

**Genetic analysis of metapopulation processes in *Silene latifolia***

Peter Damian Fields  
Whitesburg, KY

Bachelor of Arts, Alice Lloyd College, 2006

A Dissertation presented to the Graduate Faculty  
of the University of Virginia in Candidacy for the Degree of  
Doctor of Philosophy

Department of Biology

University of Virginia  
December, 2013

Committee Members:  
Douglas Taylor (Advisor)  
Edmund Brodie III (First reader)  
David Carr  
Henry Wilbur  
Janis Antonovics

## ABSTRACT

Molecular population genetics is one of the fastest growing areas of biological research, being of central importance to understanding human history, diversity, and the potential for personalized medicine. The field uses the fact that historical processes such as change in population size, rates and pathways of migration, and natural selection, all leave distinct footprints in an organism's DNA. Modern population genetics involves a rich arsenal of statistical tools that use DNA sequence data to make inferences about the evolutionary history of organisms and the genetic basis of their traits. My dissertation focuses on identifying the importance of selective and non-selective forces at different, hierarchically nested, biological levels in spatially structured (meta-)populations. In particular, my dissertation research explores how the evolutionary process is affected by the reality that populations of organisms are clustered in space. My dissertation research focused on a long-term metapopulation (25+ years now) of the angiosperm *Silene latifolia* and its sterilizing fungal pathogen *Microbotryum violaceum*. I have attempted to develop a multifaceted research program that includes fine-scale iterative sampling, where every individual within a extant population up to an expected genetic self-assignment asymptote of 50 individuals, from approximately 1/3 of 800 extant populations, distributed across three of nine metapopulation sub-sections: approximately 2000 individuals collected in 2008, 2010, and 2012. I have developed a high-throughput multi-plexed microsatellite genotyping protocol (20 microsatellites, from a panel of roughly 80 candidate loci, identified to encompass sufficient information for 1) assignment of individuals identified as colonists

through an annual census, and 2) identification of recent migration amongst extant populations). Finally, I have worked to develop statistical genetic software that allow the use of molecular marker data derived from different genomic regions, even if they have very different evolutionary histories and rates of mutation. This dissertation has derived a number of interesting inferences concerning the consequences of spatial population structure. Molecular marker specific evolutionary processes, and its concomitant variation, will determine both the type and accuracy of population genetic inferences available. The expectation of an inverse relationship between fitness and  $F$  will not always hold, and this discrepancy might be attributed to processes associated with population structure. Historical contingency will play a dominant role in determining the role of selective processes determining the quantity and distribution of neutral molecular variation, as well as asynchrony in the co-evolutionary process in spatially structured host-pathogen systems. Mean levels of cyto-nuclear disequilibrium are generally quite stable over up to seven generations of the focal plant species, though specific associations are quite labile and are highly affected by drift-like processes. And finally, intra-demic selection can be active in “everyday”, natural metapopulations, through hard selection acting on differential colonization, where different levels of selection are not necessarily in opposition to one another. Taken together, my dissertation aims to provide a unique window into the current and historical factors distributing population genetic diversity, but also how this structure will affect the evolution of the system itself. Continued long-term monitoring, iterative sampling, application of newer sequencing technologies, and the development of newer

analytical methodologies will continue to show interesting biological dynamics taking place as the result of spatial population structure.

## ACKNOWLEDGMENTS

There are many people that must be thanked for the help and guidance they have provided, allowing me to finish this dissertation. Though the exact origin of the adage “It takes a village” is unknown, I find it an apt description nonetheless. From the very beginning of my graduate work I have received enormous support from my advisor, Douglas Taylor. Just as important though was the environment Doug cultivated in the lab, including individuals that I can now count as close friends and colleagues. These include, but are certainly not limited to, Brian Sanderson, Dan Sloan, Dexter Sowell, Quinn McFrederick, Steve Keller, and Vijay Panjeti.

In addition to these colleagues at the University of Virginia, I had the great pleasure of traveling to the University of Grenoble, France, in order to work with Oscar Gaggiotti and Matthieu Foll on aspects of population genetics software packages. Even more, this initial opportunity provided the potentiation for a collaboration with John Pannell at the University of Lausanne, Switzerland, for a review of recent population genetics methods used to understand selection in subdivided plant populations, and a collaboration with Fred Austerlitz at Le Muséum national d'Histoire naturelle, Paris, France, in order to quantify dynamics of migration and colonization in the focal metapopulation of *Silene latifolia*. In the midst of these collaborations I also had the great pleasure of visiting a number of departments across Europe and the US, each time heartened by caliber of research taking place in the field of population genetics and evolutionary biology.

Throughout the past couple of years each of my committee members, Butch Brodie, David Carr, Henry Wilbur, and Janis Antonovics, have contributed to the improvement of this dissertation. I think the argument could be made that few committees were arranged in a manner as consistent with the particular themes of focus in one dissertation. I received additional support from the University of Virginia community, including Wendy Crannage, Katherine Holcomb, Jackie Huband, and Mark Morgan (and the University of Virginia Alliance for Computational Science and Engineering generally). My work would have been impossible if not for the support at Mountain Lake Biological Station. Financial support for my research originated from many sources, including the Office of the VPR, the Graduate School of Arts and Sciences, the Department of Biology, the National Science Foundation, and the Embassy of France in the United States, Office for Science and Technology.

I wish to also thank those individuals outside of the academic purview that also helped me grow in other ways. In particular I must thank Tim Brelinski and, though it saddens me to think he is no longer with us, Hugh Dornan. When I first moved into the “House of Yes” I was as they say quite green in matters of the kitchen. And while the HOY is no longer in existence, I will take the ideas it represented wherever I go.

I have to thank my undergraduate advisor, Richard Stiehl, for introducing me to the works of Stephen J. Gould, for showing me that research in evolutionary biology can be a career, and for being a friend throughout my time at Alice Lloyd College.

Without question, the individual that has been most important in keeping me sane over the past couple of years is my partner Andrea Berardi. I cannot fathom where I'd be without her love and support.

More than anyone else in the world, I must thank my mother for all that I have today. She has given so much, and ask for nothing more than that I be successful and happy. She is truly one of the strongest people I have ever known. I wish more people could have a parent as great as her, for in such a reality the world would be a better place for all.

## TABLE OF CONTENTS

<b>Title Page .....</b>	<b>1</b>
<b>Abstract .....</b>	<b>2</b>
<b>Acknowledgements .....</b>	<b>5</b>
<b>Table of Contents .....</b>	<b>8</b>
<b>Introduction .....</b>	<b>10</b>
 <b>Chapter 1:</b> Evolution in subdivided plant populations: concepts, recent advances and future directions .....	 <b>19</b>
 <b>Chapter 2:</b> Determinants of genetic structure in a nonequilibrium metapopulation of the plant <i>Silene latifolia</i> .....	 <b>88</b>
 <b>Chapter 3:</b> Molecular-marker specific determinants of population genetic structure in a metapopulation of <i>Silene latifolia</i> .....	 <b>119</b>
 <b>Chapter 4:</b> Population history and differential consequences of inbreeding and outcrossing in a plant metapopulation .....	 <b>150</b>
 <b>Chapter 5:</b> Variation in the phenotypic and population genetic consequences of co- evolution in a host and pathogen metapopulation .....	 <b>191</b>
 <b>Chapter 6:</b> Patterns of cyto-nuclear linkage disequilibrium in <i>Silene latifolia</i> : genomic heterogeneity and temporal stability .....	 <b>220</b>



<b>Chapter 7:</b> Population determinants of migration and evidence for hard selection during colonization in a plant metapopulation .....	<b>250</b>
 <b>Appendix 1:</b> Isolation and characterization of microsatellite loci in the white campion, <i>Silene latifolia</i> (Caryophyllaceae) .....	<b>301</b>
 <b>Appendix 2:</b> Transcriptome resources for two highly divergent <i>Silene latifolia</i> populations .....	<b>310</b>
 <b>Appendix 3:</b> LinkC: A software for the measurement of variance component estimation of linkage disequilibria in subdivided populations .....	<b>318</b>
 <b>Appendix 4:</b> Modification in BayeScan to incorporate GESTE-like environment association .....	<b>324</b>

