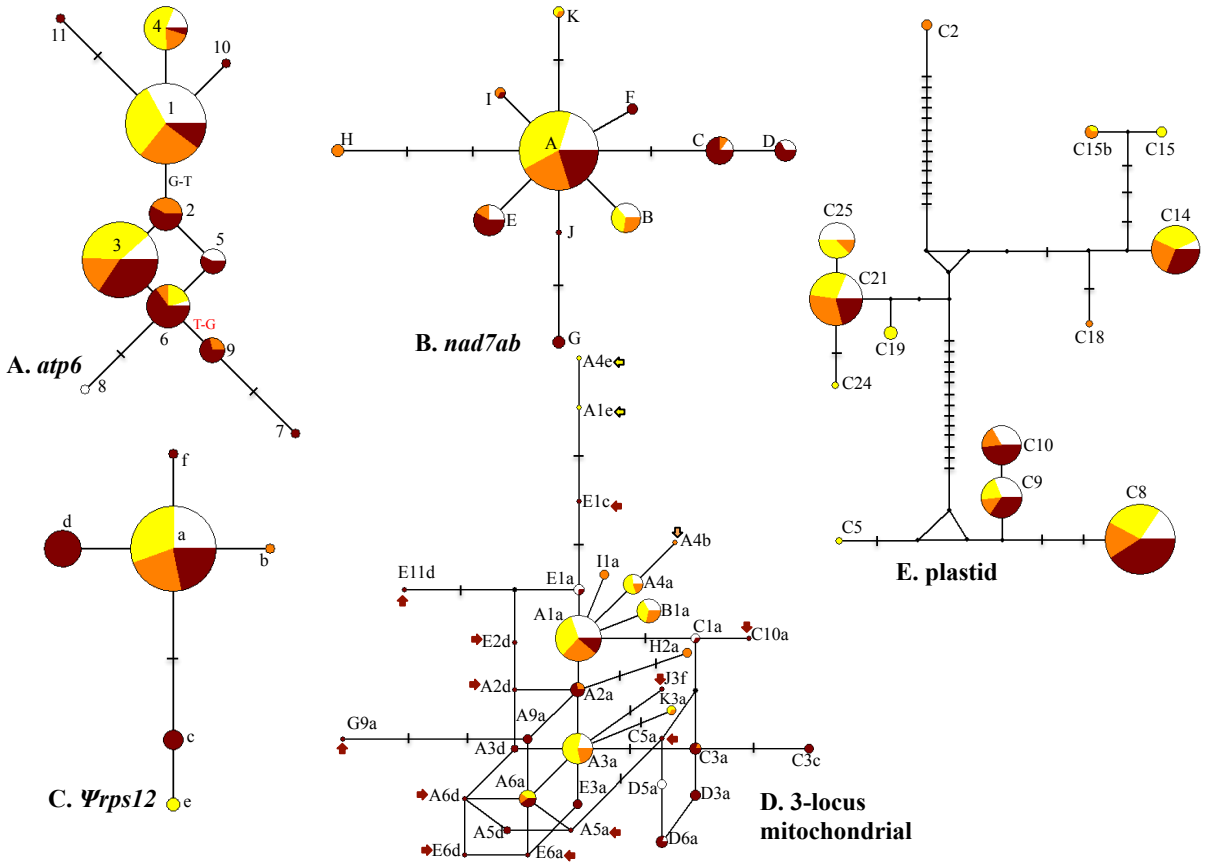


Figure 2.2. Median-joining networks for individual mitochondrial alleles (A-C), 3-locus mitotypes (D), and plastid haplotypes (E). The size of each circle is proportional to the frequency of each unique allele/haplotype (arrows point to very small circles). Black dots (median vectors) and dashes (mutations) represent hypothetical alleles/haplotypes. Pie colors indicate proportions of populations with different sex ratio quartiles containing that allele/haplotype: 0% (white), 0.1-10% (yellow), 10.1-35% (orange), >35% (maroon). The reversal event (T-G) in *atp6* is marked.

Decay of LD between plastid and mitochondrial genomes

I found relatively weak LD between plastid and mitochondrial genomes in *L. siphilitica* as expected if the two cytogenomes were not strictly maternally co-inherited and/or clonally replicated in this species. This contrasts with the general expectation of strong cytoplasmic association in flowering plants. Average pairwise |*D′*| between 'common' plastid haplotypes and 3-locus mitotypes was 0.49 (Table 2.S4A). Likewise, average pairwise |*D′*| between common plastid haplotypes and individual marker haplotypes ranged from 0.4–0.5 (Table 2.S4B–D) and did not vary significantly among markers (i.e., the null hypothesis that the two sets of |*D′*| values were similar was not rejected; Kruskal–Wallis tests, $P > 0.05$). In addition to LD tests, plastid–mitochondrial discordance was also indicated by positive tests of recombination between the plastid marker and each of the mitochondrial markers (4-gamete test; Table 2.2). Moreover, I observed a higher number of plastid–mitochondrial cytotypes than expected assuming no recombination (Figure 2.3). The level of plastid–mitochondrial LD did not vary with population sex ratio; values of |*D′*| did not vary significantly among population sex-ratio quartiles (Kruskal–Wallis tests $P> 0.05$).

Evidence of recombination in mitochondrial genome

Four lines of evidence suggested that mitochondrial recombination has occurred in *L. siphilitica* as expected if populations of this species had multiple mitotypes (corresponding to multiple CMS types) thus, increasing the chances of heteroplasmy. First, as for plastid–mitochondrial association, LD between each pair of mitochondrial markers was decayed: average |*D′*| ranged from 0.4–0.62 (Table 2.S5 A–C) and did not vary significantly among marker pairs (Kruskal–Wallis test, $P > 0.05$). Second, 4-gamete tests predicted at least one recombination

Table 2.2. Four-gamete tests showing recombination events within and between one plastid and three mitochondrial markers. Number of segregating sites is shown in the second column. Values in the remaining columns indicate the number of pairs of sites, showing 4-gametic types (outside parentheses) and the minimum inferred number of recombination events (inside parentheses). Values on the diagonal indicate recombination within the marker, while those below the diagonal indicate recombination between marker loci.

	No. of segregating sites	***nad7ab***	***atp6***	***Ψrps12***	**plastid**
nad7ab	7	0 (0)			
atp6	11	7 (1)	3 (2)*		
Ψrps12	6	3 (1)	10 (1)	0 (0)	
plastid	32	43 (1)	82 (1)	74 (1)	6 (2)

* Recombination was not found when excluding sites with possible homoplasy

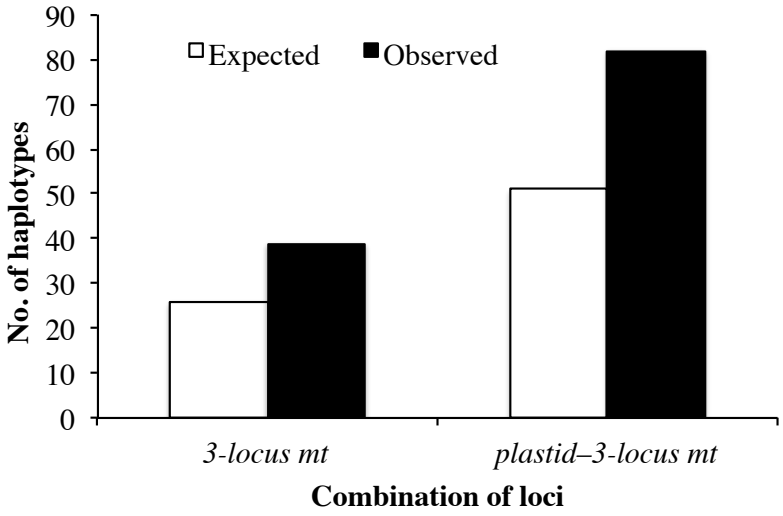


Figure 2.3. Number of 3-locus mitotypes (3-locus mt) and plastid–mitochondrial cytotypes (plastid–3-locus mt) expected under the assumption of no recombination or homoplasy (gray) and the numbers actually observed (black). Expected number of haplotypes was determined using an extension of 4-gamete rule (see McCauley 2013 and Methods text).

Table 2.3. Spearman rank correlations relating population sex ratio (SR, proportion females) to haplotype diversity (*H, Hd*) and nucleotide diversity (π). Genetic diversity metrics exclude indels and exclude monomorphic populations having only one haplotype. *P*-values: < 0.05*; < 0.01**; < 0.001***

			Correlation coefficients		
Marker	**N**	**No. Pop**	**SR *vs. H***	**SR *vs. Hd***	**SR *vs.* π**
plastid	266	61	0.08	0.08	0.15
nad7ab	265	61	0.52***	0.52***	0.52***
atp6	240	60	0.31*	0.26*	0.29*
Ψrps12	236	59	0.31*	0.31*	0.29*
3-locus mt	216	56	0.37**	0.40**	0.40**

event between each pair of mitochondrial markers, and at least two within *atp6* (Table 2.2). There was also a slight decay of LD among nucleotide sites in *atp6* (mean $|D'|$= 0.86), consistent with intragenic recombination in *atp6*; such decay was not found in *nad7ab* or *Ψrps12* (mean $|D'|$= 1). Third, networks for both 3-locus mitotypes and *atp*6 contained loops, consistent with recombination (Figure 2.2A, D). Finally, my observation of more 3-locus mitotypes than the number expected when assuming no recombination is also consistent with mitochondrial intergenic recombination (Figure 2.3).

When excluding the potentially homoplasious mutations in *nad7ab* and *atp6* (Table 2.S3), 4-gamete tests showed no evidence of recombination within *atp6*, but the results of intergenic recombination did not change. The 4-gamete test also showed a minimum of two recombination events within the plastid marker (Table 2.2). However, LD among plastid nucleotides was high (mean $|D'|$= 0.998), and there were no loops in the plastid network (Figure 2.2D), suggesting homoplasy.

High mitochondrial (but not plastid) diversity in high-female populations

I found a positive correlation between mitochondrial genetic diversity and population sex ratio as expected if high-female populations had higher CMS diversity (Table 2.3). However, plastid markers failed to show such a correlation consistent with a weak plastid–mitochondrial LD. Positive correlations between mitotype diversity and female frequency were primarily derived from low mitochondrial diversity in low-female populations (Figure 2.S1). When populations with no females were excluded from the analyses, correlations only remained significant for *nad7ab*. Moreover, although females had a smaller sample size (N= 63–88) in my total sample, they had higher mitotype diversity than hermaphrodites (N= 155–176; Table 2.1).

Distribution of older and newer mitotypes in L. siphilitica

Assuming common and geographically widespread mitotypes to be older while rare and geographically restricted mitotypes to be relatively young, most (85%) of the populations of *L. siphilitica* used in my study had one or more older 3-locus mitotypes (Table 2.S6). These mitotypes could be have been maintained within the species for a long period of time (potentially in association with CMS types) and migrated over time. However, 43 % (of total 60) of my populations also had one or more rare 3-locus mitotypes. These mitotypes could have been formed locally. Indeed, very few of my 3-locus mitotypes were shared among populations: among 31 rare 3-locus mitotypes observed, 26 were each unique to a single population and five were each shared between just two or three populations (Table 2.S6). This was also suggested by significant Mantel correlations between pairwise sex-ratio distances and pairwise genetic distances among populations (r= 0.20; P= 0.003). That is, populations varying in sex ratios not only vary in mitotype diversity but also in specific assemblages of mitotypes.

Mechanisms of formation of rare mitotypes

About 80% of the 3-locus mitotypes in my dataset were rare (i.e., occurring in fewer than 5 individuals and fewer than 3 populations) and likely to be recombinant (see Methods). Such a predominance of rare 3-locus mitotypes despite fewer rare alleles at individual mitochondrial loci suggests that they were likely formed by recombination. In other words, the 3-locus mitotypes were rare because they had unique combinations of common alleles rather than rare alleles at individual markers.

My data suggested that the rare 3-locus mitotypes could be formed in association with the formation of CMS types. First, the rare mitotypes were more prevalent in high-female

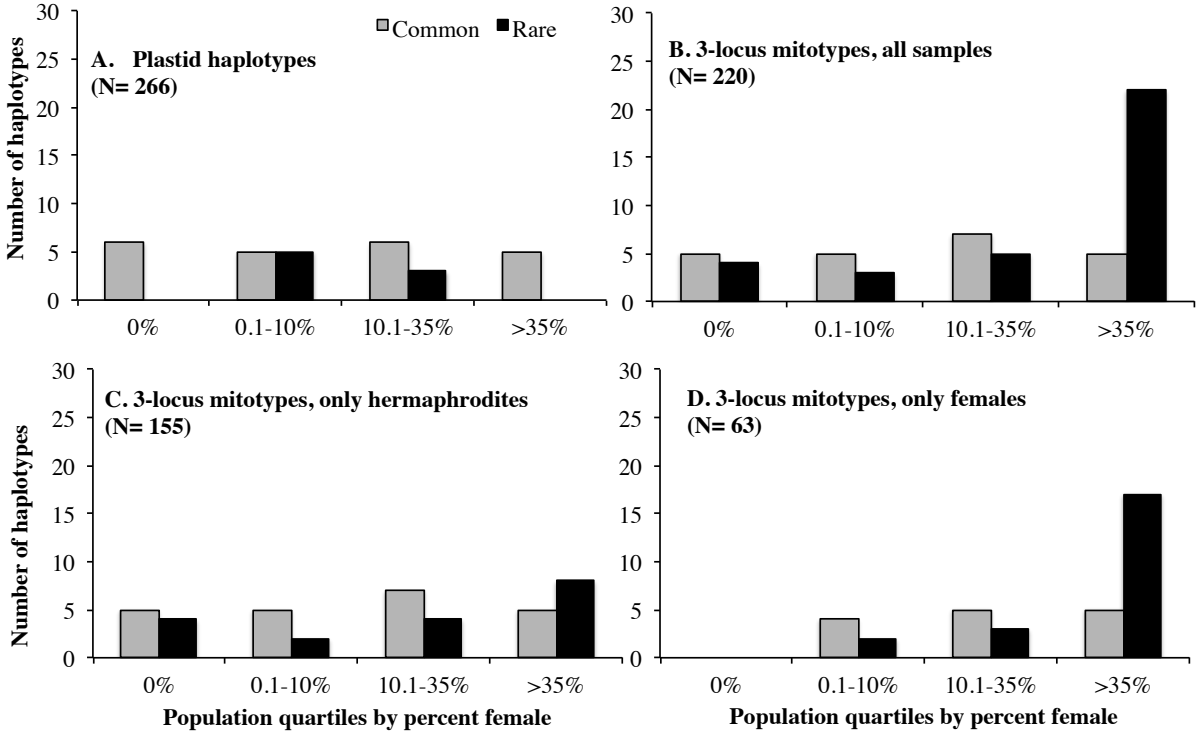


Figure 2.4. Distribution of 'rare' and 'common' haplotypes/mitotypes among 61 populations by sex ratio (percent female). Distribution with: (A) plastid haplotypes, (B-D) 3-locus mitotypes with all samples (B), only hermaphrodites (C), and only females (D). Rare haplotypes/mitotypes occur in < 2% of plants sampled (< 5 individuals and < 3 populations). Populations were divided into quartiles based on percent female. Number of populations in each quartile: 14 (0%), 18 (0.1-10%), 12 (10.1-35%) and 17 (>35%), respectively. Note: Two plants were of undetermined sex.

populations (Figures 2.2D and 2.4B). Rare plastid haplotypes did not show any association with population sex ratio (Figure 2.4A). A Chi-squared test rejected the null hypothesis that rare and common mitotypes were distributed evenly among sex-ratio quartiles (χ^2= 7.9, $P < 0.05$). Second, rare mitotypes were more common in female plants (23 rare mitotypes, N= 66) than hermaphrodites (15 rare mitotypes, N= 159) despite a much smaller sample size of females in my study (Figure 2.4). Among females, rare mitotypes were more common in the highest sex-ratio quartile of populations (SR >35%), but not in the lower quartiles (Figure 2.4D). This could indicate that the formation of rare mitotypes could be associated with feminization factors or CMS genes. Because novel CMS types are formed by mitochondrial recombination activities (Hanson & Bentolila 2004; Knoop 2004; Kubo & Newton 2008), rare 3-locus mitotypes could be formed in association with the novel CMS-type formation in high female populations.

DISCUSSION

Cytoplasmic discordance and gynodioecy

My data indicated that LD between plastid and mitochondrial genomes could weaken in gynodioecious species like *L. siphilitica* in which the expected rules of strict maternal co-inheritance and clonal replication of cytogenomes could be violated. The most obvious cause of cytoplasmic discordance in *L. siphilitica* appears to be mitochondrial recombination, particularly, intergenic recombination, as shown by my observation of numerous rare 3-locus mitotypes.

Mitochondrial recombination in gynodioecious species could be associated with the co-occurrence of multiple mitotypes (corresponding to multiple CMS types) within populations, which increases the chances of heteroplasmy in the event of paternal leakage (McCauley 2013).

Thus, populations with higher female frequencies, which were found to have higher mitotype diversity (Table 2.3), could contribute more to the weakening of cytoplasmic LD. Assuming that rare, 3-locus mitotypes are recombinant, the predominance of rare mitotypes in high-female populations supported my hypothesis that high female frequency could be associated with greater cytoplasmic discordance.

Variation in numbers of paternal leakage, heteroplasmy, and recombination among populations of gynodioecious species was reported previously (e.g., *S. vulgaris*, Welch *et al.* 2006; Pearl *et al.* 2009; Bentley *et al.* 2010; *P. lanceolata*, Levsen *et al.* 2016). However, I showed for the first time that there could be an association between atypical mitochondrial behavior (inheritance and replication) and female frequencies in populations. Although my study was not designed to test paternal leakage and heteroplasmy, I found some heteroplasmic sequences of *atp6* and *Ψrps12* in *L. siphilitica* and those sequences were more common in high-female (>20%) populations (Adhikari B, Caruso CM, Case AL, unpublished data). However, direct evidence of mitochondrial paternal leakage is still lacking in *L. siphilitica*—pedigrees showing evidence of plastid biparental inheritance (Durewicz A, Knox EB, Case AL, unpublished data) failed to show such pattern with the mitochondrial markers used in my study (Adhikari B, Case AL, unpublished data). Nevertheless, this provided another evidence for plastid–mitochondrial discordance in *L. siphilitica*.

In theory, heteroplasmy could also be caused by substoichiometric shifting (SSS), a sudden change in the stoichiometry of mitochondrial subgenomic molecules in flowering plants (see refs in Arrieta-Montiel *et al.* 2001). In crop plants, SSS has been found to be associated with the expression of cytoplasmic male sterility (Arrieta-Montiel *et al.* 2001; Tang *et al.* 2017). However, this phenomenon still remains to be explored in wild flowering plants. Detailed studies

of paternal leakage and SSS could help understand relative roles of these mechanisms in creating heteroplasmy.

Evolutionary model of gynodioecy in L. siphilitica

Evolutionary models of gynodioecy in flowering plants still remain poorly understood. Most marker-based studies have suggested that gynodioecy could be caused by long-term maintenance of multiple CMS types by balancing selection (Houliston & Olson 2006; Touzet & Delph 2009; Lahiani *et al.* 2013) except one study based on plastid markers, which reported patterns consistent with epidemic dynamics (Ingvarsson & Taylor 2002). Previous studies in *L. siphilitica* also provided results consistent with balancing selection because (1) gynodioecious *Lobelia* (including *L. siphilitica*) show higher mitotype diversity than close relatives (Delph & Montgomery 2014); (2) crossing experiments showed evidence of multiple CMS types within populations of *L. siphilitica* (Dudle *et al.* 2001; Bailey 2002; Chapter III), and (3) evidence of a cost of restoration, theoretically an important component of the selective maintenance of females in gynodioecious plants, has been well documented in *L. siphilitica* (Bailey 2002; Case & Caruso 2010; Caruso & Case 2013).

Assuming that mitotypes roughly correspond to CMS types, the positive correlation of mitotype diversity and female frequency in populations I observed in *L. siphilitica* is consistent with the occurrence of higher CMS diversity in high-female populations (Table 2.3). In addition, my observation of common mitotypes in majority of populations could indicate that gynodioecy in *L. siphilitica* could be caused by long-term maintenance of CMS types by balancing selection. However, the rare and localized mitotypes could be associated with novel CMS types that could

have invaded the populations relatively recently, although restricted distribution could, sometimes, be caused by highly limited gene flow.

The association of rare mitotypes with high female frequency and particularly with female plants suggests that rare mitotypes could be formed in association with feminizing genes, or CMS types. That is, mitochondrial recombination forming novel CMS types (Hanson & Bentolila 2004; Knoop 2004; Kubo & Newton 2008) could be associated with the formation of rare (potentially recombinant) mitotypes observed in my study. Further study to confirm this prediction requires understanding structure of CMS genes and finding an association between mitochondrial recombination and novel CMS types formation in this species.

My data are not consistent with epidemic dynamics operating in *L. siphilitica* because such a process would cause regular homogenization of mitochondrial genomes (Frank 1989; Couvet *et al.* 1998). Therefore, high mitochondrial genetic diversity and widespread distribution of mitotypes is not likely under epidemic dynamics.

My data indicated that gynodioecy in *L. siphilitica* could be caused by a mechanism similar to that of a recently proposed 'mixed model' (McCauley & Bailey 2009). According to this model, balancing selection maintains a fertile cytotype within a population while multiple CMS types competitively replace one another by partial selective sweeps. Although there is no confirmed evidence of fertile cytotypes in *L. siphilitica,* my recent crossing experiment identified three maternal families with no evidence of female plants in their entire history. These families could potentially represent fertile cytotypes (Chapter III). Alternatively, a modification of the mixed model could operate in *L. siphilitica,* in which one or more CMS types (or fertile cytotypes) are selectively maintained by balancing selection while novel CMS types also invade

(or formed *de novo*) the population periodically and replace older CMS types. Frequent formation of novel CMS types could be responsible for high female frequencies in populations.

Conclusion

In conclusion, my data showed a weak plastid–mitochondrial LD in *L. siphilitica,* which is likely associated with the occurrence of multiple mitotypes corresponding to multiple CMS types in this species. The co-occurring mitotypes could increase the chances of heteroplasmy, sorting, and mitochondrial recombination. Thus, high-female populations with higher mitotype diversity could particularly contribute to the weakening of cytoplasmic LD. My data indicated that some of the CMS types in *L. siphilitica* could be maintained for a long time by balancing selection. However, novel CMS types could also be formed periodically and potentially by mitochondrial recombination. Frequent formation of novel CMA types could cause high female frequencies in populations. Epidemic dynamics that regularly homogenize mitochondrial genetic diversity is unlikely to operate in *L. siphilitica*. It appears that some variant of the 'mixed model' as proposed by McCauley & Bailey (2009) could operate in *L. siphilitica* such that balancing selection maintains one or more CMS types (and/or a fertile cytotype) within a population while novel CMS types could be formed periodically by mitochondrial recombination, thus increasing female frequencies. Further steps to better understand gynodioecy in *L. siphilitica* should address whether rare mitotypes are recombinant and if they are formed in association with novel CMS-type formation. It would also be interesting to seek more direct evidence of paternal leakage and heteroplasmy and to understand if these events are more common in high-female populations.

ACKNOWLEDGEMENTS

I thank my advisor Dr. Andrea Case for her support during the entire period of this research and for her comments on earlier versions of this manuscript. I thank Madson HJ for providing plastid sequences and DNA samples, and Knox EB for providing DNA samples for outgroup species. I thank Knox EF and Delph LF for providing helpful comments on the manuscript. This project was supported by a start-up fund from the Department of Biological Sciences, Kent State University to Case AL, an NSF grant (DEB #0842280) to Case AL & Caruso CM, and a Kent State University Graduate Student Senate research award to Adhikari B.