## INTRODUCTION

The study of geographic population structure and its consequences was introduced by Sewall Wright, in particular the construction of the Island Model (WRIGHT 1931). In the Island Model, populations are assumed to be stable and interconnected by migration, with population differentiation ( $F_{ST}$ ) generated by drift and diminished by gene flow. Specifically, population differentiation is defined by

$$F_{ST} = \frac{1}{4Nm + 1}$$

where  $N$  is population size and  $m$  is the rate of migration amongst populations. Wright's Island Model includes a number of limiting assumptions, including no selection or mutation, as well as constant population size, equal contributions of each population to the migrant pool, where migration is random, and, importantly, equilibrium has been reached. In order for analytical models of geographic population structure to accommodate realities of natural populations of biological interest, a number of additional refinements to the Island Model were required. In particular, allowance for multiple alleles was introduced by Nei (1973), with:

$$F_{ST} = \frac{1}{4Nm + 4N\mu + 1}$$

where  $\mu$  is the mutation rate of a given locus. A contrast to this equilibrium model is an idealized metapopulation (WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990), where demes experience extinction and recolonization. In metapopulations, the level of population differentiation will be described by

$$F_{ST} = \frac{\left[1 - e + e\phi\left(1 - \frac{1}{2k}\right)\right] / (2N + \frac{e}{2k})}{1 - [1 - e + e\phi\left(1 - \frac{1}{2k}\right)](1 - m)^2(1 - \frac{1}{2N})}$$

where  $e$  is the rate of population extinction,  $k$  is the number of founding propagules, and  $\phi$  is the number of sources from which the  $k$  founding propagules (WHITLOCK and McCAULEY 1990). Chapter 1 of this thesis presents a broad and detailed review of our current understanding of evolution of plants in subdivided populations, as well as the methodologies that have and will be used to detect and characterize it (PANNELL and FIELDS 2013). I was invited by John Pannell (University of Lausanne, Switzerland) to compose this review as part of the Tansley Review series for the New Phytologist. My contribution focused on bridging more traditional methodologies with those that have been derived recently.

As described in Chapter 1, models of how geographic population structure affects the evolutionary process are quite numerous, while the number of empirical studies is quite limited. My dissertation focuses on the angiosperm *Silene latifolia*, and in particular a metapopulation located in Giles and Craig Counties, Virginia, USA. This region has been the subject of a 25+ year study of population dynamics and genetic structure in more than 800 spatially distributed populations (ANTONOVICS *et al.* 1994; McCAULEY 1994; McCAULEY 1997; McCAULEY *et al.* 1995; RICHARDS 2000; RICHARDS *et al.* 1999; RICHARDS *et al.* 2003). An annual census of an approximately 25x25-km area adjacent to Mountain Lake Biological Station has been conducted since 1988, recording the location, gender, of *S. latifolia* individuals along ~150km of predominantly roadside habitat. The structure of the data and how it was collected is reported in Antonovics *et al.* (1994). Briefly, the roadside habitat is divided up into ~40m segments referred to as psilons. Which psilons are occupied and the number of plants in each occupied psilon are recorded annually. Important

phase transitions, such as extinction and colonization events, are confirmed with a second census during the same season. Time since colonization (or population age) is based on the year plants were first observed in a given psilon. I identified extinction as the disappearance of plants from a psilon for a single year, though a previous study in the focal metapopulation has indicated the presence of seed bank with a possible viability of approximately four years (PERONI and ARMSTRONG 2001). The census data provide the demographic data, spatial relationships among populations, and extinction/recolonization dynamics used in this study.

I sampled plants from 77 spatially distinct populations during peak flowering in the summer of 2008, spanning  $\sim 1/3$  of the focal metapopulation. I collected leaf tissue from every plant in the population, or up to 50 individuals in the largest populations, and stored the leaves in silica gel. These individuals were genotyped at between 16 and 19 microsatellite loci. The origins of the microsatellites are described in a number of separate publications (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; MOCCIA *et al.* 2009; TEIXEIRA and BERNASCONI 2007). Chapters 2 and 3 present an application of a novel Bayesian modeling framework, the so-called *F*-model, to discern the direction and magnitude of effect of spatiotemporal features of populations that are generating population differentiation, and potentially driving metapopulation dynamics in a metapopulation of *Silene latifolia*. These chapters reveal the interesting result of differential inferences based upon molecular marker origin, and the failure of current population genetic models to accommodate such heterogeneities. I present a model in Chapter 3 based upon the identification of inbreeding avoidance in the focal species (AUSTERLITZ *et al.* 2012; RICHARDS *et al.*

1999) as well as differential explanatory power of molecular markers derived from heterogeneous genomic regions for identifying Heterozygosity Fitness Correlations (or HFCs) (OLANO-MARIN *et al.* 2011).

In addition to the population genetic analyses presented in earlier chapters, in Chapter 4 I describe a paired approach of population genetic analysis and controlled crosses in order to determine how spatial population structure leads to variation in the consequences of inbreeding. Specifically, the exact consequences of inbreeding are shown to be highly dependent on spatiotemporal character, with population isolation generating outbreeding depression over unexpected spacial and time scales. These results are consistent with recent experimental work implicating founder effects in generating large shifts in allele frequencies, and, potentially, complex epistatic associations.

Chapter 5 investigates the direct impact of metapopulation structure on natural selection for a particular phenotype, disease resistance, in *S. latifolia* to its sterilizing pathogen, *Microbotryum violaceum*. Using both direct inoculations and population genetic analysis of the 2008 sample, I show that historical presence of the pathogen leads to distinct evolutionary outcomes, with spatial and temporal variation in disease resistance.

In addition to the 2008 samples, my dissertation included a collection of samples derived from 1993 which were collected as part of earlier studies of the focal metapopulation (MCCAULEY 1994). A subset of the 2008 samples that directly overlapped with the 1993 samples were genotyped for a single-nucleotide polymorphism (SNP) in the mitochondrial gene, *atp1*, and the 1993 samples were

genotyped for 14 microsatellite loci as well as the mitochondrial SNP. By utilizing collections taking place more than +10 years apart, I show in Chapter 6 that, contrary to limited theoretical exploration, cyto-nuclear disequilibrium in a metapopulation is quite stable over up to seven generations. Because the utilized markers are most likely of neutral in regards phenotype, I show that under extinction/colonization dynamics associated with metapopulation structure, population genetic signatures previously interpreted as being indicative of selection may in fact be the result of neutral, drift-like processes (FIELDS *et al.* 2013).

Following the 2008 sample, I conducted repeated samples of the same populations in 2010 and 2012, as well as sampling new populations (or colonists) in 2009, 2011, and 2013 (data from 2012 and 2013 are not included in this dissertation). In Chapter 7, having developed a panel of highly informative molecular markers and utilizing these iterative sampling routine of populations over multiple generations of the focal species, we were able to measure and determine the causal factors of migration and colonization success. Most significantly, I uncovered evidence for hard selection (WALLACE 1975) determining colonization success, as modulated through a disproportionately larger proportion of colonists originating from populations of higher genetic diversity, controlling for population size.

The appendices in describe a number of resources potentially of use to both *Silene* researchers, and population geneticists generally. Appendices 1 and 2 describe molecular resources for the focal species, both microsatellites and SNPs, respectively. Appendices 3 and 4 describe pieces of software that can be used for

partitioning of linkage disequilibrium in spatially structured populations (LinkC) and the estimation of the determinants of population structure, taking into account heterogeneous marker processes such as mutation rate variation (GESTE v.3), respectively.

Molecular population genetics is one of the fastest growing areas of biological research, being of central importance to understanding human history, diversity, and the potential for personalized medicine. The field uses the fact that historical processes such as change in population size, rates and pathways of migration, and natural selection, all leave distinct footprints in an organism's DNA. Modern population genetics involves a rich arsenal of statistical tools that use DNA sequence data to make inferences about the evolutionary history of organisms and the genetic basis of their traits. My research improves and employs high throughput analytical methods in molecular population genetics, including those that identify genetic variants responsible for variation in traits (so called genome wide associations studies, or GWAS) and those that detect histories of population growth or patterns of mating (as in the quantification of the genetic mixing between Neanderthals and humans).

## **LITERATURE CITED**

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010 Permanent genetic resources added to Molecular Ecology resources database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in *Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.
- AUSTERLITZ, F., G. GLEISER, S. TEIXEIRA and G. BERNASCONI, 2012 The effects of inbreeding, genetic dissimilarity and phenotype on male reproductive success in a dioecious plant. *Proceedings of the Royal Society B-Biological Sciences* **279**: 91-100.
- FIELDS, P. D., D. E. MCCAULEY, E. V. MCASSEY and D. R. TAYLOR, 2013 Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability. *Heredity*.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.
- MCCAULEY, D. E., 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.



- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- NEI, M., 1973 Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences* **70**: 3321-3323.
- OLANO-MARIN, J., J. C. MUELLER and B. KEMPENAERS, 2011 Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci. *Molecular Ecology* **20**: 4028-4041.
- PANNELL, J. R., and P. D. FIELDS, 2013 Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist*: n/a-n/a.
- PERONI, P. A., and R. T. ARMSTRONG, 2001 Density, dispersion and population genetics of a *Silene latifolia* seed bank from southwestern Virginia. *Journal of the Torrey Botanical Society* **128**: 400-406.
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.
- RICHARDS, C., S. CHURCH and D. MCCAULEY, 1999 The Influence of Population Size and Isolation on Gene Flow by Pollen in *Silene alba*. *Evolution* **53**: 63-73.
- RICHARDS, C. M., S. N. EMERY and D. E. MCCAULEY, 2003 Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity* **90**: 181-186.

- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.
- WADE, M. J., and D. E. McCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WALLACE, B., 1975 Hard and Soft Selection Revisited. *Evolution* **29**: 465-473.
- WHITLOCK, M. C., and D. E. McCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.
- WRIGHT, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97-159.

Chapter 1: Evolution in subdivided plant populations: concepts, recent advances and future directions<sup>1</sup>

---

<sup>1</sup> Formatted for submission as a coauthored manuscript: Pannell, J.R. and P.D. Fields

**Abstract**

Research into the evolution of subdivided plant populations has long involved the study of phenotypic variation across plant geographic ranges and the genetic details underlying that variation. Genetic polymorphism at different marker loci has also allowed us to infer the long and short-term histories of gene flow within and among populations, including range expansions and colonisation-extinction dynamics. But the advent of affordable genome-wide sequences for large numbers of individuals is opening up new possibilities for the study of subdivided populations. In this review, we consider what the new tools and technologies may allow us to do. In particular, we encourage researchers to look beyond description of variation and to use genomic tools to address new hypotheses, or old ones afresh. Because subdivided plant populations are complex structures, we caution researchers away from adopting simplistic interpretations of their data, and to consider the patterns they observe in terms of the population genetic processes that have given rise to them; here the genealogical framework of the coalescent will continue to be conceptually and analytically useful.

Running Head: Evolution in subdivided populations

## INTRODUCTION

Plant populations are often separated from one another by areas of unsuitable habitat over which migration and gene flow is limited. Even populations that occupy apparently homogeneous habitat over large areas can be structured because of limited dispersal and local mating (NEIGEL 1997). Groups of individuals occupying different parts of a species' range can thus end up evolving relatively independently of one another under the influence of drift and local selection. It is becoming increasingly clear that even within continuous populations environmental heterogeneity can bring about fine-scaled genetic structure with the evolution of local adaptation (AUDIGEOS *et al.* 2013).

The extent to which plant populations are genetically divergent depends on the balance between processes that drive them apart and those that homogenize them. For neutral loci, surprisingly small amounts of gene flow can prevent much genetic divergence between demographically stable populations (HARTL and CLARK 1997; SLATKIN 1985; SLATKIN 1987; WRIGHT 1931). For loci under selection, genetic divergence can be restricted or enhanced, relative to the neutral case, depending on whether different phenotypes are selected in different populations or whether the same phenotype is selected globally. An important corollary is that because action of selection is expected to vary among loci, the magnitude of population divergence will be locus-specific. This complicates how we describe population structure, but the differences in population structure among loci provide us with a powerful means to tease apart the effects of natural selection from drift and the effects of demographic processes or history.

Variation in divergence among loci reminds us not only that plant populations are fragmented geographically, but also that the genome is fundamentally fragmented, too. As Darlington (quoted in LEWONTIN 1980) put it, “the really important small populations are the little bits of chromosomes that are populations within which recombination cannot occur”. Here, Darlington was referring to inverted chromosomal regions in which recombination is completely suppressed, so that there are effectively two different populations of genes at the same locus that do not mix by recombination. But even loci in genomic regions that continue to recombine may be in gametic disequilibrium, i.e., may be associated with each other non-randomly. Whether locus-specific divergence evolves in response to selection for local adaptation will thus often depend on genetic correlations among traits (ETTERSON and SHAW 2001).

With rapid advances in sequencing technologies and high-throughput data analysis, subdivided plant populations are providing new opportunities to study the evolutionary forces influencing genetic divergence across the genome. Population genetic studies have benefitted from using an increasing number of individuals and loci (LASCoux and PETIT 2010). Guichoux *et al.* (2011) recently identified approximately 8,000 published population genetic analyses utilizing simple sequence repeats (SSRs) and single nucleotide polymorphism (SNPs) in 2009 alone, many assessing the presence and consequences of population subdivision. Continued advances in second-generation sequencing technologies and analytical methods promise to accelerate these trends (CAO *et al.* 2011; WANG and HEY 2010). The time is ripe to consider what these studies can tell us about the evolution of

plants across their fragmented landscapes, and what sorts of questions we might now address.

In this article, we review advances in our understanding of the evolution of subdivided plant populations from a conceptual point of view, beginning with a non-technical discussion of effective population sizes, migration, and the characterisation and interpretation of genetic structure, which has often been measured in terms of the statistic  $F_{ST}$ . We discuss the utility of  $F_{ST}$  as a basis for inferring the demographic and selective history of populations, we highlight new studies that are moving beyond the use of  $F_{ST}$  to infer a population's evolutionary history, and we ask how new data, and the new ways of dealing with it, allow us to understand the distribution of genetic variation across subdivided plant species.

## **EFFECTIVE POPULATION SIZES, GENETIC DRIFT AND MIGRATION**

The effective population size,  $N_e$ , is a parameter that enters into many expressions in population genetics, often as a product with other parameters such as the absolute mutation rate ( $N_e u$ ), the absolute migration rate ( $N_e m$ ), the selection coefficient ( $N_e s$ ), and the recombination rate ( $N_e r$ ). It is thus an important scaling parameter for the evolutionary forces that affect a population's evolution.  $N_e$  is sometimes loosely defined in terms of the number of individuals contributing genetically to future generations. But how far into the future, and contributing what? While the number of breeding individuals will influence  $N_e$ , it is more useful to think of the effective population size as a parameter that determines the extent to which the population is subject to genetic drift: the smaller  $N_e$ , the greater the extent to which genetic

measures that interest us will be affected by drift. With this in mind, it should be no surprise that there are several effective sizes, each addressing a different effect of drift. The inbreeding effective size determines how quickly populations become inbred through simple random mating (relatives are more likely to mate with one another in populations with small  $N_e$ ); the variance effective size determines the extent to which allele frequencies fluctuate from generation to generation (allele frequencies fluctuate more in populations with small  $N_e$ ); the mutation effective size determines how quickly the genetic diversity at a particular locus should equilibrate to a new drift-mutation equilibrium (populations with small  $N_e$  maintain less diversity). With the focus on drift, it should also be clear why  $N_e$  is locus-specific: loci under selection (or those linked to them) will be subject to different fluctuations in frequency to those unaffected by selection.

The effect of drift on genetic diversity is complicated by population subdivision. To see this, it is helpful to consider drift from the perspective of the coalescent, i.e., by tracing lineages backwards in time through their genealogy (HEIN *et al.* 2005; HUDSON 1990; WAKELEY 2009) (Figure 1). In a population subdivided into small populations (or demes) linked by little migration, individuals in each deme will be more closely related to one another, on average, than to individuals in other demes. Thus, the initial coalescence of lineages sampled from the same deme will tend to be rapid, occurring at a rate determined by the local inbreeding effective size of the deme (e.g., all the coalescence events among lineages sampled in deme 1 in Figure 1). Nevertheless, as long as the migration rate into the sampled deme is not zero, there is a chance that one or more of the lineages in our sample are recent



migrants (e.g., lineages sampled in deme 2 of Figure 1); in this case, tracing back to the common ancestor of our sample requires us to follow lineages until they find themselves in the same deme again, at which point they will coalesce at a rate given by the inbreeding effective size of that deme (SLATKIN 1991). There are thus two phases to the coalescent in a subdivided population, each pointing to a different effective population size: the local inbreeding effective size (the short-term rate of coalescence) and the eigenvalue effective size of the whole metapopulation (the long-term rate of coalescence, which will be inversely related to the migration rate). These two phases have been termed the ‘scattering phase’ (referring to the migration of lineages out of sampled demes as one traces their ancestry into the past) and the ‘collecting phase’ (the migration of lineages back the same demes prior to their ultimate coalescence) by Wakeley (2000; 2001). The two phases can in principle be discerned by ‘skyline’ plots (STRIMMER and PYBUS 2001) of the consecutive waiting times until coalescent events for the whole sample (scaled appropriately; Figure 1), or by assessing the frequency spectrum of nucleotide differences for sequences sampled from the individuals concerned (c.f. Figure 2).