REFERENCES

Adhikari B, Wallace LE (2014) Does the *Platanthera dilatata* (Orchidaceae) complex contain cryptic species or continuously variable populations? *Plant systematics and evolution* 300: 1465–1476

Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 158: 851–864

Bailey MF (2002) A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* 56: 2178–2186

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48

Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* 168: 39–50

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519–520

Bentley KE, Mandel JR, McCauley DE (2010) Paternal leakage and heteroplasmy of mitochondrial genomes in *Silene vulgaris*: evidence from experimental crosses. *Genetics* 185: 961–968

Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA* 92: 11331–8

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148

Bragin AG, Ivanov MK, Fedoseeva LA, Dymshits GM (2012) Analysis of mitochondrial DNA heteroplasmy of fertile and male-sterile sugar beet plants (*Beta vulgaris*). *Russian Journal of Genetics: Applied Research* 2: 53–57

Breton S, Stewart DT (2015) Atypical mitochondrial inheritance patterns in eukaryotes: a review. *Genome* 58: 423–431

Burt A, Trivers R (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, MA

Caruso CM, Case AL (2007) Sex ratio variation in gynodioecious *Lobelia siphilitica*: Effects of population size and geographic location. *Journal of Evolutionary Biology* 20: 1396–1405

Caruso CM, Case AL (2013) Testing models of sex ratio evolution in a gynodioecious plant: female frequency covaries with the cost of male fertility restoration. *Evolution* 67: 561–566

Case AL, Caruso CM (2010) A novel approach to estimating the cost of male fertility restoration in gynodioecious plants. *New Phytologist* 186: 549–557

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46: 27–39

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75: 1443–1458

Couvet D, Ronce O, Gliddon C (1998) Maintenance of nucleo-cytoplasmic polymorphism in a metapopulation: the case of gynodioecy. *American Naturalist* 152: 59–70

Darracq A, Varré J-S, Touzet P (2010) A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics* 11: 233

Delph LF (1999) Sexual Dimorphism in Life History. *In*: Geber MA, Dawson TE, Delph LF (eds.) *Gender and Sexual Dimorphism in Flowering Plants*. Springer, Berlin, Heidelberg

Delph LF, Kelly JK (2014) On the importance of balancing selection in plants. *New Phytol* 201: 45–56

Delph LF, Montgomery BR (2014) The evolutionary dynamics of gynodioecy in *Lobelia*.

International Journal of Plant Sciences 175: 383–391

Desplanque B, Viard F, Bernard J, Forcioli D, Saumitou-Laprade P, Cuguen J, Van Dijk H (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. *maritima* (L.): The usefulness of both genomes for population genetic studies. *Molecular Ecology* 9: 141–154

Dudle DA, Mutikainen P, Delph LF (2001) Genetics of sex determination in the gynodioecious species *Lobelia siphilitica*: Evidence from two populations. *Heredity (Edinb)* 86: 265–276

Dufaÿ M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99: 349–356

Dumolin-Lapègue S, Pemonge MH, Petit RJ (1998) Association between chloroplast and mitochondrial lineages in oaks. *Molecular Biology and Evolution* 15: 1321–1331

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567

Fragoso LL, Nichols SE, Levings III CS (1989) Rearrangements in maize mitochondrial genes. *Genome* 31: 160–168

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist* 133: 345–376

Gouyon PH, Vichot F, Van Damme JMM (1991) Nuclear-cytoplasmic male sterility: single point

equilibria versus limit cycles. *The American Naturalist* 137: 498–514

Govindarajulu R, Parks M, Tennessen JA, Liston A, Ashman TL (2015) Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany* 102: 544–554

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16 Suppl 1: S154–69

Houliston GJ, Olson MS (2006) Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics* 174: 1983–1994

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–589

Ingvarsson PK, Taylor DR (2002) Genealogical evidence for epidemics of selfish genes. *Proceedings of the National Academy of Sciences, USA* 99: 11265–11269

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet.* 46: 123-139

Knox EB (2014) The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proceedings of the National Academy of Sciences, USA* 111: 11097–11102

Kruskal WH, Wallis WA (1952) Use of Ranks in One-Criterion Variance Analysis. *Journal of American Statistical Association* 47: 583–621

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8 (1): 5–14

Lahiani E, Dufaÿ M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P (2013) Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity (Edinb)* 111: 157–64

Levsen N, Bergero R, Charlesworth D, Wolff K (2016) Frequent, geographically structured heteroplasmy in the mitochondria of a flowering plant, ribwort plantain (*Plantago lanceolata*). *Heredity (Edinb)* 117: 1–7

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458–472

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452

Madson HJ (2012) The effects of stochastic processes on sex-ratio variation in gynodioecious *Lobelia siphilitica* L. (Lobeliaceae). MS Thesis submitted to Kent State University, Kent OH, USA

Manicacci D, Atlan A, Rossello JAE, Couvet D (1998) Gynodioecy and Reproductive Trait Variation in Three *Thymus* Species (Lamiaceae). *International Journal of Plant Sciences*. 159 (6): 948–957

Maynared Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetic Research* 23: 23–35

McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist* 200: 966–977

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* 104: 611–620

McCauley DE, Bailey MF, Sherman NA, Darnell MZ (2005) Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* 95: 50–58

McCauley DE, Ellis JR (2008) Recombination and linkage disequilibrium among mitochondrial genes in structured populations of the gynodioecious plant *Silene vulgaris*. *Evolution* 62: 823–832

Milligan BG (1992) Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. *American Journal of Botany* 79: 1325–1328

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383–404

Nagata N (2010) Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. *J Plant Res* 123: 193–9

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia Univesity Press, New York

Olson MS, McCauley DE (2000) Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proceedings of the Royal Society of London B: Biological Sciences* 267: 1801–1808

Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist* 130: S6-S29

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *Journal of Molecular evolution* 28: 87–97

Pearl SA, Welch ME, McCauley DE (2009) Mitochondrial heteroplasmy and paternal leakage in natural populations of *Silene vulgaris*, a gynodioecious plant. *Molecular Biology and Evolution*. 26: 537-45

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689–701

Polzin T, Vahdati Daneshmand S (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* 31: 12–20

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution* 19: 708–717

Posada D, Crandall KA (2001) Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology and Evolution* 16: 37–45

Reboud X, Zeyl C (1994) Organelle inheritance in plants. *Heredity* 72: 132–140

Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution* 2: 229–232

Sakai AK, Weller SG (1999) Gender and sexual dimorphism in flowering plants: a review of terminology, biogeographic patterns, ecological correlates, and phylogenetic approaches. *In*: Geber MA, Dawson TE, Delph LF (eds.) Pg. 1–31. *Gender and sexual dimorphism in flowering plants*. Springer, Berlin Heidelberg

Schnabel A, Asmussen MA (1989) Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. *Genetics* 123: 199–215

Sears BB (1980) Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. *Plasmid* 4: 233–55

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika 52*: 591–611

Shykoff JA, Kolokotronis SO, Collin CL, López-Villavicencio M (2003) Effects of male sterility on reproductive traits in gynodioecious plants: a meta-analysis. *Oecologia* 135: 1–9

Smith SE (1989) Influence of parental genotype on plastid inheritance in *Medicago sativa.* *Journal of Heredity* 80: 214–217

Stadler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences, USA* 99: 11730–11735

Štorchová H, Olson MS (2004) Comparison between mitochondrial and chloroplast DNA variation in the native range of *Silene vulgaris*. *Molecular Ecology* 13: 2909–2919

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and Evolution* 10: 512–526

Tarayre M, Thompson JD (1997) Population genetic structure of the gynodioecious *Thymus vulgaris* L. (Labiatae) in southern France. *Journal of Ecology and Evolution* 10: 157–174

Touzet P, Delph LF (2009) The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics* 181: 631–644

Tsujimura M, Mori N, Yamagishi H, Terachi T (2013) A possible breakage of linkage disequilibrium between mitochondrial and chloroplast genomes during Emmer and Dinkel wheat evolution. *Genome* 56: 187–193

Wade MJ, McCauley DE (2005) Paternal leakage sustains the cytoplasmic polymorphism underlying gynodioecy but remains invasible by nuclear restorers. *The American Naturalist* 166 (5): 592–602

Webb CJ (1999) *Empirical Studies: Evolution and Maintenance of Dimorphic Breeding Systems* (MA, Geber, TE Dawson & LF Delph, eds), pp. 61–95. *Gender and Sexual Dimorphism in Flowering Plants*. Springer, Berlin, Heidelberg

Welch ME, Darnell MZ, McCauley DE (2006) Variable populations within variable populations: quantifying mitochondrial heteroplasmy in natural populations of the gynodioecious plant *Silene vulgaris*. *Genetics* 174: 829–837

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058

SUPPLEMENTARY INFORMATION

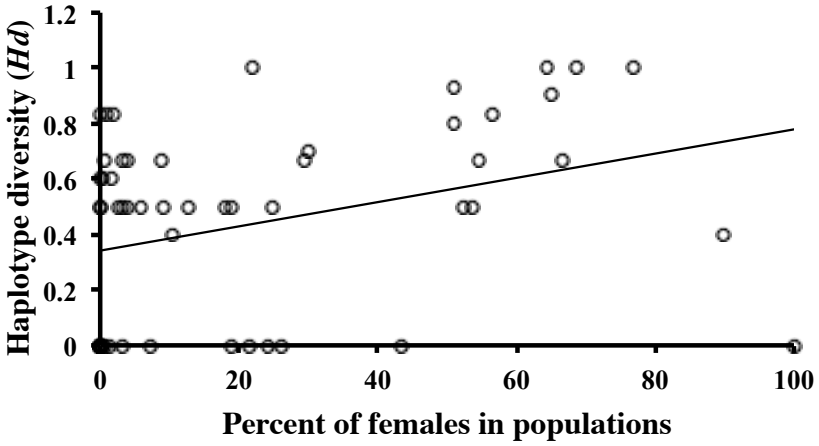


Figure 2.S1 Spearman's rank correlation between 3-locus mitotypes and population sex ratio (percent female) among 56 populations (N=220).

Table 2.S1 Distribution, sex ratio (percent female), and sample information of populations used in the current study. Percent females were averaged if data from multiple years were available (sampling period between 1999 and 2013). Total sample (N = 266) included 87 females, 177 hermaphrodites, and 2 plants of undetermined sex.

State/	Population			Sex ratio	No. of	No. of successful sequences				
Province	code	Latitude (N)	Longitude (W)	(% females)	samples	*plastid*	*atp6*	*Ψrps12*	*nad7ab*	*3-locus mt*
IA	CDI	41.69	-93.72	30.1	5	5	5	5	5	5
	CERA	41.68	-92.87	3.4	4	4	4	3	4	3
	FW	41.57	-92.57	24.7	4	4	4	4	4	4
	KR	41.71	-92.79	9.0	4	4	4	4	4	4
	RT	41.71	-92.86	4.0	4	4	4	3	4	3
IL	BC	37.31	-88.53	100.0	4	4	4	3	4	3
	COL	39.41	-88.09	76.9	4	4	3	3	4	2
	CR	39.42	-88.10	68.8	6	6	4	5	6	3
	CRS	39.23	-89.80	1.3	4	4	4	4	4	4
	FC	37.38	-89.08	54.5	3	3	3	3	3	3
	HH	41.21	-89.32	26.3	5	5	3	3	5	2
	ILCR	40.31	-91.04	53.6	5	5	4	4	5	4
	JPH	37.63	-89.34	86.7	4	4	1	4	3	0
	LP	39.43	-88.43	0.0	2	2	2	1	2	1
	RTW	42.19	-88.01	19.0	4	4	4	4	4	4
	SRP	42.50	-89.25	8.7	4	4	4	4	4	4
	WIH	37.53	-89.12	0.0	4	4	4	4	4	4
IN	FRD	39.14	-86.40	21.9	5	5	4	3	5	2
	GT	40.73	-86.50	0.0	4	4	3	3	4	2
	MR	40.79	-85.21	43.6	4	4	4	2	4	2
	NH	38.10	-87.95	56.6	5	5	4	5	5	4
	STW	39.13	-86.38	83.3	5	5	2	2	5	1
	WN	40.89	-85.00	66.7	4	4	3	4	4	3
	YW	39.22	-86.34	65.0	7	7	7	7	7	7
KY	ANG	37.49	-84.23	0.0	4	4	4	4	4	4
	CAC	38.89	-84.37	0.0	5	5	4	5	5	4
	GLA	37.04	-85.92	17.9	5	5	4	5	5	4
	KY	37.45	-84.31	90.0	5	5	5	5	5	5
	SFL	37.39	-84.67	51.0	5	5	5	5	5	5
	TIL	37.87	-87.51	64.4	4	4	3	3	4	2

Contd...

State/ Province	Population code	Latitude (N)	Longitude (W)	Sex ratio (% females)	No. of samples	No. of successful sequences				
						plastid	*atp6*	*Ψrps12*	*nad7ab*	*3-locus mt*
MA	BC	42.06	-73.35	3.3	4	4	4	4	4	4
MI	MS	42.77	-85.36	0.0	5	5	5	5	5	5
	OS	42.60	-85.39	0.0	4	4	4	4	4	4
MN	BT	44.73	-94.40	19.0	5	5	5	5	5	5
	CCFC	45.40	-93.16	0.0	4	4	4	4	4	4
	SF	44.77	-93.37	0.0	4	4	4	4	4	4
	TKF	44.06	-95.30	0.0	4	4	4	4	4	4
MO	ELS	39.13	-90.77	2.7	4	4	4	4	4	4
	ELSII	39.10	-90.75	5.9	4	4	4	4	4	4
	MAM	38.12	-90.67	12.9	5	5	5	4	5	4
	NL	39.61	-91.40	0.0	4	4	4	4	4	4
OH	BVI	41.45	-83.79	52.2	5	5	5	4	5	4
	BVII	41.45	-83.79	21.5	4	4	4	4	4	4
	CP	41.54	-83.84	0.5	4	4	4	4	4	4
	HR	39.51	-84.72	50.9	6	6	6	6	6	6
	KYR	40.61	-81.60	24.3	4	4	4	4	4	4
	MB	41.17	-81.20	2.1	4	4	4	4	4	4
	PS	41.64	-83.43	41.6	4	4	2	1	4	1
	STS	40.49	-81.98	1.6	5	5	5	5	5	5
	WBC	41.56	-83.85	0.7	4	4	4	3	4	3
ON	MRA	43.23	-80.01	29.3	5	5	4	4	5	4
	TJP	43.37	-81.00	3.4	4	4	4	4	4	4
	WL	42.58	-81.64	0.9	4	4	4	4	4	4
VA	CHE	37.21	-79.56	4.0	4	4	4	4	4	4
	FIN	37.67	-78.80	0.0	4	4	4	4	4	4
	MOO	38.57	-78.27	0.0	3	3	3	3	3	3
	PVR	37.55	-79.56	0.0	4	4	3	4	4	3
WI	PVC	43.11	-89.81	4.9	4	4	3	2	4	1
WV	BB	37.92	-80.27	10.5	6	6	5	5	6	5
	BUC	39.53	-79.64	7.2	4	4	4	4	4	4
	SD	38.14	-81.16	0.3	5	5	5	5	5	5

Standard abbreviations for US states and the Canadian province are used. Percent females = females/(females+ hermaphrodites). 3-locus mt = multilocus mitochondrial (concatenated) sequences

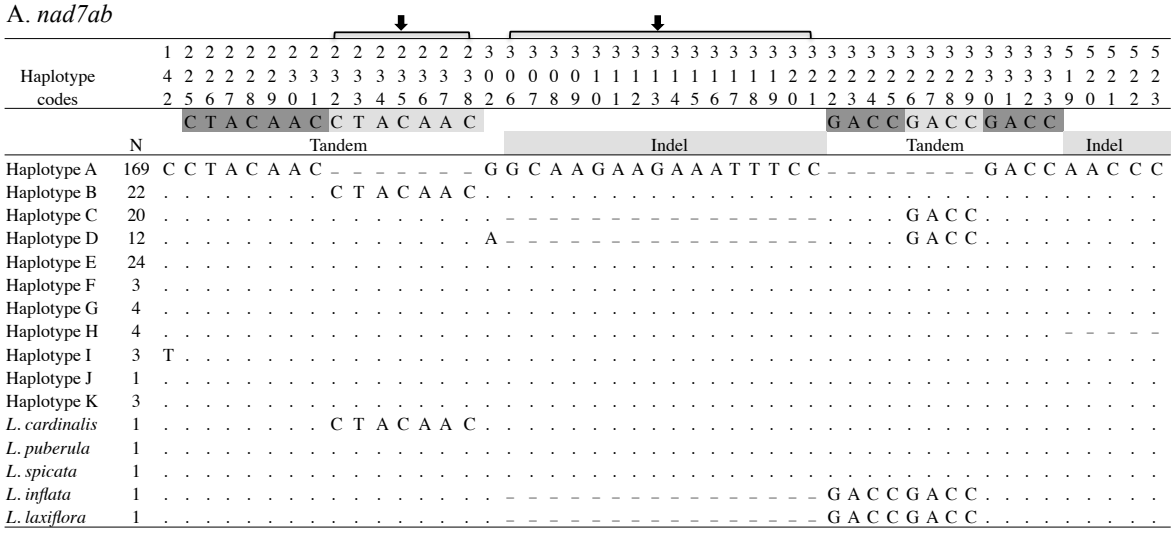
Table 2.S2 Primers used in the current study, including annealing temperatures and the product sizes.

Ta= annealing temperature.

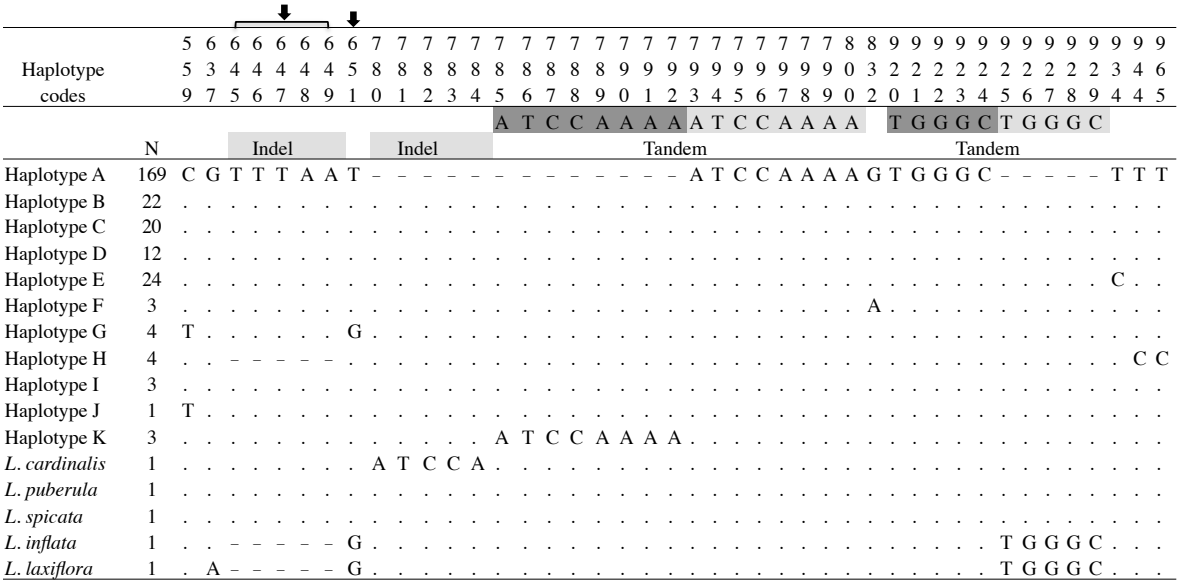
Marker	Direction	Sequence	Ta (˚C)	Product size (bp)	Reference
Plastid					
426	F1	CAAATAAGAGTATATGCACGAATAGC	53	872-1014	HJ Madson (2012)
297	R1	AAAAAAGCATAGGCCTCGGG			
427	F2	GCAACGATTYGATAAGCCGC	53	1299-1825	HJ Madson (2012)
331		CCGTTCCGGTGTTGCCCTACC		573-1099	(alternative if 427 did not work)
15		AGTCATTGGTTCAGTCGGTA		1178-1704	(alternative if 427 and 331 did not work)
425	R2	GGGTTTTTGAAGTTCCATCGG			
330		GGTACGGGCTATTGCCGCTGG		643-682	(alternative if 425 did not work)
Mitochondrial					
nad7ab	F	ACCTCAACATCCTGCTGCTC	52	1049	Dumolin-Lapegue et al. (1997); *nad7/1-2*
	R	CGATCAGAATAAGGTAAAGC			
atp6	F	ACTCGTACAGGAAGGACTCTC	52	808	Designed for this study
	R	CAACTTTGATGGAGATTTGTAGC			
Ψrps12	F	CGGATCGGGAGTAACCACTA	56	413	Designed for *nad3*, *Mimulus guttatus* (AL Case)
	R	TCCAGAGGCATCTTCCATTC			

Table 2.S3 Variable sites present in unique alleles/haplotypes found in individual markers. The base positions start from the start codon in *atp6* (B) and after the forward primers in *nad7ab* (A) and *Ψrps12* (C). Outgroup sequences are not included for *psbK-rps16* (D) because even the closest outgroup species *L. cardinalis* and *L. puberula* had much shorter sequences than *L. siphilitica,* resulting in large alignment gaps. Potentially homoplasious mutations are marked by arrows. Indels are considered as single nucleotide insertion or deletion events for *psbK-rps16.*

A. *nad7ab*



Contd..



B. *atp6*

	N	251	273	304	412	427	430	439	408	437	456	461
Haplotype 1	109	C	G	C	C	G	G	C	C	G	T	C
Haplotype 2	12	.	.	.	.	.	.	.	.	T	.	.
Haplotype 3	61	.	.	.	.	.	.	.	.	T	.	A
Haplotype 4	21	.	.	A	.	.	.	.	.	.	.	.
Haplotype 5	7	G	.	.	.	.	.	.	.	.	.	A
Haplotype 6	20	G	.	.	.	.	.	.	.	T	.	A
Haplotype 7	1	G	.	.	T	.	A	.	.	.	.	A
Haplotype 8	1	G	.	.	.	.	.	.	A	T	C	A
Haplotype 9	7	G	.	.	.	.	.	.	.	T	.	.
Haplotype 10	1	.	A	.	.	.	.	.	.	.	.	.
Haplotype 11	1	.	.	.	.	A	.	T	.	.	.	.
L. cardinalis	1	.	.	.	.	.	.	.	.	.	.	.
L. puberula	1	.	.	.	.	.	.	.	.	.	.	.
L. spicata	1	G	.	.	.	.	.	.	.	.	.	.
L. inflata	1	.	.	.	.	.	.	.	.	.	.	A
L. laxiflora	1	.	.	.	.	.	.	.	.	T	.	A
Amino acid sequence		S	P	N, K	P, S	G, E	V, I	Q, Stop	L	S	C	F, L

C. *Ψrps12*

	N	174	1Indel	143	144	145	152	154	155	229	238	290	380
Haplotype a	216	T	G	C	C	C	A	G	G	C	T		
Haplotype b	1	.	.	.	.	.	.	.	.	.	C		
Haplotype c	4	.	.	.	.	.	G	.	.	T	.		
Haplotype d	14	.	.	.	.	.	.	T	.	.	.		
Haplotype e	2	A	.	.	.	.	G	.	.	T	.		
Haplotype f	1	.	.	.	.	.	.	.	C	.	.		
L. cardinalis	1	.	.	.	.	.	.	.	.	.	.		
L. spicata	1	.	–	T	T	A	.	.	.	.	.		
L. inflata	1	.	.	.	.	.	.	.	.	.	.		
L. laxiflora	1	.	.	.	.	.	.	.	.	.	.		

Contd..