While patterns in DNA sequence diversity provide a potentially powerful means of reading the history of a sample and the population from which it was drawn, quite different processes can give rise to very similar patterns. For example, moderate genetic bottlenecks of single populations can distort the shape of the coalescent in ways similar to the effect of population subdivision (Figure 2): the rate of coalescence is increased during a bottleneck, but lineages that fail to coalesce during the bottleneck may require extended periods to coalesce prior to it,

similar to the long coalescent times observed for lineages sampled from different populations. Some of the complexities of the structured coalescent can be avoided by sampling just a single individual from each deme, though this will limit the conclusions that might be drawn about population subdivision itself (WRIGHT and GAUT 2005). (In principle, it is possible to detect a genetic signature of population subdivision by sampling individuals from only a single deme if the sample includes genes that share no local common ancestor, because their ancestors migrated into the sampled deme.)

In species in which deme sizes vary over time, especially in the extreme case where demes become extinct and are later recolonized (a 'metapopulation'), the effective population size can be dramatically reduced below the total metapopulation census size, depending on the migration rate (GILPIN 1991; MARUYAMA and KIMURA 1980; PANNELL and CHARLESWORTH 1999; WAKELEY 2001; WAKELEY and ALIACAR 2001; WHITLOCK and BARTON 1997). In a metapopulation in which the extinction-recolonization rate exceeds the migration rate, a lower effective population size is expected, and we should see reduced diversity both within demes and in the species as a whole (PANNELL and CHARLESWORTH 2000b; SLATKIN 1977; WHITLOCK and BARTON 1997). These predictions are nicely illustrated by comparisons of the genetic structure for maternally versus bi-parentally inherited genes: the former may retain the signature of colonisation if seed dispersal is limited, while genetic structure for the latter is eroded by pollen dispersal. Such contrasting patterns have been found in both herbs (DE CAUWER *et al.* 2010; e.g., MCCAULEY 1994; MCCAULEY 1997; MCCAULEY 1998) and trees (e.g., PETIT *et al.* 1997).

The above rules of thumb are useful, but they belie the potential demographic complexities of a metapopulation. Even the simplest models that assume demes of similar size and extinction probability include several parameters that all have an important effect on the effective size and genetic diversity, including the migration and extinction rates, the number of demes and their sizes, the number of founding individuals, and the extent to which they come from the same source deme or a mix of different demes (PANNELL 2003; PANNELL and CHARLESWORTH 2000a; WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990). Because different processes can affect particular summary statistics in similar ways, our challenge is to devise sampling and analysis that allow us to distinguish them, e.g., to move beyond the use of single statistics that only summarise part of the pattern.

### **POPULATION DIFFERENTIATION, AND HOW BEST TO MEASURE IT**

Because evolutionary dynamics will depend upon a species' past, current, and future population genetic structure (MEIRMANS 2012), identification of the magnitude of population genetic differentiation has been a central component of biological research. Population differentiation across the genome as a whole may indicate low historical levels of gene flow, while differences in population differentiation among traits or loci may point to a history of divergent selection and local adaptation (see next section). Measures of population differentiation are thus arguably more interesting than are estimates of the species' effective population size, and are more straightforward to estimate.

$F_{ST}$  is the most familiar and widely employed measure of population genetic differentiation, not least because it is embedded so deeply in the theoretical population-genetics literature (GUO *et al.* 2009; ROUSSET 2013; WHITLOCK 2011a). Nevertheless, its use and interpretation need care, not least because  $F_{ST}$  has been defined and derived in ways that are equivalent only under certain assumptions.  $F_{ST}$  was first introduced by Sewall Wright as an index of inbreeding, to capture the notion that individuals in the same deme tend to be more closely related to one another than to individuals of other demes, and that dispersal or migration among demes will tend to break up these relationships. Because all individual gene copies trace back to a common ancestor (i.e., all individuals are ultimately related), it is natural that an index of inbreeding imposed by population structure should incorporate estimates of co-ancestry within populations and in the population as a whole, as  $F_{ST}$  does. We should also expect  $F_{ST}$  as an inbreeding coefficient to be expressible in terms of within-deme and species-wide coalescence times. Indeed, Slatkin (1991) showed that

$$F_{ST} = \frac{t_T - t_S}{t_T},$$

where  $t_S$  and  $t_T$  are the expected times to coalescence of two genes sampled from the same population and from the global population, respectively; an estimate of this parameter will be accurate when the mutation rate is low.

Whitlock (2011a) has usefully referred to this inbreeding, coalescent,

perspective of  $F_{ST}$  as  $F_{ST,coal}$  to emphasise its distinction from other definitions; Rousset (2013) has labeled it  $C_{ST}$ . Importantly,  $F_{ST,coal}$  is defined without reference to gene diversity and is independent of the mutation rate; it depends only on processes that affect the shape of the underlying genealogy, such as dispersal and population size (including fluctuations caused by bottlenecks and expansions). For demographic and evolutionary inference, it thus tends ultimately to be  $F_{ST,coal}$  that we wish to estimate, though it is less useful if we wish to identify demes of especially distinctive allele composition at loci of particular interest for conservation purposes, in which case Jost's  $D$  might be more useful (Jost 2008a).

Of course, we can rarely know the history of co-ancestry for a population and thus have to turn to genetic markers for help. This is where difficulties with  $F_{ST}$  arise, because the distribution of variation at genetic marker loci depends not just on the demographic processes of drift and migration, but also on mutation. For loci with two alleles,  $F_{ST}$  was defined by Wright (1943) in terms of the variation among populations in allele frequencies. Later, Nei (1973) derived an expression for  $F_{ST}$ , which he termed  $G_{ST}$ , that is applicable to loci with multiple alleles. A proper multi-allelic estimate of  $F_{ST}$  based on allelic variance was introduced by Weir and Cockerham (1984).  $G_{ST}$  is roughly equivalent to  $F_{ST}$  as an expression of the variance of allele frequencies; it also approximates  $F_{ST,coal}$  (and thus the genealogical structure of the population at the sampled loci, as determined by drift and migration) as long as the mutation rate is low.

Because  $F_{ST}$  is a relative measure of diversity, it presents a number of difficulties for interpretation. First, processes that reduce  $N_e$  locally, including

background selection in non-recombining regions of the genome or inbreeding, will necessarily lead to high  $F_{ST}$  at those loci (CHARLESWORTH 1998). As we discuss below, this is particularly relevant for interpreting patterns of variation in  $F_{ST}$  across the genome. In cases where genetic diversity has been reduced by drift locally, Charlesworth (1998) has thus warned against relative measures of differentiation such as  $F_{ST}$  for genomic regions with different levels of recombination or different mating systems (and thus potentially different  $N_e$ ) and argued in favor of using absolute measures of differentiation, such as the difference between within- and between-population diversity. A second difficulty occurs when  $F_{ST}$  is calculated for loci with high allelic diversity, where the upper bound for  $F_{ST}$  is substantially below one, such that populations that are strongly differentiated (e.g., they shared few alleles) can have low  $F_{ST}$  values (JOST 2008a). In general,  $F_{ST}$  can thus be unhelpful as a measure of genetic differentiation, particularly when one is keen to compare different species or different loci that show markedly different levels of genetic diversity (and for which the upper bound for the differentiation statistic differs as a result) (CHARLESWORTH 1998; HEDRICK 2005b; JOST 2008a). A related difficulty arises when  $F_{ST}$  is based on loci with high mutation rates, which potentially obscure the genealogical relationships among individuals because of homoplasy, so that  $F_{ST}$  fails to reflect the coalescent history of the sample and the demographic processes influencing it. Microsatellites are much more prone to this problem than, for example, SNPs.

New measures of population differentiation have been proposed address the problem of high allelic diversity, the relative merits of which have been discussed at

length (ROUSSET 2013; RYMAN and LEIMAR 2009; WANG 2012; WHITLOCK 2011b). Hedrick (2005a) introduced a standardized measure,  $G'_{ST}$ , by dividing by the maximum possible value of  $G_{ST}$  for the observed allele frequencies globally, and Jost (2008b) proposed a new measure of population differentiation,  $D$ , that has the intuitively appealing property of reaching a maximum when each allele is private to a single population (JAKOBSSON *et al.* 2013); new software has been developed to estimate these alternatives to  $F_{ST}$  (CRAWFORD 2010; MEIRMANS and HEDRICK 2011; WINTER 2012). Nevertheless,  $F_{ST}$  has the advantage over indices such as  $D$ , because it is a well-defined parameter that is connected to the theoretical literature allowing demographic and evolutionary inference (ROUSSET 2013; WHITLOCK 2011b).

It is well known that  $F_{ST}$  can be related to the number of individuals migrating between sub-populations ( $Nm$ ) according to the equation  $F_{ST} \approx 1/(4Nm + 1)$  (WRIGHT 1931). As Whitlock and McCauley (1999) re-emphasised, the utility of this expression, however, requires not only that populations have reached drift-migration equilibrium, but also that all demes have the same constant size and equal migration rates. Such conditions are probably rarely met by real species. In a metapopulation, for example, rapid population turnover is predicted to increase genetic differentiation among demes (WADE and MCCAULEY 1988). Population turnover can be incorporated into models, but one quickly faces the problem of over-parameterisation of models that are biologically plausible, even though adequate sampling can overcome some of the difficulties (see STÄDLER *et al.* 2009).

Estimators formalized around the so-called  $F$ -model may provide insights beyond those revealed by  $F_{ST}$ . The  $F$ -model, which is likelihood-based approach that

defines  $F_{ST}$  as a parameter of the full distribution of allele frequencies (BALDING and NICHOLS 1995; BHATIA *et al.* 2013; GAGGIOTTI and FOLL 2010; KARHUNEN and OVASKAINEN 2012; NICHOLSON *et al.* 2002) accommodates differences in population size and migration rates across a species range (GAGGIOTTI and FOLL 2010), and thus have advantages over  $F_{ST}$  that estimates a single “global” value of differentiation. Foll and Gaggiotti (2006) introduced a hierarchical formulation that uses population-specific measurements to obtain priors for  $F_{ST}$ , and that then estimates the proportional contribution of population-specific drift and migration to characterise population genetic structure. For example, to estimate the contribution of extinction and recolonization to population structure, the  $F$ -model can be applied to a long-term data set recording the size and spatial distribution of demes along with knowledge of their demographic history and age structure. Jay *et al.* (2012) recently applied an  $F$ -model based approach to assess how ecological characteristics of twenty alpine plant species determined each respective population’s shared co-ancestry, and allowing a prediction of how future climate change might alter the magnitude and distribution of population divergence.

### **$F_{ST}$ AS A BASIS OF INFERRING LOCAL ADAPTATION: NEUTRAL GENES VS. PHENOTYPES**

Individuals from different populations of a species often differ phenotypically. To discern whether such phenotypic variation is the result of drift vs. natural selection, Spitze (1993) introduced the  $F_{ST}$ - $Q_{ST}$  comparison.  $Q_{ST}$  is a phenotypic analogue to  $F_{ST}$  and estimates the additive genetic variation among populations for the trait in



question as the total genetic variation among populations divided by the among-population variation plus twice the additive genetic variation within populations. If trait divergence is due solely to random processes,  $Q_{ST}$  and  $F_{ST}$  should be similar; in contrast,  $Q_{ST} > F_{ST}$  or  $Q_{ST} < F_{ST}$  should reflect divergent or globally stabilizing selection, respectively. Yu et al. (2011) recently adopted this approach to understand the basis of variation in flower size and inflorescence variation in the dioecious herb *Silene latifolia*.

*S. latifolia* is sexually dimorphic for a number of traits, the extent to which varies among populations (e.g., males produced smaller, more numerous, flowers than females) (DELPH *et al.* 2011; STEVEN *et al.* 2007). Yu et al. (2011) measured among- and within-population quantitative genetic variation and covariation in calyx width (the most sexually dimorphic trait), and compared this variation to global  $F_{ST}$  derived from microsatellite markers. Interestingly, the ratio of phenotypic to neutral genetic differentiation was 4.2 for males and only 0.4 for females, suggesting that selection on only one of the sexes may be largely responsible for the degree of among-population divergence in calyx width – or that selection has taken place on other traits that are genetically correlated with calyx width (Figure 3). A multivariate form of this classic test, derived by Martin et al. (2008) and Chapuis et al. (2008), allows for the inclusion of such potential among-trait covariances.

The direct  $F_{ST}$ - $Q_{ST}$  comparison is useful only when  $Q_{ST}$  estimates the additive genetic component of phenotypic divergence between populations, and is not influenced by phenotypic plasticity. Measuring traits in the field is thus problematic, because individuals from different populations may express different phenotypes in

response to environmental cues (e.g., see PUJOL *et al.* 2008; WHITLOCK and GUILLAUME 2009). Ideally, plants need to be measured growing in a common garden or glasshouse, as in the study of *S. latifolia* by Yu *et al.* (2011). However, Antoniazza *et al.* (2010) proposed an approach by which the distribution of among-population variation in phenotypes measured in the field, or  $P_{ST}$ , might be substituted for  $Q_{ST}$ . Conclusions from such an analysis must still remain somewhat limited, but the use of independently derived estimates of trait heritability may alleviate much of the concern presented by the  $P_{ST}$  approach (ANTONIAZZA *et al.* 2010),

Because  $F_{ST}$ - $Q_{ST}$  comparisons clearly require estimates of  $F_{ST}$ , its limitations are also relevant here. For example, because  $F_{ST}$  is biased downward for loci with higher mutation rates, we ought to expect an enrichment of studies inferring divergent selection among populations based on microsatellites compared with isozymes (EDELAAR *et al.* 2011). To deal with this issue, Edelaar (2011) suggested the use of an estimator of neutral genetic divergence that corrects for molecular marker heterozygosity, such as  $G'_{ST}$ , or  $D_{EST}$  (though theory that relates these statistics to  $Q_{ST}$  still needs to be developed) or to use estimators that are not affected by the mutation rate. Another solution is simply to avoid the problem by estimating  $F_{ST}$  using markers with lower mutation rates, such as SNPs or allozymes (EDELAAR *et al.* 2011).

Ovaskainen *et al.* (2011) and Karhunen and Ovaskainen (2012) have recently taken a new perspective on the standard  $F_{ST}$ - $Q_{ST}$  comparison (Figure 4). Their approach uses an extended  $F$ -model-based estimator, the admixture  $F$ -model (AFM), to construct a matrix of population co-ancestry, simultaneously disentangling the

role of local drift and gene flow on the basis of population-specific deviations in allele frequencies. The parameters of the co-ancestry matrix can be used to estimate population means for quantitative traits under neutral expectations. Because the full matrix of population associations is accounted for, the effects of drift and selection can also be identified even if  $F_{ST} \neq Q_{ST}$ . This method, which bolsters the continued relevance of  $F_{ST}$ , will be ideal for understanding the role of neutral and selective processes in species range expansions and biological invasions (Figure 4).

### **INFERRING LOCAL ADAPTATION: NEUTRAL VS. SELECTED GENES**

Genetic differentiation between populations can be the result of differential selection among habitats or of genetic drift, potentially enhanced by demographic processes (LUIKART *et al.* 2003). Because demographic processes affect all loci, while selection should affect only loci responsible for fitness or closely linked loci, comparisons among loci provide a potentially revealing way to distinguish between the two types of process. One useful approach involves comparison among loci for outliers in their  $F_{ST}$  against the full, observed distribution of  $F_{ST}$  across the genome (BEAUMONT and NICHOLS 1996; LEWONTIN and KRAKAUER 1973; PRUNIER *et al.* 2012). Outlier loci with unusually high  $F_{ST}$  are then candidates for genomic regions involved in local adaptation. Another idea is to generate a null distribution for genetic differentiation across the genome by simulating differentiation statistics for independent loci conditioned on the heterozygosity actually observed, e.g., using a coalescent framework (THORNTON and JENSEN 2007). This approach is particularly useful for cases where the number of loci in a dataset is modest (i.e., where the data

are insufficient to provide a robust distribution), but it has the drawback that incorrect demographic models can lead to the identification of wrong loci (OETJEN and REUSCH 2007).

It is important to recall that processes other than local adaptation can give rise to variation in  $F_{ST}$  across the genome. For instance, Charlesworth (1998) suggested that high  $F_{ST}$  observed in regions of low recombination in *Drosophila* could be accounted for by processes such as background selection and may have nothing to do with local adaptation. Similarly, in a recent study of patterns of genomic divergence among four species of sunflower, Renaut et al. (2013) found that genomic islands of high divergence did not correspond to lower effective gene flow (as might have been expected for loci under selection for local adaptation); rather, because such regions tended to also have low recombination rates, forces reducing  $N_e$  were more likely to be responsible.