D. *psbK-rps16*

(plastid)	Indel	Indel	Indel	Indel	Indel	Indel	Indel	Indel	Indel	Indel	Indel
Haplotype C10	C T C G T C A C - - A - G T G A A C A T T G C A C A T A G G G G C T A C C A T G C C C G										
Haplotype C18	. . G . . . . . . T . T G A . G . C . - - G C A . . C A A C C T G C G G . . . . A G .										
Haplotype C14	. . G - . . . . . T . T G A C G . C . - - G C A . . C A A C C T G C G G . . . . A . -										
Haplotype C8	. . . . . C . . . . . G T . . . . A G .										
Haplotype C9	. . . . . A . .										
Haplotype C15b	. G G . . . . . A T . T G A C G . C . - - G C A - C C A A C C T G C G G . . . . A - -										
Haplotype C21	. . G . . . . . T . T G A . G . C . - - G C A . . C A . A C T G C G . G . . . A . .										
Haplotype C15	. . G . A . . . . A T . T G A C G . C . - - G C A - C C A A C C T G C G G . . . . A - -										
Haplotype C25	. . G A . . . . . T . T G A . G . C . - - G C A . . C A . A C T G C G . G . . . A . .										
Haplotype C2	T . G . . . C - A . . . T G C C G . . . . A G . . . . C . . . C T G C G A . G A T A . .										
Haplotype C19	. . G . . . . . . T . T G A . G . C - . . G C A . . C A . A C T G C G . G . . . A . .										
Haplotype C24	. . G . . T . . . . T . T G A . G T C - - G C A . . C A . A C T G C G . G . . . A . .										
Haplotype C5	. . G . . . . - - - . . . . . G . . . . . A . .										

Indels are considered as single nucleotide insertion/deletion events in plastid data

Table 2.S4 Values of standardized linkage disequilibrium (D′) between pairs of 'common' alleles/haplotypes between: A) plastid and 3-locus mitotypes; B-D) plastid haplotypes and individual mitochondrial marker alleles. 'common'= found in ≥ 5 individuals and ≥ 3 populations. Alleles and haplotypes are ordered by decreasing frequencies. The alleles/haplotypes match those in Figure 2.2.

A.

plastid–3-locus mt	**A1a**	**A3a**	**B1a**	**A4a**	**A6a**	**A2a**	**C3a**	**D6a**
C8	0.16	-0.67	-0.31	-0.32	-1.00	0.01	-0.27	-1.00
C21	0.00	0.03	0.06	0.15	-0.48	-0.25	0.01	0.26
C14	-0.06	0.06	-1.00	-0.61	0.76	-1.00	0.05	0.05
C9	-0.20	0.05	-1.00	0.15	-1.00	0.35	-1.00	-1.00
C10	-0.06	-0.70	-0.53	-0.39	-0.02	-1.00	0.33	0.33
C25	-0.86	0.31	0.47	-1.00	-1.00	-1.00	-1.00	-1.00
Ave. |D′|= 0.49								

B.

***nad7ab*–plastid**	**C8**	**C21**	**C14**	**C9**	**C10**	**C25**
A	-0.07	0.14	0.34	0.33	-0.28	-0.35
B	-0.42	0.06	-1.00	-1.00	-0.54	0.50
C	0.20	-0.45	0.17	-1.00	0.00	-0.45
E	0.33	-0.77	-0.21	-0.62	0.08	-1.00
D	-1.00	-0.08	-0.47	0.25	0.35	-1.00
Ave. |D′|= 0.45						

C.

***atp6*–plastid**	**C8**	**C21**	**C14**	**C9**	**C10**	**C25**
1	0.19	-0.04	-0.26	-0.37	-0.02	0.17
3	-0.28	-0.19	0.11	0.19	-0.37	0.27
4	0.05	0.07	-0.71	0.08	-0.54	-1.00
6	-0.50	-0.18	0.46	-1.00	0.05	-1.00
2	-0.44	0.19	-1.00	0.34	-1.00	-1.00
5	0.39	-1.00	-1.00	-1.00	0.36	-1.00
9	-0.52	0.48	-1.00	-1.00	0.04	-1.00
Ave. |D′|= 0.5						

D.

***Ψrps12*–plastid**	**C8**	**C21**	**C14**	**C9**	**C10**	**C25**
a	-0.25	0.76	-0.08	-0.02	0.20	1.00
d	0.30	-0.60	-0.11	0.11	-0.40	-1.00
Ave. |D′|= 0.4						

Table 2.S5 Values of standardized linkage disequilibrium (D′) between pairs of 'common' mitochondrial alleles. 'Common' = found in ≥ 5 individuals and ≥ 3 populations. Alleles are arranged in order of decreasing frequency. The alleles/haplotypes match those in Figure 2.2.

A.

nad7ab–atp6							
	1	**3**	**4**	**6**	**2**	**5**	**9**
A	0.05	-0.07	1.00	-0.11	-0.01	-0.36	-0.15
B	1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
C	-0.61	0.61	-1.00	-1.00	-1.00	0.08	-1.00
E	-0.17	-0.26	-1.00	0.05	0.02	-1.00	0.23
D	-1.00	0.11	-1.00	0.36	-1.00	0.40	-1.00
Ave. |D′|= 0.62							

B.

nad7ab–Ψrps12		
	a	**d**
A	0.18	-0.03
B	1.00	-1.00
C	-0.10	-1.00
E	-0.27	0.30
D	1.00	-1.00
Ave. |D′|= 0.59		

C.

atp6–Ψrps12		
	a	**d**
1	0.67	-0.81
3	-0.16	0.03
4	-0.03	-1.00
6	-0.03	0.11
2	-0.09	0.14
5	-0.22	0.25
9	1.00	-1.00
Ave. |D′|= 0.4		

Table 2.S6 Distribution of 39 three-locus mitochondrial haplotypes among 225 individuals and 61 populations. Concatenated haplotype names indicate allelic combinations of individual markers *nad7ab* (A-K), *atp6* (1-11), and *Ψrps12 (*a-f) respectively. Dashes in haplotype names (e.g., E9-) indicate missing sequences. Haplotypes could be inferred for 5 individuals despite missing sequences while 41 samples remained undetermined. The alleles/haplotypes match those in Figure 2.2.

3-locus mt haplotype	No. of samples	No. of populations	3-locus mt haplotype	No. of samples	No. of populations
A1a	73	33	I1a	2	1
A3a	33	16	A1e	1	1
B1a	21	8	A2d	1	1
A4a	16	9	A4b	1	1
A6a	10	5	A4e	1	1
A2a	7	5	A5a	1	1
C3a	5	4	A6d	1	1
D6a	5	4	C10a	1	1
D3a	4	1	C5a	1	1
E1a	4	3	E11d	1	1
A9a	3	1	E1c	1	1
C1a	3	2	E2d	1	1
C3c	3	1	E6a	1	1
D5a	3	1	E6d	1	1
E3a	3	1	E8-	1	1
H2a	3	1	F6-	1	1
K3a	3	2	G1-	1	1
A3d	2	2	G9a	1	1
A5d	2	2	J3f	1	1
E9-	2	1	-	-	-

CHAPTER III. DIVERSITY AND COMPLEXITY OF SEX DETERMINATION IN GYNODIOECIOUS *LOBELIA SIPHILITICA* L. (CAMPANULACEAE)

ABSTRACT

Cytoplasmic male sterility (CMS) represents an interesting example of cytonuclear incompatibility in flowering plants. Mitochondrial CMS genes result in the loss of pollen production, yielding female phenotype unless accompanied by matching nuclear restorer(s) that suppress CMS. Theoretical models propose that within-population CMS polymorphism (and that of specific restorers) may result in the expression of male sterility, a.k.a. gynodioecy—the co-existence of females and hermaphrodites within a population. However, mechanisms maintaining CMS polymorphism within gynodioecious populations are not clearly understood. Also, the cause(s) of highly variable female frequencies among populations of these some gynodioecious species remain to be identified. Here, I conducted an extensive crossing experiment using 30 maternal families from 12 populations of gynodioecious *Lobelia siphilitica* to assess the diversity of CMS genes within and across populations to evaluate evidence that long-term maintenance and/or frequent-replacement dynamics has contributed to CMS polymorphism, and to examine potential causes of variable female frequencies among populations. I found at least six CMS types within *L. siphilitica*, with five populations having multiple (two to four) CMS types. Interestingly, CMS types identified from crosses corresponded to mitotypes defined by

mitochondrial markers. Some of the CMS types appear to be widely distributed among populations, indicating that they may have been maintained within the species for a long time. Other CMS types, however, appear to be younger because they had restricted geographic distribution, and nuclear restorers for these CMS types were rare or absent. Also, while the genetic models of restoration for some CMS types fit Mendelian expectations, other CMS types required complex restoration genetics to explain offspring sex ratios. In sum, the long-term maintenance of multiple CMS types could result in gynodioecy in *L. siphilitica*, while high CMS diversity, the frequent formation of novel CMS types, and/or complex restoration genetics could contribute to high female frequencies in populations.

Keywords: cytoplasmic male sterility, male fertility restoration, sex ratio, *Lobelia siphilitica*

INTRODUCTION

Proper functioning of eukaryotes requires precise coordination between nuclear and cytoplasmic (mitochondria and plastid) functional genes. Changes in one partner should be accompanied by compensatory changes in the other in order to maintain compatibility (Frank 1989; Burt & Trivers 2006; Budar & Roux 2011). However, some of the cytoplasmic genes evolve at a higher rate, making it difficult for the nuclear genome to compensate the changes, leading to cytonuclear incompatibility with significant phenotypic consequences, such as loss or reduction in fertility or viability (Rand *et al.* 2004; Sambatti *et al.* 2008; Johnson 2010). Flowering plants with gynodioecy represent the natural expression of a well-studied cytonuclear incompatibility and thus provide an ideal opportunity to understand mechanisms that maintain the associated genetic polymorphisms (Eckardt 2006; Fishman & Willis 2006).

Gynodioecy is a dimorphic sexual system in which male-sterile (female) and male-fertile (hermaphroditic) individuals co-occur within natural populations (Darwin 1877; Renner & Ricklefs 1995). This most often results when cytoplasmic male sterility (CMS) genes located in mitochondrial genomes interfere with pollen production in some individuals, making them functionally female (Correns 1906). CMS genes are widespread among flowering plants (Laser & Lerston 1972; Carlsson *et al.* 2008; Gobron *et al.* 2013). However, in most angiosperms, CMS genes are 'cryptic' because compatible nuclear restorer alleles that suppress their action spread to fixation within populations, restoring pollen production and hermaphroditism (Richards 1997; Geber *et al.* 1999; Case *et al.* 2016). CMS phenotypes are expressed in nature when compatible nuclear restorers are not fixed within populations, forming gynodioecy. Gynodioecy has been documented in less than 1% of flowering plant species (Godin & Demyanova 2013), but this relatively small number of species represents ca. 21% of families, indicating that gynodioecy has evolved numerous times in diverse angiosperm lineages (Caruso *et al.* 2016).

Theoretical models predict that polymorphism of CMS genes and nuclear restorers are required for cytonuclear gynodioecy. However, models differ fundamentally regarding the specific mechanisms that maintain CMS polymorphism. Predictions about the initial steps of CMS evolution are common to all genetic models. Specifically, a new CMS gene invades and spreads 'selfishly' through a population because it is maternally inherited and feminizing (Burt & Trivers 2006). Females, being obligatory outcrossers, could gain further fitness benefit by avoiding inbreeding depression that hermaphrodites may suffer when they inbreed (Thompson & Tarayre 2000; Delph 2004; Chang 2007). However, females suffer from pollen limitation when CMS genes (and thus females) become too frequent (McCauley & Taylor 1997; Frank & Barr 2001; Alonso 2005; Zhang *et al.* 2008). Thus, the high female frequency that results from the

invasion and spread of a novel CMS gene should create strong selection for a compatible nuclear restorer, which is then expected to spread through the population, restoring male fertility. At this point, models of gynodioecy differ in their assumption of how CMS polymorphism is maintained over the longer term.

According to a first type of models, nuclear restorers are expected to sweep to fixation where a target CMS gene is present because of the fitness benefits of producing pollen. If so, all plants in the population should eventually be restored hermaphrodites that carry a common CMS type (e.g., Case *et al.* 2016). Under these models, gynodioecy will only be observed when another CMS type invades, and only until another restorer sweep occurs and restores monomorphic hermaphroditism. Over time, multiple CMS types could periodically invade a population and replace existing CMS types in an "epidemic" fashion (Frank 1989; Couvet *et al.* 1998). But at any point in time, there should be few CMS types maintained within populations and gynodioecy should be transient (McCauley & Bailey 2009). According to a second type of models, a newly selected restorer would not necessarily spread to fixation. Instead, its spread would slow once it reached a reasonably high frequency (Charlesworth 1981; Gouyon *et al.* 1991; Bailey *et al.* 2003). This is expected to occur whenever hermaphrodites carrying excess or 'silent' restorers (that is, restorers not paired with their target CMS type) suffer a fitness cost (reviewed in Charlesworth 1981; Delph *et al.* 2007). In this "cost of restoration" scenario, restorers experience a type of balancing selection called negative frequency-dependent selection, and cycle through higher and lower frequencies without going to fixation or being eliminated from the population (reviewed in Delph & Kelly 2014). Consequently, multiple CMS types and restorers can be maintained within a population at intermediate frequency over a long period of time. Finally, according to a recent 'mixed' model, a fertile (non-CMS) cytotype could be

selectively maintained in a population for a long period of time while multiple CMS types competitively replace one another by partial selective sweeps (McCauley & Bailey 2009).

Distinguishing among alternative models of CMS–restorer polymorphism is difficult because multiple CMS types could co-occur within a population at a given point in time under any of these models. One could, however, potentially distinguish among these models based on the time of maintenance of individual CMS types within a population. Balancing selection associated with a cost of restoration predicts long-term maintenance of individual CMS types. Thus, CMS types are expected to be relatively older than the CMS types in populations experiencing frequent replacement (epidemic dynamics). By contrast, CMS types are expected to be younger if they are formed *de novo* by mitochondrial recombination (Hanson & Bentolila 2004).

Inferring relative times of maintenance of CMS types is challenging because the genes themselves have not been identified in species with natural CMS–restorer incompatibility. Also, because CMS genes are chimeric, containing sequences of multiple known or unknown sources, inferring relative times of origin by using a 'molecular clock' is challenging (Bromham & Penny 2003; Kumar 2005; Zuckerkandl & Pauling 1965). Therefore, an alternative strategy to infer relative time of occurrence of a CMS type is to study their geographic distribution. That is, CMS types that are widespread are expected to be relatively older then those that are restricted in distribution unless gene flow is limited.

In addition, the frequency of females is highly variable among natural populations of some gynodioecious species, which sometimes reaches an extreme (e.g., *Lobelia siphilitica*, 0–100% females, Caruso & Case 2007 and unpublished data). In theory, high female frequencies within populations could be caused by a high diversity of CMS genes (Delph & Kelly 2014).

High CMS diversity could potentially increase the mismatches between CMS types and compatible nuclear restorers, thus increasing the number of females in a population. In addition, high female frequency could also be caused if some CMS types are difficult to restore either because they are new for the population such that compatible nuclear restorers are rare or absent (Frank 1989), or because they have complex genetics of restoration (Ehlers *et al.* 2005; Bailey & Delph 2007).

In cases where the CMS genes have not been characterized, crossing studies are useful tools to infer CMS types (and nuclear restorers) present in parental plants based on progeny sex ratios. Previous crossing studies have reported CMS polymorphism within gynodioecious species, with up to four CMS types at species level (e.g., de Haan *et al.* 1997b; van Damme *et al.* 2004) and up to three at population level (e.g., Damme *et al.* 2004; Dufay *et al.* 2009). However, these data are not sufficient to test whether the CMS types are maintained within a species for a long period of time (balancing selection) or they are frequently replaced (epidemic dynamics). Also, empirical evidence for fertile cytotypes has been reported in only two gynodioecious species (*Plantago lanceolata*, de Haan *et al.* 1997b; *Beta vulgaris* ssp. *maritima*, Dufay *et al.* 2009). Thus, it is unclear whether fertile cytotypes are more common components of the cytoplasmic polymorphism associated with gynodioecy. Moreover, while previous studies have found evidence of complex genetics of restoration of male fertility in gynodioecious species (e.g., Belhassen *et al.* 1991; Koelewijn & van Damme 1995; Charlesworth & Laporte 1998; Dudle *et al.* 2001; van Damme *et al.* 2004; Garraud *et al.* 2011), the association of complex restoration genetics and high female frequencies in populations has not been thoroughly explored.

In this study, I used an extensive crossing experiment using 30 maternal families

representing 12 geographically widespread populations of the gynodioecious wildflower *Lobelia siphilitica* L. (Campanulaceae) that vary widely in frequencies of females to observe CMS polymorphism, to test models that maintain CMS polymorphism, and to find potential causes of very high female frequencies some in populations. First, I assessed the level of CMS polymorphism within and among populations of *L. siphilitica* and tested for evidence of fertile cytotypes. Previous crossing studies based on two populations of *L. siphilitica* have identified three CMS types including two CMS types from a high-female population (Dudle *et al.* 2001; Bailey 2002). However, recent studies based on mitochondrial markers found a much higher diversity of mitochondrial haplotypes (hereafter 'mitotypes') both within and across populations of this species (Delph & Montgomery 2014; Chapter II). Because CMS genes are located in mitochondrial genomes, I expected to find a comparable level of CMS diversity as shown by mitotypes.

Second, I asked whether the CMS polymorphism in this species is caused by long-term maintenance of CMS types by balancing selection or by frequent invasion and periodic replacement of novel CMS types (epidemic dynamics). Because actual CMS genes have not been identified in *L. siphilitica*, we could not directly determine their ages. Thus, I used a proxy strategy to understand times of occurrences of CMS types based on their distribution: a widespread geographic distribution indicates that CMS types could have been maintained within the species for a long time (balancing selection). This is especially true given very low levels of cytoplasmic gene flow in this species (mitochondrial $F_{ST} = 0.65$, Chapter II; plastid $F_{ST} = 0.75$, Madson 2012). By contrast, a localized distribution could indicate that CMS types are frequently formed and/or periodically replaced (epidemic dynamics). The marker-based study suggested that some of the mitotypes in this species could to be geographically widespread while others

could be very restricted in distribution (Chapter II). Here, I tested this pattern by using CMS types inferred by controlled crosses.

Finally, I evaluated potential causes of highly variable female frequencies among populations of *L. siphilitica.* Specifically, I examined whether high diversity of CMS genes, the frequent formation of novel CMS types and/or complex restoration genetics could cause high female frequencies in populations. Because the marker-based study found a positive correlation between mitotype diversity and female frequencies in populations (Chapter II), I expected to find a higher number of CMS types in high- than low-female populations. Also, given the evidence of complex genetics of restoration found for a CMS type in this species (Dudle *et al.* 2001), I expect to find additional evidence of this phenomenon in my detailed study.

MATERIALS AND METHODS

The study species

Lobelia siphilitica L. (Campanulaceae) is a gynodioecious wildflower native to eastern North America, usually growing in moist habitats (Johnston 1991). Each plant can produce dozens of blue flowers, ca. 3-cm-long, on an apical spike. Anthers and filaments are fused together forming a tube that covers the pistil (Crowl *et al.* 2016). The color of the anther tubes differs between the sexes, as the flowers are dark-purple with fertile pollen in perfect flowers and whitish with no pollen in pistillate flowers. The species is self-compatible; however, because perfect flowers are protandrous, opportunities for autonomous selfing are low (Johnston 1992). Primary pollinators of *L. siphilitica* are *Bombus* species (Beaudoin Yetter 1989). Sex-

determination in this species is cytonuclear (Dudle *et al.* 2001; Bailey 2002), and sex ratios of natural populations vary from 0 to 100% females (Caruso & Case 2007 and unpublished data).

Pollination procedure

I performed all pollinations in the Herrick Conservatory greenhouse at Kent State University, Kent, Ohio in fall 2012. I emasculated perfect flowers designated to be dams at or just before anthesis by removing anther cylinders using a pair of clean fine forceps. To prevent pollen contamination, I placed dome caps made of Parafilm over the emasculated flowers prior to pollination. The stigmas of *Lobelia* flowers are bi-lobed and only receptive when the two lobes separate to expose stigmatic papillae. Stigmas of perfect flowers were receptive 1–2 days after emasculation and those of pistillate flowers were receptive 1–2 days after anthesis. I placed pollen from a donor flower on a glass slide by squeezing the anthers with forceps. I saturated the stigmas of the recipient flowers, replaced the dome caps, and removed them after the stigmas lost their receptivity (turned brown), usually one day after pollination. I rinsed the slide and the forceps thoroughly with 70% ethanol and dried between each cross. I collected leaf tissue samples from each parent for haplotyping.

Sampling of populations and families

In order to better assess the diversity and distribution of CMS genes (and corresponding restorers) in this species, I chose 12 populations spanning several hundred kilometers across four U.S. states that varied dramatically in female frequency (Table 3.1). In order to uncover as many CMS types as possible, I chose to sample a large number of high-female populations: 10 of the 12 populations I studied had > 20% female. I used a total of 30 maternal families, 1–5 families per population. Each family represented descendants of open-pollinated seeds collected from a single maternal plant in the wild.

Table 3.1 Geographic locations and sex ratios of 12 populations of *Lobelia siphilitica* used in the current study. Standard U. S. state acronyms are used. Population sex ratios are percent female obtained by averaging census data from multiple years (when available).

U. S. states	Population code	Latitude	Longitude	Population sex ratio (% female)	IDs of maternal families
OH	BV	41.449	-83.788	21.5	2, 6, 8
	PS	41.641	-83.434	41.6	1, 4, 6, 8
	HR	39.508	-84.716	50.9	5, 6
IN	YW	39.219	-86.343	65.0	2, 3, 4, 8, 9
	Y2	39.165	-86.344	32.2	3
	Y3	39.218	-86.341	55.7	2, 3
	Y4	39.174	-86.341	37.9	1, 3
	MR	40.791	-85.207	43.6	2, 4
IL	CR	40.314	-91.043	53.6	2, 3, 5
IA	CD	41.689	-93.72	30.1	3, 4
	KR	41.705	-92.786	9.0	2
	RE	41.705	-92.864	4.0	2, 3, 6

I chose focal maternal families based on the results of a preliminary crossing experiment (Case AL & Caruso CM, unpublished data), which identified three distinct sex segregation patterns: i) families that produced only female progeny with multiple sires, raising the possibility that restorers for their CMS types are rare or absent; ii) families that never produced any female progeny, possibly because their restorers are common or they have fertile cytotypes; iii) families that segregated progeny of both sexes, suggesting that their CMS types and matching restorer alleles were polymorphic within or across populations.

Crossing design

I performed a total of 100 cross pollinations, and 23 self pollinations (hermaphrodites only). Cross-pollinations were split between two types—nested for female dams and reciprocal for hermaphrodite dams. Maternal families were assigned to cross type based on sex-segregation patterns. My goal was to maximize the chance of producing progeny of both sexes, because such crosses are more informative about sex-determination than those producing single-sexed progenies.