Another general approach that contrasts with  $F_{ST}$  scans is to identify candidate loci by seeking associations between allele frequencies and particular habitats (or across environmental gradients), with comparisons made against the distribution of allele-habitat associations over the whole genome (BIERNE *et al.* 2011; HEDRICK *et al.* 1976). Genetic-environment correlations have been particularly useful in determining the co-variation of particular allelic variants and climatic variables, thereby pointing to possible genes responsible for adaptation to variation in temperature, moisture availability or variables that co-vary with latitude and longitude. A particular revealing example is provided by Eckert et al.'s (2010) study

of genetic-environment correlations across the species range of Loblolly Pine (*Pinus taeda*) in North America (ECKERT *et al.* 2010).

In their study of Loblolly Pine, Eckert *et al.* (2010) sought associations between genetic loci and environmental variation across the entire species range. By statistically removing correlations due to co-ancestry and range expansions, the authors identified five loci that were correlated with aridity gradients. All five loci are known to have stress-related functions in *Arabidopsis thaliana*, but, revealingly, none of these loci were among the 24 loci identified from the same data set through  $F_{ST}$  outlier analysis. More recently, Frichot *et al.* (2013) applied a similar approach to a subset of the same dataset and identified genes associated with wound repair and immunity, photosynthetic activity, carotenoid biosynthesis, cellular respiration, carbohydrate metabolism, and responses to heat, salt, and oxidative stress. The approach taken in these studies is powerful, because it is capable of revealing even small environmental correlations for loci, specified *a priori*, that are likely to be targets of selection along gradients. Selection on quantitative traits can bring about large changes in phenotypes due to only small changes in allele frequencies at many loci (BARTON and TURELLI 1989), and it is satisfying that these sorts of associations can still be found with appropriate sampling.

Analysis of genetic-environmental associations can point to particular environmental factors that might have been involved in the selective process. For example, in a study of local adaptation in pines, this time with the Mediterranean conifers *Pinus pinaster* and *P. halepensis*, Grivet *et al.* (2011) identified different sets of genes in each species as likely targets of species (with only one locus in common

between the two). Surprisingly, they identified temperature as having been the most probably driver of selection as opposed to, say, precipitation, which one might suspect as being important in a mediterranean climate. Another revealing example is provided by data from balsam poplar (*Populus balsamifera*). Keller et al. (2012) combined tests that identified  $F_{ST}$  outliers (Arlequin, EXCOFFIER *et al.* 2009; BayeScan, FOLL and GAGGIOTTI 2008) with tests for significant correlations between allele frequencies and environmental factors (COOP *et al.* 2010; GÜNTHER and COOP 2012). While their study identified 14 genes that showed signatures of local adaptation, only two showed statistical significance for both  $F_{ST}$  outliers and an association with one or more environmental variables (Figure 5). This sort of inconsistency can be revealing, because it brings into sharper focus model assumptions and points to the most effective methodology for determining loci responsible for adaptation to heterogeneous environments. Only a few such studies have been performed to date, but it is becoming clear that methods correlating allele frequencies with environmental variables are likely to be revealing (DE MITA *et al.* 2013; SCHOVILLE *et al.* 2012). Such methods can be made more robust to the residual effects of demography and population structure through the use of Latent Factor Mixed Models, which estimate and remove the effects of unknown hidden factors (FRICHOT *et al.* 2012).

In principle, genetic markers associated with environmental variables or  $F_{ST}$  outliers may simply be in linkage disequilibrium (LD) with a selected locus and not be the locus itself. Encouragingly, it has become clear that LD around selected loci decays relatively rapidly with distance along the chromosome (BARTON 1979;

BARTON and BENGTTSSON 1986), even for selfing species in which gametic disequilibrium decays more slowly (NORDBORG *et al.* 1996). Markers identified by the approaches reviewed here may thus often be very close to a selected locus, particularly if divergent selection among different habitats has been strong over long periods of time, and as long as an appropriate model of subdivision or demographic history has been assumed (because subdivision directly increases gametic disequilibrium) (BIERNE *et al.* 2011). This explains why it was possible to identify plausible genes under selection for local adaptation in the *Pinus* species cited above, as well as loci implicated in local adaptation to serpentine soils in *Arabidopsis lyrata* (TURNER *et al.* 2010; TURNER *et al.* 2008).

### **EFFECTS OF SUBDIVISION ON INBREEDING AND INBREEDING DEPRESSION**

Because plants tend to mate locally, population subdivision will tend to increase the rate of inbreeding. As a consequence of the corresponding increase in homozygosity, the increased expression of recessive deleterious alleles can lead to the short-term purging of inbreeding depression within populations (THRALL *et al.* 1998; WHITLOCK 2002). However, over the longer term, drift can overcome selection in small and isolated demes, fixing deleterious alleles locally. This reduces not only the mean fitness of the population, but also the level of inbreeding depression, because inbred and outbred individuals all express the same fixed deleterious recessive mutations. Variation in inbreeding depression in the plant *Mercurialis annua* are probably explained by such effects of drift in small populations, in this case following species range expansion (Figure 6). Because the Iberian Peninsula was recolonized by *M.*

*annua* from the south, northern populations are less genetically diverse (OBBARD *et al.* 2006) and show substantially lower inbreeding depression than southern ones (Figure 6). It is not known whether these populations have also fixed their genetic load of deleterious mutations, but genome-wide molecular data might allow such effects of drift to be detected, e.g., through the detection of fixed differences at non-synonymous sites.

The effects of drift on inbreeding depression and the fixation of genetic load in small isolated populations, predicted by a negative association between inbreeding depression and  $F_{ST}$  (Figure 7), can ultimately only be reversed by migration among demes. This process of ‘genetic rescue’, which has been documented for both plants (e.g., WILLI and FISCHER 2005; WILLI *et al.* 2005) and animals (e.g., EBERT *et al.* 2002; ESCOBAR *et al.* 2008; SACCHERI *et al.* 1998), has important implications both for our understanding of the genetic architecture of population subdivision, as well as for conservation. For instance, managers of threatened species have tended to advocate sourcing seeds for species introductions from populations as nearby and genetically similar as possible. A recent study by Pickup *et al.* (2013) indicates that this may often be misguided. These authors analyzed the fitness of individuals produced by artificial crosses within and between populations of the Australian perennial herb *Rutidosia leptorrhynchoides* and found that fitness was increased by crossing with individuals from large, genetically diverse populations that were not necessarily local (Figure 8). Importantly, the consequences of crosses will often vary among populations, and conservation



efforts should often be framed in a metapopulation context, bearing in mind differences in population age, size and history of migration.

In any finite population, the relatedness between pairs of mating individuals will vary: the higher the relatedness, measured by the pairwise inbreeding coefficient  $F$ , the more likely it is that progeny will be homozygous at a given locus (WRIGHT 1932). The associated increased expression of deleterious recessive alleles, or (probably less commonly) the expression of overdominant loci will then cause lower fitness in the progeny of parents with higher  $F$  (CHARLESWORTH and WILLIS 2009), i.e., we expect a positive relationship between the expression of inbreeding depression and  $F$  among individuals in a population. In the most classical sense, inbreeding depression is thus characterized by an inverse relationship between  $F$  and an individual's fitness, or of a pair's offspring. Inferences derived from  $F$  can be made directly (through molecular genotyping) or indirectly (using controlled crosses).

Although there is much evidence for inbreeding depression in natural populations, the expectation of an inverse relationship between fitness and  $F$  for a given parental pair has not been supported by a number of recent empirical experiments. The discrepancy might be attributed to processes associated with population structure. For example, in the plant *Ranunculus reptans*, inbred offspring may be equally fit or fitter relative to other individuals, and the relationship between  $F$  and fitness varies between populations (WILLI *et al.* 2005). Similarly, a 5-generation serial inbreeding experiment with *Mimulus guttatus* showed that the

relationship between total flower production and the degree of inbreeding varied significantly among populations and families (DUDASH *et al.* 1997). In this experiment, the higher extinction probability of inbred lines meant that the purging of genetic load could be accomplished more readily by selection among lines, rather than selection among individuals within lines.

The causes of variation in the  $F$  – fitness relationship among populations or families must include some variance in the distribution of recessive, or nearly recessive, mutations, brought about by variance in population age, demographic history, genetic drift, founder effect, historical gene flow, bi-parental inbreeding, and other past opportunities for the purging or fixation of deleterious recessive alleles. These and related processes can occur at very local scales, even within continuous populations, and may feed back to affect dynamics associated with population structure such as gene flow, population growth and persistence. Such variance in inbreeding effects therefore reflects the effects of population structure in the inbreeding process.

One study on the perennial rosemary scrub (*Hypericum cumulicola*) suggests that population size, age, and isolation will have a significant effect on determining the fitness consequences of certain crosses (Figure 9). In order to determine population characters known to affect the consequences of inbreeding, Oakley et al. (2012) combined molecular marker-based estimators of migration and estimates of relative effective size for several natural populations of *H. cumulicola*. In a study that resonates with that of Pickup et al. (2013) of *Rutidosis leptorrhynchoides*, cited above, fitness assays conducted on the products of hand pollinations of within- and

among-population crosses (including crosses with self-pollen, outcross pollen from a different individual within the same population, and outcross pollen from each of two different subpopulations), revealed that heterosis was significantly greater for small populations relative to large ones, and inbreeding depression tended to be smaller (Figure 9).

### **CURRENT TECHNOLOGIES – FROM GENOME SEQUENCING TO RAD-TAGS**

Advances in sequencing technologies are creating new opportunities for testing hypotheses using genomic information in both well-established models such as *Arabidopsis thaliana* and uncharacterized systems and at new scales of biological organization (HORTON *et al.* 2012; SLOAN *et al.* 2012). *Arabidopsis thaliana* started out as a model for the study of developmental and molecular genetics, but it has now been accepted as a model by a broader research community that includes ecologists and evolutionary biologists (Figure 10). This contrasts with the sneers one used to hear in bars at ecology and evolution conferences that ‘*Arabidopsis thaliana* is not a real plant’! Indeed, *Arabidopsis thaliana* and its relatives offer us a broad set of genetic tools for genomic analysis of population subdivision (GAUT 2012), as the increasingly large multi-population, partial or whole-genome scans for polymorphism illustrate (ARANZANA *et al.* 2005; ATWELL *et al.* 2010; BAKKER *et al.* 2006; CLARK *et al.* 2007; FOURNIER-LEVEL *et al.* 2011; HORTON *et al.* 2012; NORDBORG *et al.* 2005a; NORDBORG and WEIGEL 2008; PLATT *et al.* 2010b). Application of these tools in *A. thaliana* has revealed signatures of population demographic history, subdivision and selection that were hitherto not possible.

While early studies of population structure across the range of *A. thaliana* failed to find much evidence for isolation by distance, such patterns have now been revealed by analysis of whole-genome resequencing (NORDBORG *et al.* 2005b; PLATT *et al.* 2010a). High throughput analyses have also revealed a characteristic signature of reduced diversity at high latitudes indicative of bottlenecks associated with range expansion (LEWANDOWSKA-SABAT *et al.* 2010), as has been found in other species (HEWITT 2000). The reduction in genetic diversity during range expansion has implications both for the expression of inbreeding depression, as well as for populations' responsiveness to selection (PUJOL and PANNELL 2008). It is thus possible to interpret patterns of genetic diversity in *A. thaliana* in the context of what we can infer about the species demographic and phylogeographic history.

Increasingly detailed analysis of the structure of genomic variation over the geographic distribution of *A. thaliana* has revealed evidence for selection on a wide range of traits. Horton *et al.* (2012) recently genotyped a global sample of over a thousand individuals of *A. thaliana* using a 250,000 SNP chip. They used statistics that measure the haplotype structure across the genome and the allele frequency spectrum, allowing the distinction between past and ongoing selection and between selection on new mutations and standing genetic variation (MAYNARD SMITH and HAIGH 1974; NIELSEN *et al.* 2005; TOOMAJIAN *et al.* 2006). They found that  $F_{ST}$  was the only statistic that pointed to selection on defense-related regions of the genome, a pattern inconsistent with a model of repeated selective sweeps on defense genes and more consistent with long-term balancing selection at these loci (BAKKER *et al.* 2008).  $F_{ST}$  scans also revealed population differentiation in genomic regions

associated with flowering time, a trait expected to be under differential selection in different environments (HORTON *et al.* 2012). A recent selection study of *A. thaliana* using reciprocal transplant experiments across Europe (LOWRY 2012) has revealed direct evidence for local adaptation, particularly implicating freeze tolerance, and it can be anticipated that application of genomic tools to the genetic material produced by this study will point to where in the genome, and how, selection has acted over short periods of time.

Much of the progress made with *A. thaliana* as a model has been facilitated by the availability of a reference genome (KAUL *et al.* 2000), but for many plants that have particularly large and repetitive genomes, full genome sequencing is still not economically feasible. Nevertheless, many questions can now be addressed through the application of genotyping-by-sequencing approaches (GBS), which allow for a targeted fraction of the genome (or a reduced representation library) to be sequenced; these include the use of targeted restriction enzymes in order to reduce genome complexity, capture probes, or transcriptome-based analysis (DAVEY *et al.* 2011; NARUM *et al.* 2013). GBS allows the genetic analysis of species with little or no genomic information and with a full range of genome sizes (NARUM *et al.* 2013), and is rapidly being adopted to address questions across a range of taxa (see Molecular Ecology's special issue, Genotyping by Sequencing in Ecological and Conservation Genomics).

Currently the most rapidly advancing GBS approach is perhaps restriction-site-associated DNA sequencing (RAD-tag or RAD-seq, BAIRD *et al.* 2008; though also see ELSHIRE *et al.* 2011). The RAD-tag approach involves a genome-wide survey of

nucleotide diversity of regions flanking restriction sites, and allows the simultaneous detection and genotyping of thousands of genome-wide SNPs (WAGNER *et al.* 2013). High costs of multiplexing prevented genotyping of population or pooled samples for initial iterations of the method, but emerging pipelines such as double-digest RADseq (ddRADseq) now allow cheaper polymorphism discovery and genotyping for large samples by multiplexing digested samples (PETERSON *et al.* 2012). Currently, ddRADseq offers the most feasible approach for generating genomic data necessary for inferences about population structure, especially when its consequences (such as local adaptation) are not extreme, though the possibility of allele dropout can compromise its potential (GAUTIER *et al.* 2013). Pipelines for the analysis of data derived from RADseq are beginning to be published (e.g. STACKS, CATCHEN *et al.* 2011; UNEAK, LU *et al.* 2013). With these sequencing and analytic tools, a combination of ddRADseq and genetic-environment correlations offer an effective framework for future studies of the consequences of population subdivision in plants.

### **WHITHER NOW - NEW WINE IN OLD SKINS?**

The study of the evolution of local adaptation in plants has a long history, going back to investigations that have become classics at a wide range of spatial scales (e.g., reviewed in KAWECKI and EBERT 2004). Most of this work focused on phenotypes, but current technologies are opening up new possibilities for working on the genetics of local adaptation for species with larger and more difficult genomes than *A. thaliana*, including polyploids. These studies will include systems that have not been

investigated before, but much might also be gained from using new approaches to rake over the coals of earlier work, for which fundamentals have already been worked out. The evolution of heavy-metal tolerance on contaminated mine tailings in the grass *Anthoxanthum odoratum* (e.g., ANTONOVICS *et al.* 1971; BRANDON 1990) provides a good example a system that might now benefit from revisiting, but there are of course many others. The research program on *A. odoratum* is exemplary, because it represents a case study in which several key milestones have been achieved in understanding the genetic basis of adaptation and speciation. These are loosely enumerated in Table 1, following Brandon (1990). Table 1 also summarizes the details of how these milestones were reached in the study of *A. odoratum* specifically.

Early reciprocal transplant experiments with *A. odoratum* revealed strong selection acting between metal-tolerant and metal-sensitive genotypes under field conditions, and manipulation of the soil environment demonstrated that this was due specifically to the effects of the heavy metals, i.e., points 1 and 2 in Table 1 were well covered by ecological experimentation. It is in points 3 to 5 that classical methods have only provided partial understanding of the details of the evolution of heavy-metal tolerance in this species. To some extent, variation in tolerance among individuals and populations has been attributed to gene flow among populations on different soils, but much remains to be learnt about which genes and genome segments are involved in the adaptation, and about their fate in contaminated and uncontaminated environments. Although the widespread distribution of non-tolerant phenotypes suggest that tolerance is a derived trait (ANTONOVICS 1966;

ANTONOVICS *et al.* 1971; BRANDON 1990), direct phylogenetic evidence for the polarity of adaptation is still lacking, and it is not known whether tolerance evolved independently in some populations or whether it evolved once and was then successfully exported to other populations.