Nested one-way crosses were made between a common hermaphrodite sire and multiple female dams (Figure 3.1A). My prediction from nested crosses was that female dams having different CMS types should produce significantly different progeny sex ratios when pollinated by a common sire (Dudle *et al.* 2001). I used 18 maternal families from six populations as nested dams. I chose dams from families that segregated sufficient hermaphrodite progeny in preliminary crosses. I used seven hermaphrodites from seven families and five populations as common sires. Four sires were from within the dam families but three were from new families. I used hermaphrodite sires from multiple populations in order to increase the chances of

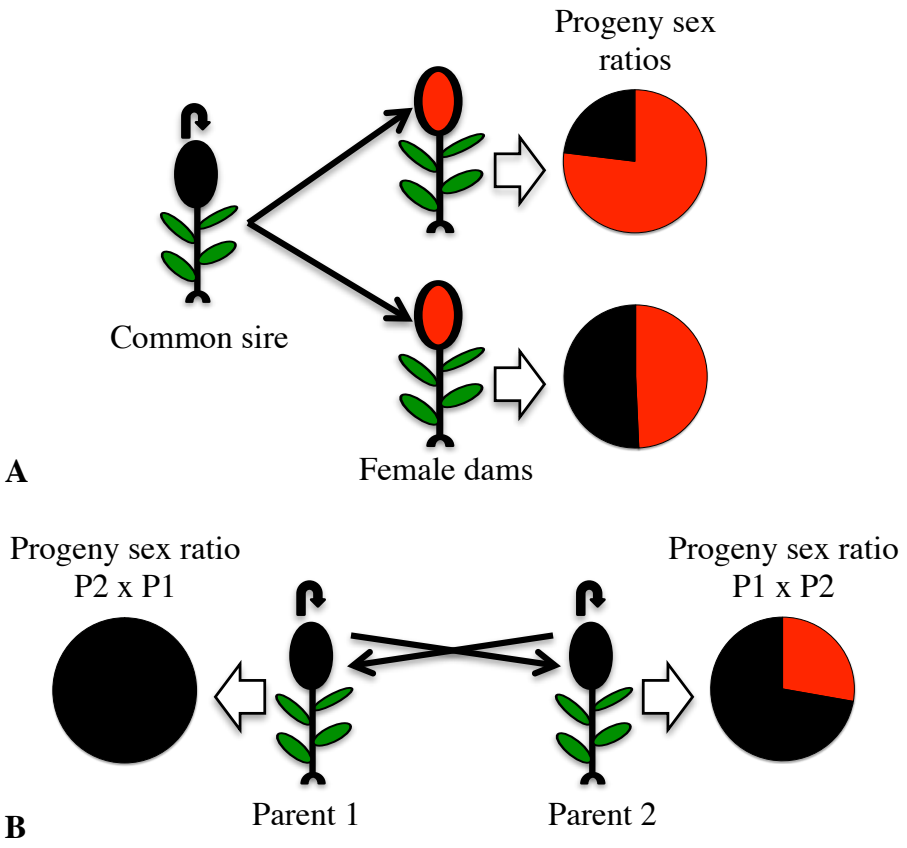


Figure 3.1 Schematic showing each type of crossing design. A. Nested one-way cross between a common hermaphrodite (black) sire and two female (red) dams. B. Reciprocal crosses between pairs of hermaphrodites. Black arrows indicate transfer of pollen. Pie charts show hypothetical progeny sex ratios. This result would indicate that the two females (in A) or two hermaphrodites (in B) do not share the same CMS type as each other. Curved arrows indicate self pollinations of hermaphrodites: if self pollinations produced any female progeny, the parent was assumed to carry a CMS gene.

obtaining heterogeneous sex segregations useful to identify the most CMS types. Although I could not practically cross each sire with each dam, I obtained a total of 60 nested crosses.

In **reciprocal crosses** between pairs of hermaphrodites, pollen from one hermaphrodite was used to fertilize a second hermaphrodite and vice versa (Figure 3.1B). My prediction from reciprocal crosses was that two hermaphrodites with different CMS types should produce significantly different reciprocal progeny sex ratios (van Damme 1983; Koelewijn & van Damme 1995; de Haan *et al.* 1997b). I used 16 hermaphrodites from 13 maternal families and 10 populations to carry out 20 reciprocal crosses (a total of 40 sets of progenies). Reciprocal parents were made using families that never or rarely produced female progeny in preliminary crosses. I opted for inter-population crosses to maximize the chances of obtaining female progeny because geographic separation could increase chances of CMS–restorer mismatching (Belhassen *et al.* 1991; Gigord *et al.* 1998; but see Emery & McCauley 2002; Bailey & McCauley 2005). Inter-population crosses would also be useful for understanding CMS-sharing among populations. Similar reciprocal progeny sex ratios, however, do not necessarily indicate that the parents share the CMS types because different CMS types with similar level of restoration by respective restorers can yield such similarly (van Damme *et al.* 2004).

Finally, I performed **self pollinations** of all the hermaphrodites used in this study (N=23; Figure 3.1). A hermaphrodite that produced any female progeny upon selfing was assumed to carry a CMS gene and at least one non-restoring allele. Otherwise, the hermaphrodite could be either highly restored or have a fertile cytoplasm.

Rearing of progeny

I collected mature seed from each cross into a separate envelope, stored dry at 4˚C. Most fruits produced > 200 seeds; in a few cases < 50 seeds/fruit were produced. I grew progeny in

two cohorts because of the space limitation. Previous greenhouse studies have shown that progeny sex ratios in *L. siphilitica* are consistent across growouts (Caruso & Case 2013). In the first cohort (February 2013), I grew progeny from 104 pollinations including all 40 reciprocal crosses, 41 (of 60 total) nested crosses, and 23 self-pollinations. I cold stratified seeds on wet filter paper at 4˚C for four weeks (60 seeds/cross or pollination, when available) to break dormancy. I transferred the seeds to 10 x 10 cm pots filled with Fafard Super Fine Germination Mix (Conrad Fafard Inc., Agawam, MA, USA) for germination and bottom-watered. I transplanted seedlings to cone-tainers (20.5 cm long x 4 cm top circumference; Stuewe & Sons Inc., Tangent, OR, USA) filled with Sunshine growing mix #1 (Sun Grow Horticulture Canada Ltd., Seba Beach, AB Canada) once they had 4–5 true leaves. I grew progeny in greenhouse bays with 16 hours of supplemental light, daily watering to saturation, fertilizer every other week (N-P-K, 10-30-20) and pesticides as needed. I divided the first cohort between two greenhouses: the Holden Arboretum (Kirtland, OH, USA) and the Herrick Conservatory, Kent State University (Kent, OH, USA). Plants were randomly assigned among six blocks per location, and randomly arranged within each block.

In the second cohort (February 2015), I grew progeny of 33 crosses (29 nested and four reciprocal), which included 19 nested crosses not grown previously and 14 crosses (10 nested and four reciprocal) re-planted to increase progeny sample size. I followed the same protocol as above except that all the progeny in second cohort were grown in the Herrick Conservatory.

Validating maternal inheritance of mitochondrial genomes

My predictions about sex-determining genes from crosses were based on the assumption of maternal inheritance of mitochondrial genomes (and thus CMS types), which is sometimes questioned in gynodioecious species (reviewed in McCauley 2013; Chapter II). In order to

validate maternal inheritance in the families used in this study, I sequenced all crossing parents at two variable mitochondrial marker loci—the first intron of *nad7* (*nad7ab*) and the *atp6* gene and tested if individuals belonging to the same maternal families shared the two-locus mitochondrial haplotypes (hereafter 'mitotypes') jointly defined by these markers. With maternal inheritance, I expected individuals within the same maternal families to share mitotypes. In addition, I sequenced 10 progeny each from five nested crosses in which dams and sires had different mitotypes for *nad7ab* and compared the mitotypes of parents and progeny. With maternal inheritance, I expected all progeny to be homoplasmic for the dam's mitotype.

I carried out all PCR reactions under standard reaction conditions: 25 μL in volume, containing 2.5 μL of template DNA and with a primer-specific annealing temperature (see Chapter II). Purification and Sanger sequencing of amplicons was performed by Macrogen USA (Rockville, MD, USA) using ABI3730XL (Applied Biosystems, Foster City, CA, USA). I manually aligned sequences in Sequencher 5.2.3 (Gene Codes, Ann Arbor, MI, USA) and edited as needed.

Scoring and analysis of progeny sex ratios

I determined plant sex by examining anther morphology and the presence/absence of pollen on 5–10 flowers per individual. Plants that produced only pistillate flowers were considered female, those producing only perfect flowers were considered hermaphrodites; the few plants that produced both pistillate and perfect flowers were considered gynomonoecious. Plants rarely produced flowers that were not clearly pistillate or perfect, but these were noted.

Because progeny sex ratios did not differ significantly between grow-outs (Fisher's exact test, $P > 0.05$), I pooled progeny of those crosses grown in two cohorts for analysis. For nested crosses, I used two types of heterogeneity tests. First, I used Chi-squared contingency tests with

simulations in R v. 3.2 (R Core Team 2015) to test heterogeneity in progeny sex ratios among all the female dams pollinated by the same sires. Significant heterogeneity in progeny sex ratios suggests that there are multiple CMS types among female dams used in the crossing. Likewise, I tested the heterogeneity among the sires that were used to pollinate common female dams. Significant heterogeneity here suggests that the sires had different restorer genotypes. Second, I used Fisher's exact tests with simulations in R to distinguish whether any two female dams pollinated by a common sire had significantly different progeny sex ratios and thus different CMS types. For reciprocal crosses, I used Fisher's exact tests with simulations in order to distinguish whether pairs of hermaphrodite parents had different progeny sex ratios and thus different CMS types.

Association between marker-based mitotypes and CMS types

Crossing studies are labor intensive, thus limiting their scope to a relatively small number of populations. This makes finding an association between female frequency and CMS diversity using crossing study difficult. Identifying genetic markers that could represent CMS genes could be very useful to understand mechanisms of CMS polymorphisms using markers. Therefore, I tested if unique mitotypes jointly defined by *nad7ab* and *atp6* genes correspond to unique CMS types identified by crosses. These variable markers showed a high level of intraspecific variation in *L. siphilitica*, each yielding 11 mitotypes, and that diversity at these markers was positively correlated with the percent of females in their population of origin, potentially indicating an association with CMS genes (Chapter II). I used the same sequencing protocol as above.

Geographic distribution of CMS types

I inferred the distribution of CMS types among populations based on the distribution of matching restorers. Because specific restorer(s) are expected to evolve in response to their

respective CMS types (CMS–restorer co-evolution), I assumed that a population that has proper restorer for a CMS type also has that CMS type. Although pollen can travel far geographically, a restorer is not expected to persist in a population unless it also has a matching CMS type because such a 'silent' restorer would incur fitness cost (Bailey 2002; Dufay *et al.* 2008; Del Castillo & Trujillo 2009; Case & Caruso 2010). Nested crosses were particularly useful for these tests. By contrast, reciprocal crosses were not helpful for inferring the distribution of restorers and CMS types because a hermaphrodite dam could produce hermaphrodite progeny for multiple reasons: 1) the sire has proper restorer(s) for the dam's CMS type, 2) the dam is fixed for its own restorers, and 3) the dam has a fertile cytotype. In the latter two cases, whether the sire and dam have matching CMS type and the nuclear restorer cannot be discerned.

Genetic models of male-fertility restoration

I used G-tests of heterogeneity (Sokal & Rohlf 1995) to compare the observed progeny sex ratios with the expected ratios under various models of fertility restoration. I obtained the expected ratios for crosses involving only hermaphrodites (reciprocal crosses and self pollinations) from Garraud *et al.* (2011) and computed expected ratios for nested crosses (Table 3.S1). Female dams were predicted to produce higher proportions of female progeny than hermaphrodite dams. For example, with a single dominant restorer, a hermaphrodite can have *RR* or *Rr* restorer genotypes while a female has *rr*. Thus, expected progeny sex ratio for a nested cross should be 0:1 or 1:1 (female: hermaphrodite) *versus* 1:1 or 1:3 for a reciprocal cross or a self-pollination. Some progeny sex ratios were consistent with alternate models of restoration. In such cases, I chose models that required the fewest restorer loci, dominant (over recessive) and independent action between loci (over epistatic interaction) as being the most parsimonious explanation of the data. In cases where progeny sex ratios of a single dam fit different genetic

models among multiple crosses, I chose the model that could account for all observed progeny sex ratios produced by that dam as the most parsimonious.

RESULTS

Sex segregation patterns among crosses

In both cohorts, flowering started in late August and continued until January of the following year. A total of 3447 progeny survived to flowering, while a small proportion remained vegetative and flowered in the following year. Progeny that showed purplish but empty anthers (no pollen) were considered females. Fourteen progeny (0.4%) produced both perfect and pistillate flowers within the single plant (gynomonoecious) and 59 (1.7%) had odd anther phenotypes, such as, brownish anthers, that were not clearly sterile or fertile. I excluded these progeny from the analyses.

Most crosses (107 of 123) yielded at least 15 flowering progeny. Most nested crosses produced female progeny (80% of the progeny, N=1971); only about half (27 of 60) nested crosses produced any hermaphrodites, among which five crosses produced ≤ 3 hermaphrodite progeny. Although a small number of hermaphrodite progeny is not statistically different from zero (Fisher's exact test, $P > 0.05$), it nevertheless indicates some level of restoration. The majority of progeny in reciprocal crosses (96.7% plants, N= 1099 plants) and self pollinations (95.2% plants, N= 377) were hermaphrodite. Thus, 6 of 40 reciprocal crosses and 7 of 23 self pollinations produced enough female progeny to assess the genetics of restoration.

Table 3.2 Progeny sex segregation of nested one-way crosses between common hermaphrodite sires and multiple female dams. Heterogeneity in progeny sex ratios of dam families pollinated by each sire as well as sex ratios of a common dam family pollinated by multiple sires were tested by Chi-squared contingency tests with Monte Carlo simulations (significant P-values are underlined). Each unique pattern of segregation among dam families as identified by Fisher's exact tests ($\alpha < 0.05$) is indicated by lowercase letters (across populations) or symbols (within populations).

		Pollen parents (sires)																				Heterogeneity among sires		Sex ratio patterns	Mitochondrial haplotypes	
		BV6H			BV8H			PS6H			PS8H			YW3H			RE3H			CD3H							
		Progeny sexes																				χ2	P-value		*nad7ab*	*atp6*
		F	H	Prop. F	F	H	Prop. F	F	H	Prop. F	F	H	Prop. F	F	H	Prop. F	F	H	Prop. F	F	H	Prop. F					
Seed parents (dams)	BV2F	–	–	–	–	–	–	35	36	0.493	–	–	–	–	–	–	13	17	0.433	18	24	0.428	0.5617	0.7711	a	A	1
	BV8F	–	–	–	–	–	–	41	33	0.554	–	–	–	–	–	–	14	19	0.424	24	14	0.632	3.1135	0.2249	a	A	1
	BV6F	–	–	–	–	–	–	0	45	0.000	–	–	–	–	–	–	13	14	0.481	0	10	0.000	31.471	0.0005	b	A	3
	PS8F	–	–	–	–	–	–	17	32	0.347	15	8	0.652	15	14	0.517	12	10	0.545	19	37	0.339	9.933	0.0425	d	J	1
	PS1F	–	–	–	–	–	–	7	6	0.538	16	19	0.457	–	–	–	–	–	–	–	–	–	0.251	0.756	a	G	3
	PS6F	–	–	–	–	–	–	39	14	0.736	20	1	0.952	30	0	1.000	26	0	1.000	59	1	0.983	31.316	0.0005	c	G	1
	PS4F	–	–	–	–	–	–	–	–	–	12	0	1.000	–	–	–	–	–	–	–	–	–	–	–	f	–	9
	YW2F	–	–	–	–	–	–	50	15	0.769	–	–	–	31	0	1.000	28	0	1.000	30	0	1.000	22.755	0.0005	c	G	1
	YW9F	–	–	–	–	–	–	58	1	0.983	–	–	–	30	0	1.000	26	1	0.963	22	0	1.000	1.75	0.828	f	A	–
	YW4F	–	–	–	–	–	–	34	0	1.000	16	0	1.000	–	–	–	–	–	–	–	–	–	0	1	f	–	3
	YW8F	–	–	–	–	–	–	23	0	1.000	24	0	1.000	–	–	–	–	–	–	–	–	–	0	1	f	–	–
	HR6F	–	–	–	28	8	0.778	30	0	1.000	–	–	–	30	0	1.000	–	–	–	30	0	1.000	21.356	0.0005	e	C	3
	HR5F	–	–	–	35	1	0.972	–	–	–	–	–	–	30	0	1.000	–	–	–	–	–	–	0.846	1	f	C	10
	CR2F	–	–	–	27	3	0.900	37	0	1.000	–	–	–	21	0	1.000	29	0	1.000	30	0	1.000	11.944	0.027	f	C	3
	CR5F	–	–	–	32	1	0.969	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	f	C	3
	CR3F	–	–	–	–	–	–	30	0	1.000	–	–	–	27	0	1.000	29	0	1.000	30	0	1.000	0	1	f	C	–
	MR2F	14	1	0.933	–	–	–	29	0	1.000	–	–	–	26	0	1.000	26	0	1.000	30	0	1.000	7.459	0.107	f	C	11
	MR4F	46	0	1	–	–	–	29	0	1.000	–	–	–	34	0	1.000	30	0	1.000	30	0	1.000	0	1	f	B	3
Heterogeneity among dams	χ2			3.118			10.158			281.03			42.604			131.42			127.46			211.64					
	P-value			0.25			0.017			0.0005			0.0005			0.0005			0.0005			0.0005					

F, female; H, hermaphrodite; Prop. F, proportion of females (no. of females/total no. of progeny)

nad7ab, haplotypes based on the first intron of mitochondrial *nad7* (A–K); *atp6,* haplotypes based on mitochondrial *atp6* (1–11). A total of 11 haplotypes were identified in *L. siphilitica* for each of these markers (Chapter II). Some samples did not amplify for one or both the markers despite multiple trials.

—, data not available

Table 3.3 Progeny sex segregation of reciprocal crosses between pairs of hermaphrodites. Heterogeneities between direct and reciprocal progeny sex ratios were tested by Fisher's exact tests. P-values are only shown for crosses that produced progeny of both sexes. Hermaphrodites that segregate female progeny when selfing, are in bold. Families that do not have evidence of CMS genes are in italics.

Parent 1 (P1)			Possible CMS types	Parent 2 (P2)			Possible CMS types	Progeny sexes (P1 x P2)		(P2 x P1)		Fisher's exact test P-value
	nad7ab	*atp6*			*nad7ab*	*atp6*		F	H	F	H	
BV2H				**CD4H**	A	1	a	0	26	0	29	
BV2H				CR5aH	C	3	f	0	30	0	30	
BV2H	A	1	a	*RE6H*	D	–	–	0	21	0	30	
BV2H				*Y4.1aH*	B	–	–	0	27	0	29	
CD3bH				CR5aH	C	3	f	6	18	0	23	0.022
CD3bH				KR2aH	A	1	a	15	39	0	44	< 0.0001
CD3aH	A	3	b	**RE2H**	D	1	g	7	19	0	27	0.004
CD3bH				**Y3.2H**	–	–	–	3	23	0	28	0.105
CD4H				CR5aH	C	3	f	0	29	0	28	
CD4H	A	1	a	**RE2H**	D	1	g	4	20	0	28	0.039
CD4H				*Y4.1bH*	B	–	–	0	30	0	30	
CR5aH				KR2aH	A	1	a	0	30	0	39	
CR5aH				*RE6H*	D	–	–	0	31	0	30	
CR5aH	C	3	f	*Y4.1bH*	B	–	–	0	29	0	29	
CR5bH				KR2bH	A	1	a	0	28	0	29	
KR2aH				**RE2H**	D	1	g	1	22	0	26	0.469
KR2aH	A	1	a	*Y4.1aH*	B	–	–	0	30	0	21	
KR2bH				*Y4.3H*	B	–	–	0	28	0	11	
PS6H	G	1	c	**RE2H**	D	1	g	0	3	0	31	
Y2.3H	–	–	–	*Y4.1aH*	B	–	–	0	14	0	24	

F, female; H, hermaphrodite

nad7ab, mitotypes (A–K) based on the first intron of mitochondrial gene *nad7,* and *atp6* (1–11), mitotypes based on mitochondrial gene *atp6* based on a larger study of *L. siphilitica* (Chapter II). Possible CMS types 'a', 'b', 'c', 'f' and 'g' were inferred based on the association between CMS types and mitotypes in nested crosses.

Table 3.4 Inheritance of mitochondrial alleles in selected maternal families of *L. siphilitica* based on marker *nad7ab* (10 progeny/cross; Total N= 50).

			Pollen parents (mitotype)		
			CD3 (A)	PS6 (G)	YW3 (A)
			Progeny haplotypes		
Seed parents (mitotype)	BV2	(A)	–	A	–
	PS8	(J)	J	–	J
	YW2	(G)	G	–	G

Upper case letters A, G, and J are mitotypes (*nad7ab*). A total of 11 mitotypes (A–K) have been identified in *L. siphilitica* (II). –, progeny not sequenced

Maternal inheritance of mitochondrial genomes

Two pieces of information validated my analyses of sex-ratio data in *L. siphilitica* assuming maternal inheritance of mitochondrial genomes (and thus CMS types). First, all of the parents belonging to the same maternal families shared the same two-locus *nad7ab–atp6* mitotypes (Table 3.2 and 3.3). Second, all 10 progeny of each of the 5 nested crosses had *nad7ab* mitotypes of their maternal parents (Table 3.4). Thus, if non-maternal inheritance occurred, it was rare.

High diversity of CMS genes within species and populations

I found at least six CMS types within *L. siphilitica* and they were mainly inferred from nested crosses. Progeny sex ratios produced by 18 dam families in nested crosses yielded six unique sex-ratio patterns (Fisher's exact tests, $P < 0.05$), potentially representing at least six unique CMS types (a to f; Table 3.2). The sex-ratio type 'f' either yielded no hermaphrodites (most of the families) or did so rarely. Also, all seven sires used in nested crosses yielded unique progeny sex ratios; six varied significantly suggesting that the sires had different restorer genotypes (Chi-squared tests, Table 3.2). Reciprocal crosses yielded much less heterogeneity in progeny sex ratios, mainly because of the limited sex segregation; only three maternal families segregated progeny of both sexes (Table 3.3). Fisher's exact tests distinguished CMS types for four pairs of hermaphrodites. It is not clear whether the CMS types inferred from reciprocal crosses represent additional CMS types above the six inferred from the nested crosses because there was no overlap between maternal families yielding informative sex segregation in reciprocal crosses and those in nested crosses.

Five out of 12 populations studied were inferred to have multiple (2–4) unique CMS types based on the statistically distinguishable patterns of sex segregation in their progeny (Table

3.2, 3.3). Nested crosses revealed two CMS types each in populations BV and HR. In addition, population YW could have three CMS types because maternal families with sex-ratio types 'c' and 'f' could not be restored by pollen from a sire from within the population (family YW3), indicating that the latter family could carry a third CMS type within YW. Alternatively, this family could carry a fertile cytotype and have no restorers. Population PS could have four CMS types based on four progeny sex-ratio types (Table 3.2). Finally, reciprocal crosses revealed two CMS types in CD because one maternal family within this population produced females while the other did not when pollinated by a common sire, suggesting that the two families could have different CMS types (Table 3.3). Each of these five populations (BV, CD, HR, PS, and YW) that were found to have multiple CMS types had high female (more than 20% females) frequencies (Table 3.1).

Potential evidence of fertile cytotypes

Three maternal families (BV2, PS6, and Y2.3) used in my study never produced female progeny when crossed reciprocally with multiple hermaphrodites or when they were self pollinated, indicating that these families potentially had fertile cytotypes or they represented CMS types that are highly restored across the species (Table 3.3). This is also supported by a lack of female progeny in a number of crosses involving these families traced back to the open-pollinated seeds collected from the wild (Case AL & Caruso CM, unpublished data). The other hermaphrodite families that did not produce female progeny in reciprocal crosses did so when selfed (Table 3.3) and/or in earlier crosses, indicating that they carry CMS genes (Case AL & Caruso CM, unpublished data).

Mitotypes represent CMS types

I found six *nad7ab* (A, B, C, D, G, J) and five *atp6* mitotypes (1, 3, 10, 11, 9) in this study, which yielded 11 unique two-locus mitotypes (A1, A3, B3, C3, C10, C11, D1, G1, G3, J1, –9) upon concatenation (Table 3.2, 3.3). A few samples failed to amplify or sequence unambiguously (marked as '–') for one or both markers. Although unique mitotype for one family could be inferred despite missing data (–9), mitotypes for some other families remained undetermined. I found a strong correspondence between two-locus (*nad7ab–atp6*) mitotypes and unique CMS types inferred from my crosses (Table 3.2, 3.3). Among these, CMS types 'a', 'b', 'c', and 'd' was each associated with unique mitotypes. However, some mitotypes were associated with both 'e' and 'f', and 'f' was associated with five different mitotypes, suggesting that 'f' could represent multiple CMS types that were all poorly restored in my crosses. The strong association between unique mitotypes and CMS types was also found in reciprocal crosses (Table 3.3). Mitotype D1 was unique to a reciprocal family RE2, meaning it could represent an additional CMS type in this species.

Both widespread and narrow distribution of CMS types

Sires from geographically distant populations effectively restored CMS types in several nested crosses (Table 3.5), suggesting that these CMS types could be widely distributed across the species range. For example, dams from population BV (Ohio) were restored by sires from PS (Ohio), RE, and CD (Iowa), which were 36 km, 756 km, and 827 km away, respectively. Likewise, restorers for CMS types in PS were found as far away as 365 km (Indiana) and 784–855 km (Iowa). There was little to no restoration of CMS types in 'e' and 'f' suggesting that restorers for these CMS types were either absent or globally rare, or that their restoration genetics was complex.

Table 3.5 Geographic distances (in kilometers) between pollen and seed parent populations used in nested crosses. Open cells, successful restoration; shaded cells, no evidence of restoration based on my nested crosses. Dashes (–) represent crosses not made.

		Populations of pollen parents				
		PS (Ohio)	BV (Ohio)	YW (Indiana)	RE (Iowa)	CD (Iowa)
Populations of seed parents	PS (Ohio)	0	–	365	784	855
	BV (Ohio)	36	–	–	756	827
	MR (Illinois)	176.02	140	199.93	648.67	719.25
	HR (Ohio)	260.88	230	143.66	–	798.39
	YW (Indiana)	365	–	0	618	682.42
	CR (Illinois)	656	623	420.18	217.71	272.06

Table 3.6 Observed progeny sex ratios (female: hermaphrodites) that fitted most closely to various genetic models of restoration. Expected ratios for crosses between hermaphrodites—reciprocal crosses and self-pollinations—were taken from Garraud *et al.* (2011) and adjusted for use with female–hermaphrodites (nested) crosses. Pairs of parents found to have different CMS types are in bold and the genetic models are based on this assumption.

			No. of progeny			Heterogeneity			
Cross types	Dams	Sires	F	H	Fitted ratio	G(1)	P-value	Models of restotation	CMS
Self-pollination	BV8Ha	BV8Ha	2	5	1:3	0.049	0.824	*1 dominant	a
Self-pollination	BV8Hb	BV8Hb	8	5	1:1	0.309	0.578	2 recessive, independent	
Reciprocal	**CD3H**	**Y3.2H**	3	23	1:3	2.106	0.147	2 dominant, independent	b
Reciprocal	**CD3H**	**CR5H**	6	18	1:3	2.054	0.152	2 dominant, independent	
Reciprocal	**CD3H**	**KR2H**	15	39	1:3	0.097	0.76	2 dominant, independent	
Reciprocal	**CD3H**	**RE2H**	7	19	1:3	0	1	2 dominant, independent	
Self-pollination	CD3H	CD3H	1	28	1:15	0.061	0.805	2 dominant, independent	
Reciprocal	**CD4H**	**RE2H**	4	20	1:3	0.533	0.466	2 dominant, independent	a
Self-pollination	CD4H	CD4H	1	25	1:15	0.011	0.918	2 dominant, independent	
Reciprocal	KR2H	RE2H	1	22	1:15	0.003	0.957	2 dominant, independent	a
Self-pollination	RE2H	RE2H	2	15	1:3	1.077	0.299	1 dominant	g
Self-pollination	Y3.2H	Y3.2H	2	12	1:3	0.41	0.522	1 dominant	—
Self-pollination	Y3.3H	Y3.3H	2	11	1:3	0.244	0.621	1 dominant	—
Nested crosses	BV2F	CD3H	18	24	1:1	0.597	0.44	1 dominant	a
	BV2F	PS6H	35	36	1:1	0	1	1 dominant	
	BV2F	RE3H	13	17	1:1	0.301	0.584	1 dominant	
	BV6F	RE3H	13	14	1:1	0	1	1 dominant	b
	BV8F	CD3H	24	14	1:1	2.152	0.142	*1 dominant	a
	BV8F	PS6H	41	33	1:1	0.663	0.415	*1 dominant	
	BV8F	RE3H	14	19	1:1	0.486	0.486	*1 dominant	
	CR2F	BV8H	27	3	3:1	3.337	0.068	2 dominant, epiststic	f
	CR5F	BV8H	32	1	–	–	–	quantitative	f
	HR5F	BV8H	35	1	–	–	–	quantitative	f
	HR6F	BV8H	28	8	3:1	0.038	0.846	2 dominant, epiststic	e
	MR2F	BV6H	14	1	–	–	–	quantitative	f
	PS1F	PS6H	7	6	1:1	0	1	1 dominant	a
	PS1F	PS8H	16	19	1:1	0.114	0.735	1 dominant	
	PS6F	CD3H	59	1	–	–	–	quantitative	c
	PS6F	PS6H	39	14	3:1	2.024	0.155	*2 dominant, epiststic	
	PS6F	PS8H	20	1	–	–	–	quantitative	
	PS8F	CD3H	19	37	1:3	1.813	0.178	2 dominant, independent	d
	PS8F	PS6H	17	32	1:3	1.84	0.175	2 dominant, independent	
	PS8F	PS8H	15	8	1:1	1.583	0.208	*1 dominant	
	PS8F	RE3H	12	10	1:1	0.045	0.831	*1 dominant	
	PS8F	YW3H	15	14	1:1	0	1	*1 dominant	
	YW2F	PS6H	50	15	3:1	0.047	0.829	2 dominant, epiststic	c
	YW9F	PS6H	58	1	–	–	–	quantitative	f
	YW9F	RE3H	26	1	–	–	–	quantitative	

F, female; H, hermaphrodite. Fitted ratio is the expected ratio of females: hermaphrodites, which does not differ significantly with the observed sex ratio. When the observed ratios fitted multiple genetic models, ratios requiring fewer restorer loci, dominant restorers, and independent action were preferred. Heterogeneity was tested with a G-test of goodness of fit with Yates correction for continuity (1 degree of freedom). *= the simplistic model for a cross but not sufficient to accommodate all the crosses within the family

Simple and complex genetic models of restoration

About half of the nested crosses produced only female progeny indicating that CMS type(s) present in these families lacked proper restorers (Table 3.2). By contrast, most (85%) of the reciprocal crosses and about a quarter (30%) of self pollinations produced only hermaphrodite progeny (Table 3.3). These crosses or self pollinations were not helpful in determining genetic models of restoration. Among crosses yielding progeny of both sexes, six nested crosses fit a 1:1 ratio and three reciprocal crosses fit a 1:3 ratio, both consistent with one dominant restorer (Table 3.6). The remaining crosses/self pollinations required more complex models, with two or more restorer loci (dominant or recessive) and independent or epistatic interactions to explain offspring sex ratios. Progeny of seven crosses/self pollinations yielded complex sex ratios like 14:1, 58:1, 32:1 etc., which could only be explained by quantitative models (Ehlers *et al.* 2005). Progeny sex ratios of some of the maternal families (BV8, PS6, and PS8) varied among sires and fit multiple genetic models (Table 3.6). For instance, four of five crosses within family BV8 (nested plus reciprocal) fit with a single dominant restorer; however, the fifth cross fits only with two independent recessive restorers. Thus, I invoked this latter model because it could account for all of the crosses involving this family.

DISCUSSION

High CMS polymorphism in gynodioecious species

I found a higher level of CMS polymorphism both at species (six–seven CMS types) and population (two–four CMS types) levels of *L. siphilitica* than previously reported in this species—three CMS types within species and two within a population (Dudle *et al.* 2001; Bailey

2002). The level of CMS diversity in *L. siphilitica* is also higher than the diversity previously reported in any other gynodioecious species—up to four CMS types at the species level (*P. lanceolata*, de Haan *et al.* 1997b; *P. coronopus*, van Damme *et al.* 2004) and up to three at the population level (*P. lanceolata,* van Damme *et al.* 2004; *Beta vulgaris* ssp. *maritime*, Dufay *et al.* 2009).

High CMS polymorphism in *L. siphilitica* is consistent with the observations of high mitotype diversity both within and across populations of this species (Delph & Montgomery 2014; Chapter II). In fact, CMS diversity in this species could be even higher. A large number of maternal families yielded uninformative crosses, and those families were associated with multiple *nad7ab–atp6* mitotypes. For instance, 10 maternal families identified as sex-ratio pattern 'f' were associated with five mitotypes (Table 3.2). Based on my observation of a strong association between CMS types and mitotypes (Table 3.2, 3.3), I argue that these undetermined families could actually have additional CMS types. Additional crosses are required to confirm this prediction.

In addition, I found potential evidence of fertile cytotypes in *L. siphilitica* (families RE6, Y4.1, and Y4.3, Table 3.3) while they were previously reported in only two gynodioecious species (*Plantago lanceolata*, de Haan *et al.* 1997a; *Beta vulgaris* ssp. *maritima*, Laporte *et al.* 2001; Dufay *et al.* 2009). A second generation of crosses, including some interspecific crosses, could confirm fertile cytotypes in *L. siphilitica* but was beyond the scope of my study.

My report of higher CMS diversity in *L. siphilitica* could have two explanations. First, *L. siphilitica* could have higher CMS diversity either because higher numbers of CMS types are maintained within its populations and/or novel CMS types are formed more frequently in this species. Higher CMS diversity in this species could be associated with very high frequency of

females observed in some populations (0–100% females). Alternatively, I discovered higher CMS diversity in this species because of my extensive sampling of populations, the inclusion of several high-female populations, and focus on inter-population crosses. A comparable scale of crossing experiments with inter-population crosses in other gynodioecious species could provide an answer to this question.

Mechanisms of CMS polymorphism in L. siphilitica

Consistent with the conclusions made by the marker-based study (Chapter II), my data suggested that CMS polymorphism in *L. siphilitica* could involve two different mechanisms: 1) long-term maintenance of multiple CMS types and 2) frequent invasion of novel CMS types. First, several CMS types that were found to be widely distributed among populations hundreds of kilometers apart could have been maintained within the species for a long period of time, consistent with balancing selection. Such a widespread distribution is otherwise not expected in species like *L. siphilitica* with very low levels of cytoplasmic gene flow (mitochondrial F_{ST} = 0.65, Chapter II; plastid F_{ST} = 0.75, Madson 2012). Second, the other CMS types that were never or rarely restored by pollen from multiple populations could have invaded (or formed in) the respective populations relatively recently. Because restorers are expected to evolve in response to a CMS type, they may not have evolved yet or were recently evolved making them rare. Thus, CMS types (up to five) in ten dam families used in nested crosses (currently identified as 'f') may have restricted distributions because male sterility in these families was not (or rarely) restored (Table 3.2).

The patterns of CMS polymorphism I found in *L. siphilitica*, however, are not consistent with epidemic dynamics, which is expected to homogenize CMS types within populations—old and widespread CMS types are not expected to occur under this model. Also, my data are also

not consistent with the recent 'mixed model,' which proposes a fertile cytotype to be selectively maintained for a long period of time while multiple CMS types to invade the population periodically and competitively replacing each other (McCauley & Bailey 2009). Nevertheless, my data indicated that fertile cytotypes could be a more common component of the cytoplasmic polymorphism associated with gynodioecy (e.g., Dufay *et al.* 2007; McCauley & Bailey 2009).

Causes of highly variable female frequencies in populations

My data suggested that multiple potential mechanisms could cause highly variable frequency of females in natural populations of *L. siphilitica* (0–100% females). First, high CMS diversity could cause high female frequencies in some populations—each of the five populations (BV, CD, HR, PS, and YW) that were found to have multiple CMS types had more than 20% females (Table 3.1). Additionally, given the strong association between two-locus mitotypes and CMS types inferred from crosses (Table 3.2, 3.3), the positive correlation between mitotype diversity and frequencies of females in populations could also suggest that high CMS diversity could cause high female frequencies (Chapter II).

Second, frequent invasion of novel CMS types could cause high female frequencies in some populations because nuclear restorers are rare or absent (sex ratio types 'e' and 'f,' Table 3.2). Indeed, populations containing these CMS types had more than 40% females. By contrast, among all the reciprocal crosses involving four maternal families (KR2a–b, RE2, and RE6) from two low-female populations (KR and RE, <10% females), only a single female progeny was produced, suggesting that either these populations have few CMS types or that nuclear restores for their CMS types are easily available (Table 3.3). One of these families (RE6) could actually represent a fertile cytotype (see above).

Finally, complex restoration genetics could also cause high female frequency in some

populations. Not surprisingly, the restoration of male fertility was more complex in *L. siphilitica* than assumed by most of the theoretical models (e.g., Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991; Couvet *et al.* 1998; Dufay *et al.* 2007). Based on the informative crosses, six out of 19 maternal families yielded sex-ratio patterns consistent with a single dominant restorer locus (Table 3.6). Restoration in the remaining families required at least two restorer loci with dominant or recessive alleles and with independent or epistatic actions, and some required quantitative models of restoration. Interestingly, multiple crosses involving the same maternal families (BV8, PS6, and PS8) were consistent with more than one model of restoration. Also, progeny sex ratios of crosses associated with the same CMS types (e.g., 'a' or 'c') were consistent with multiple genetic models.

Implications of strong mitotype–CMS type association

My study showed that finding appropriate markers that could represent CMS types within a gynodioecious species could help better understand mechanisms of this sexual condition. Detailed information about diversity and distribution of CMS types that would otherwise be impractical to achieve by crossing studies could be obtained by using genetic markers. In addition, mitotypes could help infer CMS types otherwise not identified by crosses alone. For example, my prediction that sex ratio pattern 'f' in my nested crosses could represent multiple CMS types, was based on the strong association between mitotypes and rest of the CMS types. Also, my prediction that mitotype D1 could represent CMS type 'g' was also based on similar reasoning.

Some level of associations between CMS types and mitotypes was reported in other gynodioecious species. In *Plantago lanceolata*, four CMS types were associated with nine RFLP mitotypes (de Haan *et al.* 1997a) and in *S. nutans,* CMS types inferred from crossing study

showed some level of correspondence with mitochondrial *cob*-based haplotypes (Garraud *et al.* 2011). In *B. vulgaris*, the fertile and male-sterile cytoplasms were associated with different mitochondrial genome types (Satoh *et al.* 2006). Thus, further study of genetic markers that could better represent CMS types would open up opportunities to better evaluate models of gynodioecious species.

Conclusion

In conclusion, I found a higher diversity of CMS genes both at the species and population levels of *L. siphilitica* than previously reported in other gynodioecious species, including potential evidence of fertile cytotypes in this species. My data suggested that some of the CMS types within populations of *L. siphilitica* could be maintained long-term by balancing selection while other CMS types could periodically invade the population or formed *de novo*. High female frequencies within populations of *L. siphilitica* could be caused by high CMS diversity, frequent invasion/formation of novel CMS types, or complex (non-Mendelian) genetics of male fertility restoration. Finally, I found a strong association between mitotypes jointly defined by mitochondrial *nad7ab* and *atp6* and CMS types inferred from crossing study. Discovery of appropriate genetic markers that could closely represent CMS types could help understand mechanisms of gynodioecy in flowering plant species that would otherwise be very difficult solely based on crossing studies.

ACKNOWLEDGEMENT

I thank my advisor Dr. Andrea Case for her support during the entire period of this research and for providing me comments on my writing. I thank greenhouse staff of the Herrick Conservatory, Kent Ohio and The Holden Arboretum, Kirtland, Ohio for their support in

growing plants. I also thank Aleecia Willis and Nicole Moscollic for their help in growing plants and scoring sex data. This project was supported by the start-up fund from the Department of Biological Sciences, Kent State University to Case AL and an NSF grant (DEB #0842280) to Case AL & Caruso CM, and an Art and Margaret Herrick Aquatic Ecology Research Award to Adhikari B.

REFERENCES

Bailey MF (2002) A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* 56: 2178–2186

Bailey MF, Delph LF (2007) Sex-Ratio Evolution in Nuclear–Cytoplasmic Gynodioecy When Restoration Is a Threshold Trait. *Genetics* 176(4): 2465–2476

Barnard-Kubow KB, So N, Galloway LF (2016) Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* 70(12): 2752–2766

Beaudoin Yetter R (1989) The expression of male sterility in *Lobelia siphilitica* L. (Campanulaceae): a life history approach. PhD dissertation, Miami University, USA

Belhassen E, Dommée B, Atlan A, Gouyon PH, Pomente D, Assouad MW, Couvet D (1991) Complex determination of male sterility in *Thymus vulgaris* L.: genetic and molecular analysis. *Theoretical and Applied Genetics* 82(2): 137–143

Budar F, Roux F (2011) The role of organelle genomes in plant adaptation. *Plant Signaling & Behavior* 6(5): 635–639

Burt A, Trivers R (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard

University Press, Cambridge, MA

Carlsson J, Leino M, Sohlberg J, Sundström JF, Glimelius K (2008) Mitochondrial regulation of flower development. *Mitochondrion* 8(1): 74–86

Caruso CM, Case AL (2007) Sex ratio variation in gynodioecious *Lobelia siphilitica*: Effects of population size and geographic location. *Journal of Evolutionary Biology* 20: 1396–1405

Caruso CM, Eisen K, Case AL (2016) An angiosperm-wide analysis of the correlates of gynodioecy. *International Journal of Plant Sciences* 177: 115–121

Case AL, Finseth FR, Barr CM, Fishman L (2016) Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B, Biological Sciences* 283: 20161493

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46: 27–39

Charlesworth D, Laporte V (1998) The Male-Sterility Polymorphism of *Silene vulgaris*: Analysis of Genetic Data From Two Populations and Comparison With *Thymus vulgaris*. *Genetics* 150(3): 1267–1282

Chou J-Y, Leu J-Y (2010) Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *Bioessays* 32:401–411

Correns C (1906) Die Vererbung der Geschlechtsformen bei den Gynodiöcischen Pflanzen. *Ber deutsche Bot Ges* 24: 459–474

Couvet D, Ronce O, Gliddon C (1998) Maintenance of nucleo–cytoplasmic polymorphism in a metapopulation: the case of gynodioecy. *American Naturalist* 152: 59–70

Crowl A A, Miles, N W, Visger, C J, Hansen K, Ayers T, Haberle R and Cellinese N (2016) A global perspective on Campanulaceae: Biogeographic, genomic, and floral evolution. *American Journal of Botany* 103: 233–245

Darwin C (1877) *Different forms of flowers on plants of the same species*. John Murray, London

de Haan AA, Koelewijn HP, Hundscheid MPJ, Van Damme JMM (1997b) The Dynamics of Gynodioecy in *Plantago lanceolata* L. II. Mode of Action and Frequencies of Restorer Alleles. *Genetics* 147(3): 1317–1328

de Haan AA, Luyten RMJM, Bakx-Schotman TJMT, van Damme JMM (1997a) The dynamics of gynodioecy in *Plantago lanceolata* L. I. Frequencies of male-steriles and their cytoplasmic male sterility types. *Heredity* 79: 453–462

Delannay X, Gouyon PH, Valdeyron G (1981) Mathematical study of the evolution of gynodioecy with cytoplasmic inheritance under the effect of a nuclear restorer gene. *Genetics* 99 (1): 169–181

Delph LF, Kelly JK (2014) On the importance of balancing selection in plants. *New Phytologist* 201: 45–56

Delph LF, Montgomery BR (2014) The evolutionary dynamics of gynodioecy in *Lobelia*. *International Journal of Plant Sciences* 175: 383–391

Dudle DA, Mutikainen P, Delph LF (2001) Genetics of sex determination in the gynodioecious species *Lobelia siphilitica*: Evidence from two populations. *Heredity (Edinb)* 86: 265–276

Dufay M, Cuguen J, Arnaud JF, Touzet P (2009) Sex ratio variation among gynodioecious populations of sea beet: can it be explained by negative frequency‐dependent selection? *Evolution* 63: 1483–1497

Dufaÿ M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99: 349–356

Eckardt NA (2006) Cytoplasmic Male Sterility and Fertility Restoration. *The Plant Cell* 18 (3): 515–517

Ehlers BK, Maurice S, Bataillon T (2005) Sex inheritance in gynodioecious species: a polygenic view. *Proceedings of the Royal Society B: Biological Sciences* 272(1574): 1795–1802

Emery SN, McCauley DE (2002) Consequences of inbreeding for offspring fitness and gender in *Silene vulgaris*, a gynodioecious plant. *Journal of Evolutionary Biology* 15(6): 1420–1901

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* 60(7): 1372–81

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist* 133: 345–376

Garraud C, Brachi B, Dufaÿ M, Touzet P, Shykoff J (2011) Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* 106: 757–764

Geber MA, Dawson TE, Delph LF (eds.) (1999). *Gender and sexual dimorphism in flowering plants*. Springer-Verlag, New York

Gigord L, Lavigne C, Shykoff JA, Atlan A (1998) No evidence for local adaptation between cytoplasmic male sterility and nuclear restorer genes in the gynodioecious species *Thymus vulgaris* L. *Heredity* 81: 156–163

Gobron N, Waszczak C, Simon M, Hiard S, Boivin S, Charif D, et al. (2013) A Cryptic Cytoplasmic Male Sterility Unveils a Possible Gynodioecious Past for *Arabidopsis thaliana*. *PLoS ONE* 8(4): e62450

Godin VM, Demyanova EI (2013) On the distribution of gynodioecy in the flowering plants. *Bot J* 93: 1465–1487

Gouyon PH, Vichot F, Van Damme JMM (1991) Nuclear–cytoplasmic male sterility: single point equilibria versus limit cycles. *The American Naturalist* 137: 498–514

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16 Suppl 1: S154–69

Houliston GJ, Olson MS (2006) Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics* 174: 1983–1994

Ingvarsson PK, Taylor DR (2002) Genealogical evidence for epidemics of selfish genes. *Proceedings of the National Academy of Sciences, USA* 99: 11265–11269

Johnson (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics* 26 (7): 317–325

Johnston MO (1991) Natural selection on floral traits in two species of *Lobelia* with different pollinators. *Evolution* 45: 1468–1479

Johnston MO (1992) Effects of cross and self-fertilization on progeny fitness in *Lobelia cardinalis* and *L. Siphilitica*. *Evolution* 46: 688–702

Koelewijn HP and Van Damme JM (1995) Genetics of male sterility in gynodioecious *Plantago coronopus*. I. Cytoplasmic variation. *GENETICS* 139(4): 1749–1758

Lahiani E, Dufaÿ M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P (2013) Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity (Edinb)* 111: 157–64

Laporte V, Viard F, Bena G, Valero M, Cuguen J (2001) The spatial structure of sexual and cytonuclear polymorphism in the gynodioecious *Beta vulgaris* ssp. *maritima*: I/ at a local scale. *Genetics* 157: 1699–1710

Laser KD Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male-sterile angiosperms. *Bot Rev* 38: 425–454

Leppala J, Bokma F, Savolainen O (2013) Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. *Genetics* 194: 697–708

Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, Delph LF, Van Rossum F (2017) Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* 71: 1519–1531

McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol* 200: 966–977

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* 104: 611–620

McCauley DE, Bailey MF, Sherman NA, Darnell MZ (2005) Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* (Edinb) 95(1): 50–8

R Core Team (2015) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org

Rand DM, Haney RA, Fry AJ (2004) Cytonuclear cooperation: the genomics of cooperation. *Trends Ecol Evol* 19: 645–653

Renner SS, Ricklefs R (1995) Dioecy and its correlates in the flowering plants. *American Journal of Botany* 82: 596–606

Richards AJ (1997) *Plant Breeding Systems*. Chapman and Hall: London

Sambatti JB1, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2008) Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol Lett.* 11(10): 1082–1091

Satoh M, Kubo T, Mikami T (2006) The Owen mitochondrial genome in sugar beet (*Beta vulgaris* L.): possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions. *Theor Appl Genet* 113: 477–484

Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. Freeman and Co.: New York

Stadler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences, USA* 99: 11730–11735

Touzet P, Delph LF (2009) The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics* 181: 631–644

van Damme JMM (1983) Gynodioecy in *Plantago lanceolata* L. II Inheritance of three male sterility types. *Heredity* 50: 253–273

van Damme JMM, Hundscheid MPJ, Ivanovic S, Koelewijn HP (2004) Multiple CMS—restorer gene polymorphism in gynodioecious *Plantago coronopus*. *Heredity* 93: 175–181

SUPPLEMENTARY INFORMATION

Table 3.S1 Models of restoration and expected sex segregation for nested crosses between female dams and hermaphrodite sires

			Expected sex segregation (female: hermaphrodite)				
Number of loci	relation between	dominance					
one locus		dominant	0:1 -/- x R/R	1:1 -/- x R/+			
		recessive	1:1 r/+ x r/r	1:0 +/+ x r/r			
		both dominant	0:1 -/- -/- x R1/R1 -/-	1:3 -/- -/- x R1/+ R2/+	1:1 -/- -/- x R1/+ +/+		
two loci	independent	one dominant and one recessive	0:1 -/- -/- x R1/R1 -/-	1:3 +/+ r2/+ x R1/+ r2/r2 or -/- -/- x R1/+ R2/+	1:1 +/+ +/+ x R1/+ -/- +/+ +/+ x R1/+ r2/r2	3:5 +/+ r2/+ x R1/+ r2/+ or r1/+ r2/+ x R1/+ r2/r2	
		both recessive	1:3 r1/+ r2/+ x r1r1 r2r2	1:1 r1/+ -/- x r1/r1 -/-	1:0 +/+ +/+ x r1/r1 -/-		
two loci	with epistasy	both dominant	0:1 R1/R1 R2/R2 x -/- -/-	1:3 R1/R1 R2/+ x +/+ R2/+	1:1 R1/R1 R2/+ x +/+ +/+ or R1/R1 R2/+ x R1/+ +/+	5:3 R1/+ R2/+ x R1/+ +/+	3:1 R1/+ R2/+ x +/+ R2/+ or R1/+ R2/+ x +/+ +/+
		one dominant and one recessive	1:1 R1/R1 r2/r2 x R1/R1 r2/+ or R1/R1 r2/r2 x R1/+ r2/+ or R1/+ r2/r2 x R1/R1 r2/+ or R1/+ r2/r2 x +/+ r2/r2	5:3 R1/+ r2/r2 x R1/+ r2/+	1:0 R1/R1 r2/r2 x R1/+ +/+ or R1/R1 r2/r2 x +/+ +/+ or R1/+ r2/r2 x +/+ +/+ or R1/+ r2/r2 x r1/r1 +/+ or R1/+ r2/r2 x r1/+ +/+		

CHAPTER IV. SEARCHING FOR CYTOPLASMIC MALE STERILITY GENES IN A GYNODIOECIOUS WILDFLOWER, *LOBELIA SIPHILITICA* L. (CAMPANULACEAE), USING NORTHERN HYBRIDIZATION ASSAYS

ABSTRACT

Cytoplasmic male sterility (CMS) genes are chimeric mitochondrial genes that disrupt viable pollen production. CMS genes are widespread among flowering plants, but are rarely expressed in nature because specific nuclear loci (restorers) suppress the action of CMS genes in most species. CMS genes are interesting biological entities, and are considered selfish genetic elements and drivers of cytonuclear conflict/co-evolution. In crop plants, they are useful tools for commercial breeding, and have been studied in detail. However, CMS genes remain largely uncharacterized in wild plants. In crop plants, CMS genes are associated and co-transcribed with some of the essential mitochondrial genes. Here, I used the basic features of CMS genes in crop plants to search for mitochondrial transcripts associated with male sterility among multiple families of a wildflower, *Lobelia siphilitica,* in which CMS genes are naturally expressed. Because CMS genes are co-transcribed with essential genes in females but are modified by nuclear restorers in hermaphrodites, I predicted that mitochondrial transcripts associated with CMS genes would differ between female and hermaphrodite siblings. I created RNA blots containing female and hermaphrodite full siblings and hybridized them with 12 essential

mitochondrial gene probes. I found sex-specific transcript heteromorphism associated with two different mitochondrial essential genes *atp4* and *atp6*, respectively, in three different families within a single population. They potentially represent two distinct CMS types. The rest of the families did not show sex-specific transcript heteromorphism. The two putative CMS types showed different patterns of sex-specific transcription, suggesting that the respective nuclear restores could act differently on different CMS types. Future work examining whole mitochondrial genomes has the potential to identify CMS genes.

Keywords: CMS, *Lobelia siphilitica*, Northern, transcript, *atp6*, *atp4*

INTRODUCTION

Cytoplasmic male sterility (CMS) genes are a group of complex mitochondrial genes that disrupt pollen production pathways, resulting in male sterility or functionally female phenotypes (Hanson & Bentolila 2004). CMS genes are considered to be widespread among flowering plants and have already been reported in more than 150 species (Carlsson *et al.* 2008; Gobron *et al.* 2013). However, CMS genes go unnoticed in a majority of species because specific nuclear loci, called 'restorers of fertility,' suppress the action of CMS genes (Charlesworth & Ganders 1979), resulting in normal pollen production or hermaphroditism (Geber *et al.* 1999). CMS genes are, therefore, discovered spontaneously (e.g., Liu *et al.* 1987, 2005; Wan *et al.* 2008) or by experimental crosses when they are uncoupled with the specific restorer alleles (Hinata & Konno 1979; Gobron *et al.* 2013). CMS genes are naturally expressed in a small percentage of flowering plant species (<1%; Godin & Demyanova 2013) resulting in a dimorphic sexual condition called

gynodioecy, in which females and hermaphroditic individuals co-occur within a population (Sakai & Weller 1999). Gynodioecy has been found in 21% of the total flowering plant families, and has evolved numerous times in different lineages (Caruso *et al.* 2016).

CMS genes are of interests to biologists for multiple reasons. First, they are important tools of commercial hybrid seed production—the discovery of novel CMS types is a boon to the hybridization programs (Havey 2004). Second, CMS genes provide suitable opportunities to understand the interaction and co-evolution between cytoplasmic and nuclear genomes in eukaryotes (Schnable & Wise 1998; Budar *et al.* 2003; Hanson & Bentolila 2004). Third, CMS genes are suitable systems to understand the evolution and biological consequences of selfish genetic elements, which are widespread among organisms (Burt & Trivers 2006; Werren 2011). Despite being interesting biological subjects, very little is known about CMS genes in natural plant populations, including their diversity and evolutionary dynamics. Our knowledge of CMS genes comes primarily from research done in crop plants thanks to their use in the commercial breeding programs.

Based on the crop research, CMS genes share some basic features: each CMS type is a chimeric open reading frame (ORFs) formed by mitochondrial recombination, has unique structure and sequence compared to other CMS types, contains sequences of one or more of the essential mitochondrial genes, and is co-transcribed with at least one essential mitochondrial genes, usually associated with respiratory pathways (Hanson & Bentolila 2004; Li *et al.* 2007; Chen & Liu 2014). Although this provides some basic information about CMS genes, the quest for a new CMS type is challenging because of several uncertainties involved.

First, each CMS gene is assumed to arise *de novo* by unique recombination activities in

the mitochondrial genomes that typically vary by CMS types, even within a species (Schnable & Wise 1998). Therefore, predicting the structure and sequence of an unknown CMS gene is challenging. Second, mitochondrial genomes in flowering plants are structurally labile, making relative positions of mitochondrial genes, including CMS genes, difficult to predict (Sloan *et al.* 2012; reviewed by Gualberto & Newton 2017). Third, identifying CMS genes by searching for mitochondrial chimeric ORFs is not practical because there are numerous chimeric ORFs in mitochondrial genomes of flowering plants, and only a few of them are known to have CMS function (Hanson & Bentolila 2004; Arrieta-Montiel & Mackenzie 2011). Thus, identifying CMS genes requires an understanding of their expression: CMS genes are expressed in females but not in hermaphrodites, in which specific nuclear restorers modify them to suppress their expression. The quest for CMS genes could be even more complicated if the characteristics of CMS genes common to crop plants are not applicable to CMS genes in other species. A recent study in *Silene vulgaris* showed that a non-coding mitochondrial transcript, which was not even located near any other transcribed genes, could be associated with CMS (Štorchová *et al.* 2012; Stone *et al.* 2017).

Flowering plants in which the CMS phenotype is naturally expressed provide an ideal opportunity to understand CMS genes because females indicate the presence of CMS. In addition, species with natural CMS expression are predicted to have a higher diversity of CMS genes (Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991; Couvet *et al.* 1998; Dufay *et al.* 2007), providing us opportunity to identify and compare multiple CMS types. Indeed, crossing experiments have found multiple CMS types in several species with natural CMS expression (e.g., *Beta vulgaris* ssp. *maritima*, van Damme *et al.* 2004; *Lobelia siphilitica*, Dudle *et al.* 2001, Bailey 2002, Chapter III; *Plantago coronopus*, van Damme *et al.* 2004; *Silene nutans*, Garraud *et*

al. 2011 and references therein); however, the actual CMS genes still remain to be identified and characterized at molecular level in these species.

The goal of this study was to identify mitochondrial transcripts associated with male sterility in multiple maternal families of a gynodioecious wildflower *Lobelia siphilitica* Campanulaceae (Figure 4.1) by using standard Northern hybridization technique (Sambrook *et al.* 1989). I assumed that maternal full siblings inherit the same CMS type based on predominantly maternal inheritance of mitochondrial genome in flowering plants (Birky 2001; Chapter III). However, female and hermaphroditic full siblings differ in their restorer genotypes as evidenced by different sex expression and should vary in the mitochondrial transcripts associated with CMS or sex expression (Case & Willis 2008).

Based on the CMS genes characterized in other species (reviewed by Hanson & Bentolila 2004; Case & Willis 2008; Chen & Liu 2014), I assumed that CMS genes in *L. siphilitica* would be co-transcribed with essential mitochondrial coding genes, and that female and hermaphroditic full siblings would differ in the size of transcripts containing CMS genes. Specifically, the CMS genes would be transcribed along with the essential mitochondrial genes in females, thus disrupting pollen development (Touzet & Meyer 2014). By contrast, in hermaphrodites, restorer gene products modify the expression of CMS genes by altering transcription or transcript processing, typically by cutting CMS genes off the bicistronic RNA molecules (Wise *et al.* 1996). Therefore, plants with (hermaphrodites) and without (females) appropriate restorer alleles should differ in the size or sequence of mitochondrial transcripts containing CMS genes. Typically, females should have CMS-associated transcripts not present in fertile full siblings. This allowed me to use essential mitochondrial genes as hybridization probes (following Case & Willis 2008).

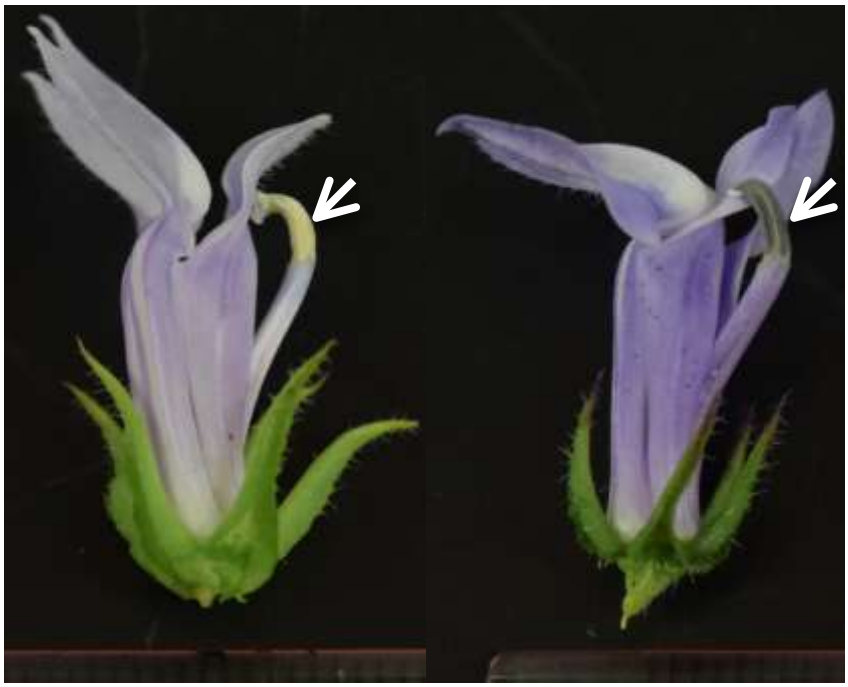


Figure 4.1 *Lobelia siphilitica*: a pistillate flower (left) with sterile (whitish) anthers (arrow) and a perfect flower (right) with fertile (dark purple) anthers (arrow).

I studied six maternal families representing three populations (1–3 families per population) of *L. siphilitica* in this study (Table 4.1). The results of my crossing study (Chapter III) allowed me to infer at least two CMS types from among these families based on the progeny sex ratios. That is, if two female dams have different CMS types, they are expected to yield significantly different ratios of female and hermaphrodite progeny, when pollinated by a single sire. However, similar progeny sex ratios do not necessarily indicate that the dams share the CMS type. Therefore, I could not infer whether the five remaining families had same or different CMS types.

The first question I asked in this study was: which essential mitochondrial genes are associated with CMS genes in *L. siphilitica*? Based on the CMS genes characterized in other species (reviewed by Hanson & Bentolila 2004; Case & Willis 2008; Chen & Liu 2014), I assumed that CMS genes in *L. siphilitica* would each be co-transcribed with an essential mitochondrial coding gene. Because it was not logistically feasible to screen every essential mitochondrial gene, I choose a subset of genes that were found to be CMS-associated in crop plants. Two of those genes (*nad7* and *atp6*) are particularly likely to be associated with CMS genes in *L. siphilitica* because the CMS types identified by my crossing experiment showed strong correspondence with the mitotypes jointly defined by the first intron of *nad7* (*nad7ab*) and the coding region of *atp6* (Chapter III). Also, mitotype diversity defined by these markers was positively correlated with frequency of females in natural populations, possibly indicating an association of *nad*7 and *atp*6 with sex expression (Chapter II).

The second question I asked was whether different mitochondrial genes in *L. siphilitica* were associated with different CMS types. Making clear prediction for this question was difficult because different mitochondrial essential genes are, sometimes, associated with different CMS

Table 4.1 Populations and families used in the current study. Standard U. S. state abbreviations are used. Population and family names are based on the crossing study (Chapter III)

U. S. states	Population codes	Latitude	Longitude	Families used in this study
OH	BV	41.449	-83.788	BV2a BV2b BV8
OH	PS	41.641	-83.434	PS1 PS6
IL	CR	40.314	-91.043	CR4

types but, at the other times, a single essential gene was associated with multiple CMS types, in crop plants (reviewed by Hanson & Bentolila 2004, Chen & Liu 2014). For example, a single essential gene *atp6* was associated with four different CMS types in rice (Chen & Liu 2014).

The final question I asked was whether the mitochondrial transcripts associated with different CMS types would also show different patterns of sex-specific transcript heteromorphism in *L. siphilitica*. Based on the studies in crop plants, restorers can interact with CMS genes in different ways and at various levels of expression—genomic, mRNA, protein or metabolic (Hanson & Bentolila 2004; Chen & Liu 2014). Also, based on the crop plants, different CMS types have different length and constituent fragments of sequences. Thus, even the CMS types sharing the restoration mechanism may have different patterns of sex-specific transcripts, such as the number, size, and position of the bands.

MATERIALS AND METHODS

Collection of tissue samples

I obtained the tissue samples for this study from a greenhouse growout by CM Caruso at University of Guelph, Ontario Canada, and from my greenhouse crossing experiment at Kent State University, Ohio (Chapter III; Table 4.1). The maternal families in Ontario and Ohio shared maternal ancestry 4–5 generations back and were expected to share CMS type based on the assumption of maternal inheritance mitochondrial genomes. Although rare non-maternal inheritance is always possible, maternal inheritance was observed in *L. siphilitica* when mitochondrial haplotypes of parents and siblings of five crosses were compared. That is, all the

siblings had mitotypes of their moms (Chapter III).

In *L. siphilitica*, blue-to-purple flowers (ca. 3-cm-long) are organized in a long apical spike. The anther and filaments are fused together forming a tube that covers the pistil. The anther tubes distinguish flower sex, being dark-purple with fertile pollen in perfect flowers and whitish with no pollen in pistillate flowers (Figure 4.1). In order to obtain plenty of high-quality RNA, I collected the entire inflorescence with several young floral buds after determining plant sex based on observing at least five mature flowers. To ensure RNA integrity, I flash-froze tissue in liquid nitrogen immediately after it was cut from the plant and stored it at –80˚C until RNA extraction.

Individuals of intermediate sex phenotypes have sometimes been found in *L. siphilitica*. Among the intermediates include gynomonoecious plants, in which both the pistillate and perfect flowers are found in the same individual, or partial steriles, in which anthers show intermediate phenotypes, such as brownish anthers, purplish but empty anthers, etc. I observed 14 (0.4 %) gynomonoecious and 59 (1.7 %) partial steriles in my growout of 3447 plants for my crossing experiment, although there was no specific association with a particular family or population (Chapter III). I excluded both types partial steriles from this study.

Northern blotting and probing

I extracted total RNA from young floral buds using RNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA), and measured the quality and quantity of RNA by using a spectrophotometer (Nanodrop Technologies, USA). Most of the RNA samples were highly concentrated (>500 ng/uL). I cleaned the less concentrated samples (<300 ng/uL) by using Qiagen RNeasy MinElute Cleanup Kit to increase concentration.

My primary goal in this study was to find mitochondrial essential genes associated with CMS genes in *L. siphilitica*. Thus, I created RNA blots using pairs of female and hermaphrodite full sibling from each of the six maternal families to screen for sex-specific transcript heteromorphism. I ran an equal amount of total RNA (4–5 μg) in each lane of a denaturing formaldehyde agarose gel along with an RNA ladder and used Ethidium Bromide to verify RNA integrity and quantity. I transferred RNA to a nylon membrane (Amersham Hybond-XL, GE Healthcare Life Sciences, Piscataway, NJ) and cross-linked RNA to the membrane by baking it at 80˚C for 2 hours.

I designed DNA probes containing one of the 12 essential mitochondrial protein-coding genes, namely, subunits of ATP synthase (*atp*), cytochrome oxidase (*cox*), and NADH dehydrogenase (*nad*; Table 4.2). I amplified probes using standard PCR protocol and purified using Qiagen QIAquick Gel Extraction Kit. I radiolabeled (α-^{32}P-dCTP) the probes by using Amersham Rediprime II Random Prime Labeling System (GE Healthcare UK Little Chalfont, Buckinghamshire, UK), and purified by using illustra NICK Columns Sephadex G-50 DNA Grade (GE Healthcare). I hybridized RNA blots with the radiolabeled probe in ULTRAhyb Ultrasensitive Hybridization Buffer (life technology) for 12 hours (42˚ C), washed in a gradient mixture of buffer made of Saline Sodium Citrate (SSC) and Sodium Dodecyl Sulfate (SDS), exposed to a Storage Phosphor Screen BAS-IP (GE Healthcare) for an average of five days (based on the radioactivity), and imaged in Typhoon FLA 9000 Fluorescent Image Analyzer (GE Healthcare).

I stripped blots up to three times, when needed, using membrane manufacturer (GE Healthcare)'s protocol, for re-probing. I created multiple blots in order to screen all 12 gene probes. Once I identified maternal families with sex-specific transcript heteromorphism, I further

tested these families by creating blots with more replicates (N= 6–10 per family) to verify the results.

Table 4.2 PCR primers sequences (forward, F, reverse, R) of 12 mitochondrial sequences used to probe the transcriptome of *Lobelia siphilitica*. All primers except *nad9* were designed for this study based on the draft sequences of *L. siphilitica*; the *nad9* primers were taken from Duminil *et al.* 2002 and the reaction conditions were readjusted for *L. siphilitica.*

Mitocohndrial protein	Gene/subunit		Primer sequences	Probe size (bp)
ATP synthase	*atp1*	F	GAAATGAACTCAAGGAGCAC	500
		R	TCCGAACAGATTTGTCCATC	
	atp4	F	AGGTAAGACTTTCAAAGCGACTCT	392
		R	GAACAACCCCCTACCCGAAC	
	atp6	F	TGGCAATCCTTGGTAGAGCT	507
		R	CGGACCGGTTAATGCAAGAAC	
	atp8	F	TTCTGGTCATGCCTTTTCCTCT	439
		R	AAAACCGATGCTTCCTTGGC	
Cytochrome oxidase b	*cob*	F	CCTCTTCCAACTCGTCCCAG	496
		R	TGGGTGTACATTCTGAGATGGA	
Cytochrome oxidase c	*cox1*	F	TCCAGTGACAGCATTCCCAC	524
		R	CTAGCCCAGAATTTGCCGGA	
	cox3	F	GCATTTCAAGGGGGTGCAAC	466
		R	CTGAAATAGTGGAGGGCGCT	
NADH dehydrogenase	*nad3*	F	TTGATCCCACTCGGTCTTCC	303
		R	TTACTCCCGATCCGAAGCAC	
	nad4L	F	TTTAGGTATTTGGGGAATCC	239
		R	TAGATTCGACAGCAATAGTC	
	nad6	F	TCCCAGTCTTTCGCAACACT	514
		R	ATTGGGTCTGTCGTCCTCCT	
	nad7	F	TGCGAGGTACCATTACGAGC	381
		R	TCCCCAATCCTTTGCTTGCT	
	nad9	F	GGTCATCTCAATGGGYTCAG	652
		R	TATAGTTGGGAGACTTTACC	

RESULTS AND DISCUSSION

The preliminary screening of six maternal families with all the gene probes showed sex-specific transcript heteromorphism in five families (Figure 4.2). The heteromorphism in these families were associated with five essential mitochondrial genes, namely, *atp6*, *atp4*, *nad3*, *nad7*, and *nad4L*. In each of these families, individuals of one sex showed extra band(s) of transcripts not present in full siblings of the other sex. Interestingly, heteromorphism in three maternal families were associated with a single essential gene *atp4* while heteromorphism in a single maternal family (CR4) was associated with two essential genes (*nad4L* and *nad7*).

Further tests of these preliminary patterns with more replicates showed consistent sex-specific transcript heteromorphism only in three maternal families, all from a single population BV (Figure 4.3, 4.4). One outlier each was found in families BV8 and BV2a, which showed the transcript patterns of the opposite sex. The extra band(s) of transcript(s) present in female but not in hermaphrodite full siblings are likely associated with sterility, that is, CMS types in the respective families.

Based on my data, there could be at least three different CMS types among six maternal families used in this study. The first CMS type could be present in family BV8 and could be associated with mitochondrial gene *atp6* because females had two extra transcripts, associated with *atp6*, not present in hermaphrodites (Figure 4.3). The extra transcripts were larger (about 1.5kb and 2.0kb) than the transcripts shared by all the full siblings.

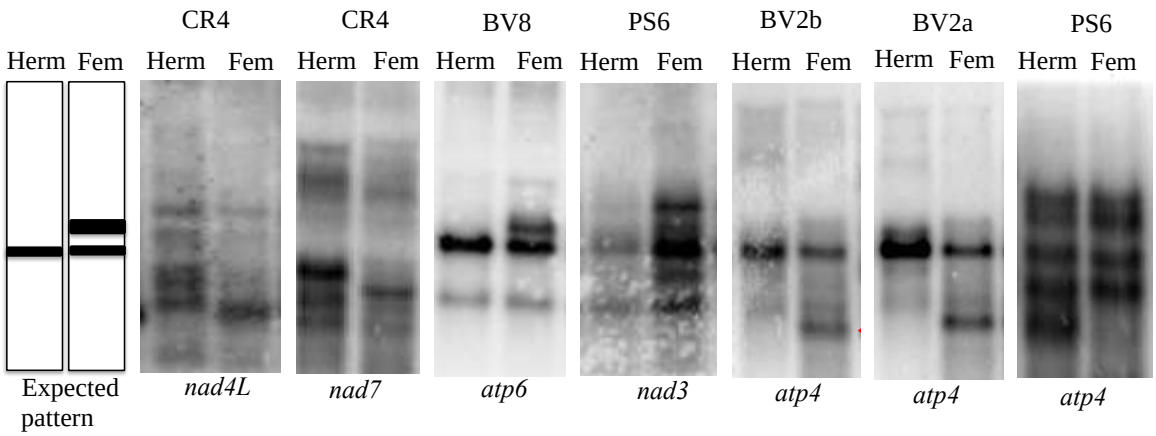


Figure 4.2 Potential cases of sex-specific mRNA associated with various mitochondrial gene probes (bottom row) found in five maternal families (top row). Expected pattern of sex-specific transcripts based on a typical method of transcription is shown. Herm, hermaphrodite; Fem, female

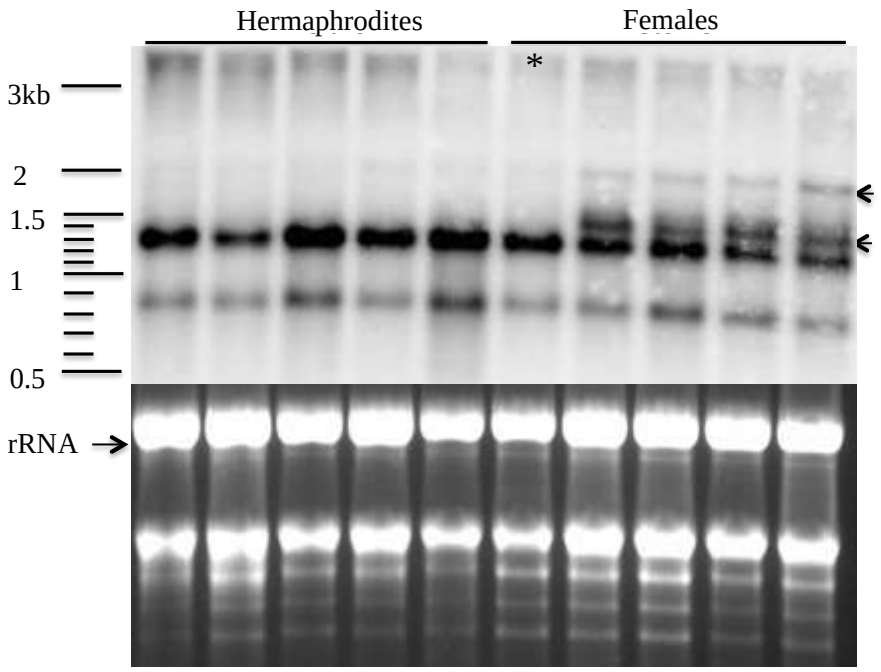


Figure 4.3 Sex-specific mRNA associated with *atp6* found in family BV8. Females have two additional transcripts (1.5 kb and 1.8 kb) not found in hermaphrodites, except for one outlier (*). The rRNA bands in the corresponding gel picture are shown.

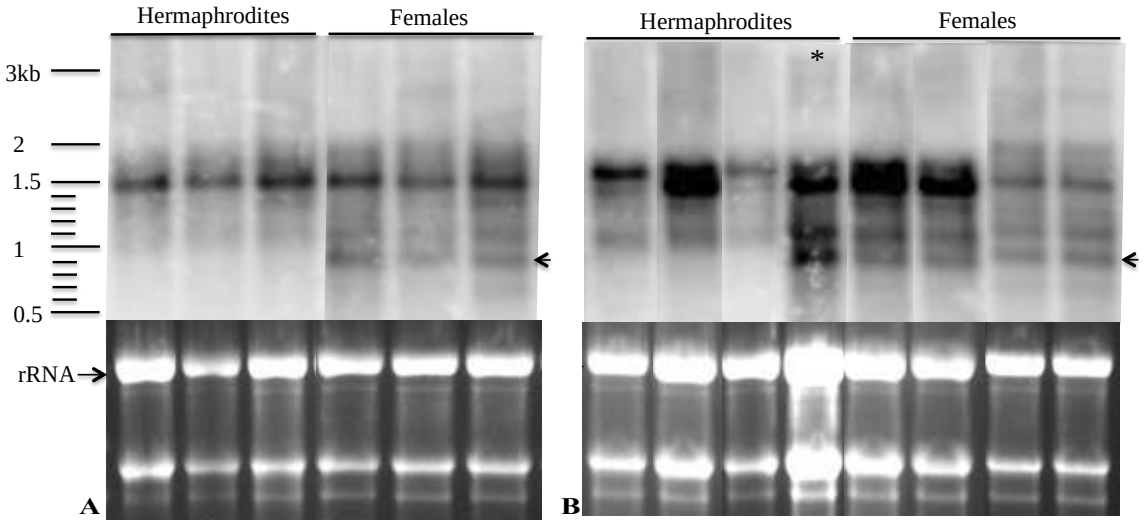


Figure 4.4 Sex-specific mRNA associated with *atp4* found in families BV2b (A) and BV2a (B). Females have one additional transcript (0.9 kb) not found in hermaphrodites, except for one outlier (*). The rRNA bands in the corresponding gel pictures are shown.

The second CMS type could be present in families BV2a and BV2b and potentially associated with essential gene *atp4* because females had one extra transcript, associated with *atp4*, not present in hermaphrodites (Figure 4.4). The extra transcript was smaller (about 0.9kb) than the bands shared by all the full siblings. Similar patterns of transcript heteromorphism found in BV2a and BV2b is consistent with their shared maternal ancestry. The CMS type(s) in the remaining families are likely different than either BV8 or BV2a/BV2b because they did not show sex-specific transcript heteromorphism with the latter families. Indeed, PS1 was inferred to have different CMS type than BV2a/BV8 by crossing study (Chapter III). My discovery of multiple putative CMS types in *L. siphilitica,* including two CMS types within a population, is consistent with the theoretical predictions that polymorphism of sex-determining genes causes natural CMS expression in flowering plants (Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991; Couvet *et al.* 1998; Dufay *et al.* 2007). In this sense, findings of this study are consistent with the discovery of multiple CMS types within populations of *L. siphilitica* (Dudle *et al.* 2001; Bailey 2002; Chapter III).

Two CMS types inferred within two BV families (BV2a and BV8), in this study, were not identified as distinct in the crossing study (Chapter III). In the crossing study, these two families showed similar progeny sex ratios when pollinated by a common sire, making it unclear whether these families have same or different CMS types (Chapter III). In crossing studies, female dams that produce significantly different progeny sex ratios when pollinated by a common sire are expected to have different CMS types. However, if they produce similar progeny sex ratios, they may or may have the same CMS type (Dudle *et al.* 2001). Thus, the two BV families, identified to have different CMS types in this study, could produce different sex-ratio patterns if additional crosses with more sires are made.

Alternatively, they have the same CMS type but incorrectly identified as distinct by this study. Not all the transcript heteromorphism in species with natural CMS expression should necessarily be CMS-associated. For example, hermaphrodites of these species may house restorer alleles that are not specific to the CMS types they carry. Such restorers could target and modify random mitochondrial transcripts causing transcript variation not associated with CMS. This could explain why I found a lot of non-sex-specific transcript heteromorphism in several maternal families in my study, such as in preliminary blots (Figure 4.2). These patterns did not hold when more replicates were added to the blots. Nevertheless, consistent patterns of sex-specific heteromorphism would not be expected among replicates if they were not associated with sex-expression (CMS).

My data also indicated that mitochondrial transcripts associated with different CMS types could be processed differently by their respective restorers. This was indicated by different patterns of sex-specific transcripts in BV8 as compared to BV2a/BV2b. The pattern of transcript in BV8 was as typically expected if the CMS transcript remains intact in females but is cut off the essential mitochondrial gene transcripts in hermaphrodites, by the nuclear restorer, resulting in an extra band of transcript larger than the transcripts shared by all the full siblings (likely essential mitochondrial gene transcript) in females but not in hermaphrodites (Figure 4.2, 4.3). However, the extra band of transcript in females was smaller (0.9kb) than the transcripts shared by all the full siblings in BV2a/BV2b suggesting a potential variation in transcript processing (Figure 4.4). Although I did not find sex-specific transcripts heteromorphism in three families used in this study (CR4, PS1, and PS6), I suspect that the respective restorers modify CMS-associated transcript differently in these families.

Explaining one outlier sample each in BV2a and BV8 blots is difficult (Figs. 4.3, 4.4).

One possible explanation could be that the outliers were only partially male sterile, and were incorrectly identified as distinct sexes. For example, if an individual was gynomonoecious—an individual that produces both the perfect and pistillate flowers in different parts of the plant—it could be misidentified when the sex is scored based on a first few flowers because they all could have a single floral sex. Because I needed to collect inflorescences for the RNA extraction, I could not wait until all the flowers were open to determine plant sex. Thus, I was not sure whether the outlier samples would produce flowers of the opposite sex at a later point in time. Interestingly, gynomonoecious plants tend to be more common in natural populations of BV (Case AL, personal communication), such a trend was not found in greenhouse-grown plants in my study. If the outliers were gynomonoecious or partial steriles, they could potentially have two types of mitochondrial genomes—one with the restored CMS type and the other without (heteroplasmy; e.g., McCauley *et al.* 2013). In this case, the RNA blot would show the transcript pattern of the mitochondrial type that is more dominant in the tissue sample.

What mitochondrial essential gene(s) are associated with the CMS types of maternal families CR4, PS1, and PS6? Note that I only used 12 essential gene probes to screen sex-specific transcript heteromorphism in this study, which is just a small subsample of more than 42 protein-coding genes that could occur in flowering plant mitochondrial genome (see Sloan *et al.* 2010; Mower *et al.* 2012; Chen *et al.* 2017). Screening with additional mitochondrial gene probes could help identify sex-specific transcripts in those families. Alternatively, the restorers for the CMS gene(s) in those families could function differently, such as by RNA editing, or restoration could occur at translational or metabolic level (Chen & Liu 2014). Detecting sex-specific transcript heteromorphism using Northern hybridizations is impossible in such situations.

In conclusion, I compared transcript profiles of female and hermaphroditic full siblings of six maternal families of *L. siphilitica,* associated with 12 mitochondrial essential gene probes, to identify three potential CMS types in this species. One of these inferred CMS types could be associated with mitochondrial gene *atp4*, and the other with *atp6*. The third (or more) CMS type could be present in maternal families that did not show transcript patterns associated with either *atp4* or *atp6*. Different inferred CMS types showed different sex-specific transcript patterns suggesting that restorers could act on the respective CMS types differently. My finding of multiple inferred CMS types within populations of this species is consistent with the theoretical prediction that CMS polymorphism causes gynodioecy in flowering plants. Further studies to characterize CMS types identified in this study could involve examining the whole mitochondrial genome for chimeric ORFs, especially near mitochondrial essential genes *atp4* and *atp6* in families BV8 and BV2a/BV2b. Further studies could also involve understanding outliers, for example, by closely examining gynomonoecious and partial sterile plants. Further work could also involve screening mRNA with additional mitochondrial protein coding gene-probes to identify sex-specific transcripts in families for which they could not be identified in this study.

ACKNOWLEDGEMENT

I thank my advisor Dr. Andrea Case for her support during the entire period of this research, for sharing her experience with Northern blotting, and for her help with writing. I thank staff of the Herrick Conservatory, Kent Ohio for their support in growing plants; Caruso CM for providing tissue samples; and Vijayaraghavan S for sharing lab space and expertise. This study was supported by an Art and Margaret Herrick Aquatic Ecology Research Award (Kent State

University) and a Graduate Student Senate Research Award (Kent State University) to Adhikari B, and a University Research Council grant to Case AL.

REFERENCES

Arrieta-Montiel M P, Mackenzie S A (2011) Plant Mitochondrial Genomes and Recombination Pg. 65–84. *In*: Kempken F (eds.) *Plant Mitochondria, Advances in Plant Biology*, Springer, New York

Bailey MF (2002) A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* 56: 2178–2186

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148

Budar F, Roux F (2011) The role of organelle genomes in plant adaptation. *Plant Signaling & Behavior* 6(5): 635–639

Burt A, Trivers R (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, MA

Carlsson J, Leino M, Sohlberg J, Sundström JF, Glimelius K (2008) Mitochondrial regulation of flower development. *Mitochondrion* 8(1): 74–86

Caruso CM, Eisen K, Case AL (2016) An angiosperm-wide analysis of the correlates of gynodioecy. *International Journal of Plant Sciences* 177: 115–121

Case A L and J H Willis (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* 62(5): 1026–1039

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46: 27–39

Charlesworth D, Ganders FR (1979) The population genetics of gynodioecy with cytoplasmic-genic male-sterility. *Heredity* 43: 213–218

Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. *Annual Review of Plant Biology* 65: 579–606

Couvet D, Ronce O, Gliddon C (1998) Maintenance of nucleo-cytoplasmic polymorphism in a metapopulation: the case of gynodioecy. *American Naturalist* 152: 59–70

Dudle DA, Mutikainen P, Delph LF (2001) Genetics of sex determination in the gynodioecious species *Lobelia siphilitica*: Evidence from two populations. *Heredity (Edinb)* 86: 265–276

Dufaÿ M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99: 349–356

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist* 133: 345–376

Garraud C, Brachi B, Dufaÿ M, Touzet P, Shykoff J (2011) Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* 106: 757–764

Geber MA, Dawson TE, Delph LF (eds.) (1999). *Gender and sexual dimorphism in flowering plants*. Springer-Verlag, New York

Gobron N, Waszczak C, Simon M, Hiard S, Boivin S, Charif D, *et al.* (2013) A Cryptic Cytoplasmic Male Sterility Unveils a Possible Gynodioecious Past for *Arabidopsis thaliana*. *PLoS ONE* 8(4): e62450

Godin VM, Demyanova EI (2013) On the distribution of gynodioecy in the flowering plants. *Botanical Journal (Russia)* 93: 1465–1487

Gouyon PH, Vichot F, Van Damme JMM (1991) Nuclear–cytoplasmic male sterility: single point equilibria versus limit cycles. *The American Naturalist* 137: 498–514

Gualberto JM, Newton KJ (2017) Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol* 68: 225-252

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16 Suppl 1: S154–69

Havey M (2004) The use of cytoplasmic male sterility for hybrid seed production. Pg. 623–634. *In*: Daniell H and Chase C (eds.) *Molecular Biology and Biotechnology of Plant Organelles*, Springer, Netherlands

Hinata K, Konno N (1979) Studies on a male sterile system having *B. campestris* nucleus and *D. muralis* cytoplasm. Breeding and some characteristics of this strain. *Jpn J Breed* 29: 305–311

Li S, Yang D, Zhu Y (2007) Characterization and Use of Male Sterility in Hybrid Rice Breeding.

Journal of Integrative Plant Biology 49: 791–804

Liu H, Fu T, Yang X (1987) Discovery and studies on Polima CMS line. *Proc 7th Int Rapeseed Cong*, Pozman, Poland: 69–78

Liu Z, Guan C, Zhao F, Chen S (2005) Inheritance and mapping of a restorer gene for the rapeseed cytoplasmic male sterile line 681A. *Plant Breeding* 124: 5–8

McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol* 200: 966–977

Sakai AK, Weller SG (1999) Gender and sexual dimorphism in flowering plants: A review of terminology, biogeographic patterns, ecological correlates and phylogenetic approaches Pg. 1–31. *In:* Geber MS, Dawson TE, Delph LF (eds.) *Gender and Sexual Dimorphism in Flowering Plants*, Berlin: Springer

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring New York

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci* 3: 175–80

Sloan DB, Müller K, McCauley DE, Taylor DR, Storchová H (2012) Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol* 196: 1228–1239

Stone JD, Storchova H (2015) The application of RNA-Seq to the comprehensive analysis of plant mitochondrial transcriptomes. *Mol. Genet. Genomics* 290: 1–9

Storchova H, Müller K, Lau S, Olson MS (2012) Mosaic Origins of a Complex Chimeric Mitochondrial Gene in *Silene vulgaris*. *PLoS ONE* 7(2): e30401

Touzet P, Meyer E (2014) Cytoplasmic Male Sterility and mitochondrial metabolism in plants. *Mitochondrion* 19: 166–171

van Damme JMM, Hundscheid MPJ, Ivanovic S, Koelewijn HP (2004) Multiple CMS–restorer gene polymorphism in gynodioecious *Plantago coronopus*. *Heredity* 93: 175–181

Wan Z, Jing B, Tu J, Ma C, Shen J *et al.* (2008) Genetic characterization of a new cytoplasmic male sterility system (hau) in *Brassica juncea* and its transfer to *B. napus*. *Theor Appl Genet* 116(3): 355–362

Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* 108: 10863–10870

Wise RP, Dill CL, Schnable PS (1996) Mutator-Induced Mutations of the *Rf1* Nuclear Fertility Restorer of T-Cytoplasm Maize Alter the Accumulation of *T-Urf13* Mitochondrial Transcripts. *Genetics* 143(3): 1383–1394

CHAPTER V. CONCLUSION AND FUTURE DIRECTION

The primary goal of this dissertation research was to understand evolutionary mechanisms by which cytoplasmic (mitochondrial) male sterility (CMS) is naturally expressed in flowering plants, resulting in gynodioecy, a dimorphic sexual system in which male-sterile (female) and male-fertile (hermaphroditic) individuals co-occur within populations. Numerous theoretical models have been proposed to explain natural gynodioecy in flowering plants and all models predict that polymorphism of CMS genes (and their nuclear restorers) is required for the maintenance of gynodioecy (Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991; Couvet *et al.* 1998; Bailey *et al.* 2003; Jacobs & Wade 2003; Dufay *et al.* 2007; McCauley & Bailey 2009). However, these models vary in specific mechanisms by which the polymorphism is maintained.

One class of models assumes that CMS types (i.e., unique CMS genes) periodically invade a population and replace an existing CMS type in an epidemic fashion (Frank 1989; Couvet *et al.* 1998). Under this scenario, populations are expected to be monomorphic for CMS types and that the CMS types are expected to be relatively younger, and thus rare, and restricted in distribution because they do not have sufficient time for migration. The other class of models assumes that multiple CMS types are maintained long-term within a population by a special form of balancing selection called 'negative frequency dependent' selection (Charlesworth 1981; Gouyon *et al.* 1991; Bailey *et al.* 2003; Dufay *et al.* 2007). Under this scenario, CMS types are

expected to be relatively older, and likely widely distributed geographically because they have had sufficient time for migration. A recent 'mixed' model considers that a fertile (non-CMS) cytotype is maintained long-term by balancing selection while successive CMS types competitively replace one another by partial selective sweeps (McCauley & Bailey 2009). Under this scenario, the fertile cytotype is expected to be relatively older (and widespread) while CMS types are expected to be relatively younger (and restricted in distribution).

Testing predictions of theoretical models of gynodioecy is challenging because the actual CMS genes have not been identified or characterized at the molecular level in any species with natural gynodioecy. This makes understanding the diversity and distribution of CMS genes challenging. Therefore, I used alternative strategies, namely, genetic markers and crossing experiments, to evaluate theoretical models of gynodioecy in *L. siphilitica*. In addition, I attempted to identify actual CMS genes in this species by comparing transcript profiles between female and hermaphroditic full siblings, which are expected differ consistently in transcripts containing functional CMS genes.

CHAPTER II: Cytoplasmic discordance is associated with sex-ratio variation in gynodioecious *Lobelia siphilitica* L. (Campanulaceae)

In this chapter, I used cytoplasmic markers to evaluate theoretical models of gynodioecy in *L. siphilitica* and also tested the level of association in patterns of genetic diversity or linkage disequilibrium (LD) between plastid and mitochondrial genomes. The first goal of Chapter II was to compare the patterns of diversity and distribution of mitochondrial haplotypes (mitotypes) in *L. siphilitica* with the patterns expected under various theoretical models of gynodioecy and to

understand potential mechanisms of highly variable female frequencies observed in populations (Caruso & Case 2007 and unpublished data). My interpretation in this project is based on the assumption that marker-defined mitotypes correspond to CMS types within the same (mitochondrial) genome, as assumed by other marker-based studies in gynodioecious species (e.g., Stadler & Delph 2002; Houliston & Olson 2006; Touzet & Delph 2009; Lahiani *et al.* 2013; Delph & Montgomery 2014).

My data indicated that gynodioecy in *L. siphilitica* could be caused both by long-term maintenance of multiple CMS types (within populations) by balancing selection, as well as frequent formation of novel CMS types, potentially by mitochondrial recombination. The long-term maintenance of CMS types and/or fertile cytotypes was suggested by the presence of common and geographically widespread mitotypes while frequent formation of novel CMS types was suggested by the occurrence of rare and geographically highly restricted mitotypes. My data, however, do not suggest that 'epidemic dynamics' are responsible for the observed restricted mitotypes in *L. siphilitica* because the mitotype polymorphism that I observed within several populations is too high to be expected under epidemic dynamics, which should homogenize mitotypes types within populations (Frank 1989; Couvet *et al.* 1998). Concluding whether a 'mixed model' of gynodioecy (McCauley & Bailey 2009) could operate in *L. siphilitica* was not possible using these data because it is not clear if some of the observed mitotypes correspond to fertile (non-CMS) cytotypes.

My data also indicated that high female frequencies in some populations of *L. siphilitica* could be caused by high diversity of CMS types within populations, including frequent formation of novel CMS types. The association between high CMS diversity and high female frequency was indicated by a positive correlation between mitotype diversity and population sex ratio

(percent female). Likewise, the association between novel CMS types and high frequency of females was indicated by the prevalence of rare (and likely recombinant) mitotypes in high-female populations and in female plants. Novel CMS types could result in high female frequencies because compatible nuclear restorers for such CMS types may not be present in the population or they could be rare.

The second goal of Chapter II was to test the level of association or linkage disequilibrium (LD) between plastid and mitochondrial genomes in *L. siphilitica* and to evaluate potential mechanisms determining the level of cytoplasmic LD. Both plastid and mitochondrial genomes in flowering plants are typically maternally co-inherited and clonally replicated (e.g., Palmer 1987; Milligan 1992; Rebound & Zeyl 1994; Dumolin-Lapegue *et al*. 1998; Birky 2001). Thus, both cytoplasmic markers are expected to show similar patterns diversity or strong LD (Maynard Smith & Haigh 1974; Schnabel & Asmussen 1989). However, the rules of strict maternal inheritance and clonal replication could be violated by mitochondrial genomes of gynodioecious angiosperms, potentially resulting in discordant patterns of diversity or weak cytoplasmic LD (reviewed by McCauley 2013). Indeed, in *Silene vulgaris*, contrasting conclusions regarding models of gynodioecy were obtained when using plastid (epidemic dynamics, Ingvarsson & Taylor 2002) *versus* mitochondrial (balancing selection, Houliston & Olson 2006; Touzet & Delph 2009) markers.

My data suggested that the linkage between plastid and mitochondrial genomes is weakened in *L. siphilitica*, and this pattern could be caused by frequent structural changes in mitochondrial genomes via inter- or intramolecular recombination. Recombination could produce novel mitotypes independent of plastid genomes, thus weakening cytopalsmic LD. Heteroplasmy, the co-occurrence of multiple mitotypes within an individual, is required for

homologous recombination to occur and this condition could develop via paternal leakage (the occasional inheritance of mitochondrial genomes via pollen). Paternal leakage of mitochondrial genomes still remains to be studied in *L. siphilitica*. I suspect that the co-occurrence of multiple mitotypes (potentially associated with multiple CMS types), especially in high female populations, could provide suitable opportunity for heteroplasmy in the event of paternal leakage. Interestingly, I found a prevalence of 'rare' mitotypes in high-female populations of *L. siphilitica* indicating that they could have been produced relatively recently by recombination.

Chapter II future direction

Some interpretations in Chapter II require further confirmation. First, my interpretation about the models of gynodioecy in *L. siphilitica* is based on the assumption that marker-defined mitotypes correspond to CMS types. I attempted to test this assumption in Chapter III to some extent. However, not all of the mitotypes identified in Chapter II could be tested for CMS types because of the smaller scope of the crossing study. Further studies could test association between mitotypes and CMS types.

Second, my inference that rare mitotypes identified in *L. siphilitica* could be recombinant needs confirmation. The mitotypes identified here came from field-collected samples; thus, distinguishing mitotypes as 'parental' and 'recombinant' was challenging. I assumed that rare and restricted mitotypes are younger and thus recombinant. Conversely, more common and widespread mitotypes are more likely to be parental, because they should require more time to spread among populations. Also, rare mitotypes are more likely to be lost from a population by drift compared to common mitotypes. My inference of relative times of occurrence of mitotypes based on their distribution is not definitive, because older mitotypes can be rare and restricted if there is no gene flow. Long-term fine-scale study of populations containing multiple mitotypes

with variable frequencies could help track parental and recombinant mitotypes, particularly if whole mitochondrial genomes are sequenced and characterized.

Third, my inference that homologous mitochondrial recombination in *L. siphilitica* could be associated with the formation of novel CMS types requires further confirmation. Chimeric CMS genes are formed by rearrangements in the mitochondrial genomes (Hanson & Bentolila 2004). However, it is still unclear whether the recombination I reported here was truly homologous and whether such a process could produce chimeric ORFs that ultimately become CMS genes, and/or if the recombinant mitotypes are formed as a byproduct of processes that form novel CMS types. Detailed studies to understand molecular mechanisms of novel CMS type formation using approaches such as mitochondrial fusion using plants with different mitotypes could help address this question.

Fourth, my interpretation that paternal leakage could result in heteroplasmy in *L. siphilitica,* which is required for mitochondrial homologous recombination, is inferential. Although I found several heteroplasmic sequences at *atp6* and *Ψrps12* (unpublished data), mitochondrial paternal leakage still needs to be documented in this species. Evidence of paternal leakage of plastid genome has been reported in this species (Durewicz 2012) using the same plastid marker that I used in Chapter II. However, the exact same pedigrees used by Durewicz (2012) did not show evidence of mitochondrial paternal leakage using the three mitochondrial markers from Chapter II (unpublished data). Leakage in one genome but not the other is consistent with weak cytoplasmic LD. Note that I did not find evidence of paternal leakage when studying patterns of inheritance of *nad7ab* in five pedigrees used in my crossing study (Chapter III). More detailed studies of pedigrees could help test whether mitochondrial paternal leakage occurs at all or is more rare than plastid leakage in *L. siphilitica*. Using approaches such as qPCR

as well as using more mitochondrial markers and/or whole mitochondrial genomes could help detect paternal leakage.

An alternative mechanism generating heteroplasmy is substoichiometric shifting (SSS)—a sudden increase in frequency of mitochondrial subgenomic molecules (see refs in Arrieta-Montiel *et al.* 2001). Mitochondrial SSS is considered to be common in flowering plants, and has been found to be associated with the functionalization of novel CMS types in crop plants (e.g., Small *et al.* 1989; Janska *et al.* 1998; Tang *et al.* 2017). However, testing this phenomenon in non-model species like *L. siphilitica* is difficult.

Finally, mitochondrial recombination may not be the sole cause of cytoplasmic discordance in *L. siphilitica*. In theory, either plastid and/or mitochondrial heteroplasmy followed by 'vegetative sorting' (Mogensen 1996; Birky 2001) could also cause cytoplasmic discordance. However, studies like mine that use one time sampling from wild populations cannot rigorously test this process. Experimental studies over multiple generations are required to evaluate the effects of vegetative sorting.

My finding of a weak cytoplasmic LD in *L. siphilitica* has implications for how we use cytoplasmic markers. Plastid and mitochondrial markers are often used interchangeably in evolutionary studies, sometimes based on convenience. For example, plastid markers are used to assess patterns in mitochondrial genomes in angiosperms because plastid genes generally have higher rates of substitution, and are thus more informative than mitochondrial genes (Wolfe *et al.* 1987; Drouin *et al.* 2008; Richardson *et al.* 2013; Zhu *et al.* 2014). However, if there is a weak association between the two cytoplasmic genomes, as I reported in *L. siphilitica,* plastid genomes may give erroneous clues about evolutionary dynamics of mitochondria. Therefore, testing cytoplasmic LD is essential before using one cytoplasmic genome as a proxy for the other.

On the other hand, weak cytoplasmic association could increase the utility of plastid or mitochondrial markers for studying evolutionary processes in each respective genome. For example, if there is high LD between cytoplasmic genomes, then selection on one genome affects the other, potentially obscuring (or erasing) the signature of evolutionary processes. Thus, having weak LD allows us to separate out signatures of evolutionary processes (e.g., selection) between the two when evolutionary processes affect them differently.

CHAPTER III: Diversity and complexity of sex determination in a gynodioecious wildflower *Lobelia siphilitica* L. (Campanulaceae)

In Chapter III, I investigated the diversity and distribution of CMS genes within and among populations of *L. siphilitica* using crossing experiments to understand evolutionary models of gynodioecy. I used progeny sex-ratio data to infer unique CMS types present in parental plants as well as to infer genetic models of restoration. First, I examined whether CMS types in this species were widely distributed across populations, as expected if they were maintained long-term by balancing selection and migrated over time or the CMS types were restricted in distribution as expected if they were relatively new. In the latter scenario, CMS genes could be periodically replaced within populations by epidemic dynamics, or they could be frequently formed *de novo*. I also tested for the evidence of fertile cytotypes in this species. Second, I evaluated possible mechanisms of highly variable frequency of females (0 to 100% females) in this species (Caruso & Case 2007 and unpublished data). I specifically tested whether high female frequency could be caused by higher CMS diversity and/or limited capacity for restoration of some CMS types.

Consistent with II, my crossing data also suggested that some kind of 'mixed' model of

gynodioecy could operate in *L. siphilitica.* That is, some of the CMS types appeared to be widely distributed among multiple populations, consistent with balancing selection maintaining CMS types within populations for a long time, while the other CMS types appeared to be restricted in distribution, as expected if such CMS types arise frequently. I did find potential evidence of a 'fertile' cytotype in *L. siphilitica*, although it needs further confirmation.

My crossing data in Chapter III also suggested that high frequencies of females in some populations of *L. siphilitica* could be caused by a high diversity of CMS genes as well as difficulty involved in restoration of CMS types. First, I found a high diversity of CMS types both at the species (at least six CMS types) and the population (two to four CMS types) level, higher than inferences made by previous crossing studies in any species with natural gynodioecy (e.g., van Damme *et al.* 2004; Dufay *et al.* 2009; Garraud *et al.* 2011 and references therein). Interestingly, all populations of *L. siphilitica* with multiple CMS types had 20% or more females, although I did not have sufficient representation of low-female populations to test this hypothesis rigorously. Second, several CMS types in high-female populations were not readily restored, indicating that matching restorers for these CMS types could be rare or absent. Also, some of these CMS types had complex genetics of restoration, involving multiple restorers, several with epistatic interactions.

Finally, I found a strong correspondence between mitotypes jointly defined by *nad7ab* and *atp6* (Chapter II) and CMS types inferred from the crossing experiment (Chapter III), suggesting that mitochondrial markers, if selected properly, could be used to document variation in CMS genes in larger samples than is possible with crossing experiments. This also provided support for the conclusions based on mitotypes in Chapter II.

Chapter III future direction

Chapter III provided crucial information about potential models of gynodioecy in *L. siphilitica*, potential causes of highly variable female frequencies, and information on the diversity and distribution of CMS types. However, some questions still remained unanswered.

First, one of my interpretations was that higher diversity of CMS types could be favoring high female frequency in populations of *L. siphilitica*. Indeed, all populations for which multiple CMS genes were identified had >20% females. However, because I studied fewer low-female populations, I cannot confirm that such populations have lower CMS diversity. Because of the exploratory nature of this study, the sampling was biased towards high-female populations in order to maximize the chances of discovering more CMS types. Because maternal families from low-female populations did not result in many informative crosses (i.e., they did not segregate progeny of both sexes), it was not clear whether these populations indeed had fewer CMS types or if I was just unable to uncover them with the mates that I had available. Future crossing studies should include more samples from low-female populations in order to address this question better, but will still be somewhat limited by space and scope.

Second, my study raised an interesting question regarding the level of diversity of CMS types in gynodioecious species. My study helped me infer six to seven CMS types within *L. siphilitica*. However, there were multiple maternal families yielding uninformative crosses, suggesting that there could be additional CMS types among those families. The level of CMS diversity I found in *L. siphilitica* is higher than the diversity previously reported in any gynodioecious studies, leading to at least two different explanations to explain this pattern: 1) *Lobelia siphilitica* has higher CMS diversity because this species maintains higher number of CMS types within populations and/or produces novel CMS types more frequently, 2)

gynodioecious species, in general, have higher CMS diversity that could be discovered by more extensive studies. These hypotheses can be tested by more extensive studies and by using more inter-population crosses.

Third, my study indicated that fertile cytotypes could be a common component of cytoplasmic polymorphism associated with gynodioecy. Although further confirmation is needed, three maternal families of *L. siphilitica* never produced female plants in their entire history—not in their native population and not in six generations of crosses in the greenhouse. These lineages could carry fertile (non-CMS) cytotypes. Alternatively, they could have CMS types but the dominant restorer could be fixed for them or they could represent rare cases of paternal leakage where both CMS and restorer came from father so the offspring were all hermaphroditic.

If these families have fertile cytotypes, the first-generation progeny (as dams) should not segregate females when backcrossed to their dads (e.g., Fishman & Willis 2006). If confirmed, *L. siphilitica* would be the third gynodioecious species shown to have fertile cytotypes co-occurring with CMS types; others are *Beta vulgaris* ssp. *maritima* (Laporte *et al.* 2001; Dufay *et al.* 2009) and *Plantago lanceolata* (de Haan *et al.* 1997). If fertile cytotypes are common, then the maintenance of gynodioecy could follow models similar to those proposed by Dufay *et al.* (2007) and McCauley & Bailey (2009).

Finally, strong correspondence between marker-defined mitotypes and CMS types identified by crossing is a crucial piece of information that could be used to better understand mechanism of gynodioecy in flowering plants. Because crossing studies are inherently labor intensive, more extensive sampling of populations and families is very challenging. However, detailed studies are possible with the help of markers that accurately represent unique CMS

types. Some level of correspondence between mitotypes and CMS types was also found in other gynodioecious species (e.g., *Plantago lanceolata,* de Haan *et al.* 1997; *B. vulgaris*, Satoh *et al.* 2006; *Silene nutans,* Garraud *et al.* 2011), indicating that some effort on finding informative markers could make future surveys of CMS diversity much easier. Findings of associations between mitotypes and CMS types also help validate the results of marker-based studies that assume mitotypes represent CMS types (e.g., Touzet & Delph 2009; Lahiani *et al.* 2013; Delph & Montgomery 2014; Chapter II).

CHAPTER IV: Searching for cytoplasmic male sterility genes in a gynodioecious wildflower, *Lobelia siphilitica* L. (Campanulaceae), using Northern hybridization assays

Studies to understand models of gynodioecy, so far, are based on inferential methods, such as cytoplasmic markers and crossing studies. Results of those indirect studies could be validated by obtaining information about the actual CMS genes, which are yet to be identified or characterized at molecular level in any gynodioecious species. Therefore, in Chapter IV, I looked for CMS genes in *L. siphilitica* using Northern hybridizations. Each CMS type is typically produced *de novo* by mitochondrial recombination. Therefore, there are several uncertainties regarding identification of CMS types, such as their sequences and specific locations in mitochondrial genomes (Hanson & Bentolila 2004). However, because CMS genes are expressed in females but not in hermaphrodites, in which compatible nuclear restorers prevent CMS expression, mitochondrial transcripts associated with CMS genes can be studied to identify them (Case & Willis 2008).

By screening mRNA blots containing female and hermaphrodite siblings of six maternal families with probes matching 12 mitochondrial essential genes, I identified sex-specific

transcript heteromorphism associated with mitochondrial *atp6* and *atp4* in three maternal families. These blots contained one or two extra transcripts in females that were not present in hermaphrodite full siblings, which could be associated with male sterility or the expression of CMS genes. Specific patterns of transcript heteromorphism associated with *atp6* and *atp4* were different in terms of the number and size of the bands, suggesting that the respective nuclear restorers could modify CMS transcripts in different ways. The remaining families did not show sex-specific transcript heteromorphism in any of the 12 essential genes I studied and could represent different CMS type(s) than the two potential CMS types inferred above.

Chapter IV future direction

In Chapter IV, I identified mitochondrial transcripts potentially associated with CMS types in three out of six maternal families of *L. siphilitica*. However, this was only a first step on the quest of identifying actual CMS types in this species. Further studies to identify candidate CMS types require searching for chimeric sequences in the associated regions of mitochondrial genomes of this species, characterizing CMS types, and confirming the action of CMS genes.

The fact that I could only identify sex-specific transcript heteromorphism in three out of six families raises questions about CMS types in the remaining families. I used only 12 essential mitochondrial gene probes to screen sex-specific transcripts, which is a subsample of ca. 42 protein-coding genes found in the flowering plant mitochondrial genomes (see Sloan *et al.* 2010; Mower *et al.* 2012; Chen *et al.* 2017), although most mitochondrial genes have never been associated with CMS. Therefore, one possibility is that CMS types in those families could be associated with other mitochondrial genes. Screening RNA blots with additional gene probes could potentially identify sex-specific transcripts in the remaining families. Another possibility is that the strategy used in this study (transcript comparison) may not be effective for finding CMS

types in the three remaining families. The current method only works if nuclear restorers suppressed CMS expression at the mRNA level. If, however, the restorer acts by other mechanisms like RNA editing, or at the proteomic or metabolic levels (Chen & Liu 2014), this approach would be ineffective.

Finally, I was able to identify six to seven CMS types in *L. siphilitica* based on my crossing study (Chapter III). Thus, further work could involve identifying and characterizing all of these CMS types. An integrative approach using genomics, transcriptomics, and metabolomics along with standard Northern analyses could be effective in identifying and characterizing CMS types in this species.

Conclusion

In the absence of information about the actual CMS genes, my dissertation research used alternative strategies, specifically cytoplasmic markers and crossing experiments, to evaluate models of natural gynodioecy in flowering plants. In addition, I attempted to search for actual CMS genes by comparing sex-specific mitochondrial transcripts. I was able to make several crucial findings concerning natural gynodioecy in flowering plants.

First, both mitochondrial markers and crossing experiments suggested that long-term maintenance of multiple CMS types within populations by balancing selection could maintain gynodioecy. However, my data suggest that novel CMS types could also be formed frequently. My study also suggested that high diversity of CMS types within populations and/or difficulty in restoration of some of the CMS types could result in high frequency of females in some populations of gynodioecious species. CMS types could be difficult to restore if they are recently formed because compatible restorers are rare or absent, and/or if the genetics of restoration is complex, requiring multiple restorer loci and their interactions. Natural selection is less effective

in changing the frequency of epistatic alleles, even when strong, so restorers with complex genetics are less likely to spread to fixation, maintaining CMS expression in nature.

Second, my data suggested that the diversity of CMS types could be higher, both at the population and species levels, in gynodioecious species than previously documented. Although higher CMS diversity could be a specific feature of *L. siphilitica*, it is equally possible that this could be a common feature of all species with natural gynodioecy. More extensive studies using inter-population crosses could help explore additional CMS types in gynodioecious species. In addition, my data suggested that fertile cytotypes could be common among gynodioecious species. Only two theoretical models have assumed the occurrence of fertile cytotype (Dufay *et al.* 2007; McCauley & Bailey 2009) and it is confirmed, so far, in only two gynodioecious species (*Beta vulgaris* ssp. *maritime,* Dufay *et al.* 2009; *Plantago lanceolata,* de Haan *et al.* 1997a).

Third, my study showed that the patterns of genetic diversity in two cytoplasmic genomes (plastid and mitochondrial) could be discordant in flowering plants with natural gynodioecy. The primary cause for this discordance could be frequent mitochondrial recombination, potentially associated with the formation of novel CMS types. Although my data showed evidence of mitochondrial homologous recombination, I did not find clear evidence of heteroplasmy. Study of heteroplasmy was beyond the scope of this study. Because paternal leakage and heteroplasmy have been reported in other gynodioecious species, they could also occur *L. siphilitica,* creating suitable conditions for mitochondrial intermolecular recombination. Studies of paternal leakage and heteroplasmy in this species would help address this question. The other possible mechanisms of cytoplasmic discordance, such as vegetative sorting, also need to be explored.

Finally, I identified sex-specific transcript heteromorphism associated with two

mitochondrial housekeeping genes *atp4* and *atp6* in three maternal families of *L. siphilitica*, indicating that at least two CMS types in this species could be associated with those genes. Further study of mitochondrial genomes of the respective samples is required to identify and characterize actual CMS types in this species. Study of CMS expression is required to confirm CMS types. Nevertheless, my study suggested that Northern blotting could be useful tool to identify CMS genes in other gynodioecious species. The project of assembling and comparing the whole mitochondrial genomes of ten different maternal families used in this study is currently underway and is expected to provide crucial insights regarding structure of CMS genes and evolution of mitochondrial genomes in *L. siphilitica*.

REFERENCES

Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 158: 851–864

Bailey MF (2002) A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* 56: 2178–2186

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148

Caruso CM, Case AL (2007) Sex ratio variation in gynodioecious *Lobelia siphilitica*: Effects of population size and geographic location. *Journal of Evolutionary Biology* 20: 1396–1405

Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* 62(5): 1026–1039

Charlesworth D (1981) A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 46: 27–39

Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. *Annual Review of Plant Biology* 65: 579–606

Chen Z, Zhao N, Li S, Grover CE, Nie H, Wendel JF, Hua J (2017) Plant Mitochondrial Genome Evolution and Cytoplasmic Male Sterility. *Critical Reviews in Plant Sciences* 36(1): 55–69

Couvet D, Ronce O, Gliddon C (1998) Maintenance of nucleo-cytoplasmic polymorphism in a metapopulation: the case of gynodioecy. *American Naturalist* 152: 59–70

de Haan AA, Luyten RMJM, Bakx-Schotman TJMT, van Damme JMM (1997) The dynamics of gynodioecy in *Plantago lanceolata* L. I. Frequencies of male-steriles and their cytoplasmic male sterility types. *Heredity* 79: 453–462

Delph LF, Kelly JK (2014) On the importance of balancing selection in plants. *New Phytologist* 201: 45–56

Delph LF, Montgomery BR (2014) The evolutionary dynamics of gynodioecy in *Lobelia*. *International Journal of Plant Sciences* 175: 383–391

Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol* 49: 827–831

Dufay M, Cuguen J, Arnaud JF, Touzet P (2009) Sex ratio variation among gynodioecious populations of sea beet: can it be explained by negative frequency-dependent selection? *Evolution* 63: 1483–1497

Dufaÿ M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99: 349–356

Dumolin-Lapègue S, Pemonge MH, Petit RJ (1998) Association between chloroplast and mitochondrial lineages in oaks. *Molecular Biology and Evolution* 15: 1321–1331

Durewicz A (2012) Innheritance of chloroplast DNA (cpDNA) in *Lobelia siphilitica*. An Honors thesis submitted to Kent State University, Kent Ohio

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* 60(7): 1372–81

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist* 133: 345–376

Garraud C, Brachi B, Dufaÿ M, Touzet P, Shykoff J (2011) Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity* 106: 757–764

Gouyon PH, Vichot F, Van Damme JMM (1991) Nuclear-cytoplasmic male sterility: single point equilibria versus limit cycles. *The American Naturalist* 137: 498–514

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell 16 Suppl 1*: S154–69

Houliston GJ, Olson MS (2006) Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics* 174: 1983–1994

Ingvarsson PK, Taylor DR (2002) Genealogical evidence for epidemics of selfish genes. *Proceedings of the National Academy of Sciences, USA* 99: 11265–11269

Jacobs MS, Wade MJ (2003) A synthetic review of the theory of gynodioecy. *American*

Naturalist 161: 837–851

Janska H, Kwasniak M, Szczepanowska J (2013) Protein quality control in organelles–AAA/FtsH story. *Biochim Biophys Acta* 1833: 381–387

Lahiani E, Dufaÿ M, Castric V, Le Cadre S, Charlesworth D, Van Rossum F, Touzet P (2013) Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity (Edinb)* 111: 157–64

Laporte V, Viard F, Bena G, Valero M, Cuguen J (2001) The spatial structure of sexual and cytonuclear polymorphism in the gynodioecious *Beta vulgaris* ssp. *maritima*: I/ at a local scale. *Genetics* 157: 1699–1710

Laser KD Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male-sterile angiosperms. *Bot Rev* 38: 425–454

Maynared Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetic Research* 23: 23–35

McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol* 200: 966–977

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* 104: 611–620

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* 104: 611–620

Milligan BG (1992) Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. *American Journal of Botany* 79: 1325–1328

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383–404

Mower JP, Case AL, Floro ER, Willis JH (2012) Evidence against Equimolarity of Large Repeat Arrangements and a Predominant Master Circle Structure of the Mitochondrial Genome from a Monkeyflower (*Mimulus guttatus*) Lineage with Cryptic CMS. *Genome Biology and Evolution* 4(5): 670–686

Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist* 130: S6-S29

Reboud X, Zeyl C (1994) Organelle inheritance in plants. *Heredity* 72: 132–140

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD (2013) The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol* 11: 29

Satoh M, Kubo T, Mikami T (2006) The Owen mitochondrial genome in sugar beet (*Beta vulgaris* L.): possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions. *Theor Appl Genet* 113: 477–484

Schnabel A, Asmussen MA (1989) Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. *Genetics* 123: 199–215

Sloan DB, Alverson AJ, Storchová H, Palmer JD, Taylor DR (2010) Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evol Biol* 10: 274

Small I, Sufolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via

substoichiometric intermediates. *Cell* 58: 69–76

Stadler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences, USA* 99: 11730–11735

Tang H, Zheng X, Li C, Xie X, Chen Y, Chen L, Zhao X, Zheng H, Zhou J, Ye S, Guo J, Liu YG (2017) Multi-step formation, evolution, and functionalization of new cytoplasmic male sterility genes in the plant mitochondrial genomes. *Cell Res* 27: 130–146

Touzet P, Delph LF (2009) The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics* 181: 631–644

van Damme JMM, Hundscheid MPJ, Ivanovic S, Koelewijn HP (2004) Multiple CMS—restorer gene polymorphism in gynodioecious *Plantago coronopus*. *Heredity* 93: 175–181

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84: 9054–9058

Zhu A, Guo W, Jain K, Mower JP (2014) Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Mol Biol Evol* 31: 1228–1236