As with many plant systems of ecological and evolutionary importance, there are still only limited molecular marker resources for *A. odoratum*, which include organellar gene sequences and genome wide AFLPs (FREELAND *et al.* 2010; FREELAND *et al.* 2012). The application of ddRADseq could now help advance our understanding of this classic example of local adaptation. For instance, signals of selection could be exposed by genotyping individuals across environmental gradients of soil types using ddRADseq or other *de novo* genotyping approaches, combined with analysis of outlier and genetic-environment associations and genotyping of mapping populations produced by crosses.

One could envisage follow-up reciprocal transplant experiments among soil types that targeted SNP polymorphisms. Candidate gene analysis might then lead to a deeper knowledge of the functional, metabolic, and regulatory roles of genes responsible for metal tolerance. Because full tolerance to heavy metals is not observed in all populations of *A. odoratum* at contaminated sites, there has likely been a complex history of selection and migration acting on the species; such complexities could be understood using a combination of molecular population genetic methods to estimate long-term (MIGRATE, BEERLI 2009; TreeMix, PICKRELL and PRITCHARD 2012) and short-term gene-flow (BIMr, FAUBET and GAGGIOTTI 2008; BayesAss, WILSON and RANNALA 2003).



The example of *A. odoratum* illustrates the sort of ecological and evolutionary questions likely to benefit from the application of GBS methods. Some time ago, David (1998) outlined an ambitious plan to better understand the presence/absence of heterozygosity-fitness-correlations (HFC) in natural populations. As described above, HFCs are likely due to the masking of recessive deleterious alleles in heterozygotes (LYNCH and WALSH 1998). However, direct determination of the genomic regions responsible for HFCs, and inbreeding depression more generally, remains elusive. Application of GBS to experimental populations using new genomic tools will likely allow us to detect the loci responsible for inbreeding and outbreeding depression, and to determine the distribution of the relevant alleles among populations. Ward et al. (2013) recently applied GBS methodologies, as well as a novel genome-independent imputation pipeline to deal with missing or erroneous data, in order to generate a linkage map of the plant species complex *Rubus idaeus* (which includes red raspberry and blackberry). High-density genotyping of the progeny of controlled crosses allowed the identification of genomic regions exhibiting segregation distortion that might be responsible for inbreeding depression. Importantly, these insights were gained without a reference genome, in a fraction of the time and costing much less than analyses relying on more traditional sequencing approaches.

## CONCLUSIONS

The exhilarating progress that is currently being made in the generation and analysis of genomic data poses both opportunities and serious challenges for our understanding of the evolution of plants across heterogeneous habitats. It is worth

stepping back to consider what can really be gained from these new tools and data beyond descriptions of diversity and its distribution, and what might be lost. Casting our eyes back over the several past decades of ecological genetic work based on earlier genetic markers is inspiring, but it also provides salutary lessons.

The advent of molecular polymorphic markers such as isozymes/allozymes, microsatellites, AFLPs, etc., opened up ways to understand genetic diversity and to use it for testing ecological and evolutionary hypotheses. There are many examples of what could be done. But the availability of these markers also prompted a great deal of relatively uninspiring work that simply described differentiation among populations in terms of  $F_{ST}$  and other statistics, often interpreting patterns in terms of inappropriate evolutionary models. Much of this work is still valuable for large-scale comparative analyses and meta-analysis, but one is left with the feeling that so much more might have been achieved than just describing patterns of differentiation and misinterpreting them in terms of gene flow.

The opportunities provided by the new tools and data are alluring, but there are also substantial technical challenges. As algorithms and pipelines become available to deal with the data, it will be important to remember that the processes that gave rise to observed patterns were potentially complex and may be inadequately interpreted using simple models. Importantly, it can be hoped that we will be able to go beyond mere description of patterns and to use the new tools in creative ways to test ecological and evolutionary hypotheses, in both comparative and experimental settings. Here, knowledge of evolutionary theory and population genetics will need to form the basis of sampling strategies. A sound understanding

of the genealogical structure of populations, the locus-dependency of gene flow and effective population sizes, and the time scales over which statistical associations between loci break down through migration and recombination will be as important as ever.

In this review, we have briefly discussed controversy over the utility of  $F_{ST}$  as a measure of differentiation, but we highlighted its advantages as a statistic with a sound grounding in theoretical population genetics, notwithstanding several caveats. It is nonetheless important to bear in mind that  $F_{ST}$  at any one locus has a large evolutionary variance, so that loci under weak selection will tend to escape detection by genome scans. Given the likely widespread importance of polygenic traits in local adaptation, this is a potentially serious limitation to the utility of  $F_{ST}$ . It is therefore exciting to note recent theoretical developments that reveal high power to detect selection on polygenic traits in terms of co-ordinated shifts in allele frequencies associated with habitat variation, which would be completely undetectable at individual loci (BERG and COOP 2013; TURCHIN *et al.* 2012). We can now look forward to the application of these and perhaps other approaches for understanding the evolution of subdivided plant populations.

**ACKNOWLEDGMENTS** We thank Janis Antonovics, Kent Holsinger, Stephen Keller, Santiago Martinez-Gonzalez, Lou Jost, Douglas Taylor (and all of the Taylor lab), Severine Vuilleumier and an anonymous reviewer for their helpful comments on

the manuscript. The opinions expressed are, however, our own. JP is funded by the Swiss National Science Foundation and the University of Lausanne. PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT and JA, and NSF-OISE# 1139716 to DT and PF.

## LITERATURE CITED

- ALCALA, N., D. STREIT, J. GOUDET and S. VUILLEUMIER, 2013 Peak and persistent excess of genetic diversity following an abrupt migration increase. *Genetics* (**in press**).
- ANTONIAZZA, S., R. BURRI, L. FUMAGALLI, J. GOUDET and A. ROULIN, 2010 Local adaptation maintains clinal variation in melanin-based coloration of European barn owls (*Tyto alba*). *Evolution* **64**: 1944-1954.
- ANTONOVICS, J., 1966 The genetics and evolution of differences between closely adjacent plant populations with special reference to heavy metal tolerance., pp. University of Wales.
- ANTONOVICS, J., and A. D. BRADSHAW, 1970 Evolution in closely adjacent plant populations VIII. Clinal patterns at a mine boundary. *Heredity* **25**: 349-362.
- ANTONOVICS, J., A. D. BRADSHAW and R. G. TURNER, 1971 Heavy metal tolerance in plants. *Advances in Ecological Research* **7**: 1-85.
- ARANZANA, M. J., S. KIM, K. ZHAO, E. BAKKER, M. HORTON *et al.*, 2005 Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. *PLoS Genet* **1**: e60.
- ATWELL, S., Y. S. HUANG, B. J. VILHJALMSSON, G. WILLEMS, M. HORTON *et al.*, 2010 Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627-631.
- AUDIGEOS, D., L. BROUSSEAU, S. TRAISSAC, C. SCOTTI-SAINTAGNE and I. SCOTTI, 2013 Molecular divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary Biology* **26**: 529-544.

- BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER *et al.*, 2008 Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* **3**.
- BAKKER, E. G., E. A. STAHL, C. TOOMAJIAN, M. NORDBORG, M. KREITMAN *et al.*, 2006 Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Molecular Ecology* **15**: 1405-1418.
- BAKKER, E. G., M. B. TRAW, C. TOOMAJIAN, M. KREITMAN and J. BERGELSON, 2008 Low levels of polymorphism in genes that control the activation of defense response in *Arabidopsis thaliana*. *Genetics* **178**: 2031-2043.
- BALDING, D. J., and R. A. NICHOLS, 1995 A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**: 3-12.
- BARTON, N., 1979 Gene flow past a cline. *Heredity* **43**: 333-359.
- BARTON, N., and B. O. BENGTSSON, 1986 The barrier to genetic exchange between hybridizing populations. *Heredity* **57**: 357-376.
- BARTON, N. H., and M. TURELLI, 1989 Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* **23**: 337-370.
- BEAUMONT, M. A., and R. A. NICHOLS, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**: 1619-1626.
- BEERLI, P., 2009 How to use migrate or why are markov chain monte carlo programs difficult to use?, pp. 42-79 in *Population Genetics for Animal Conservation*,

- edited by G. BERTORELLE, M. W. BRUFORD, H. C. HAUFFE, A. RIZZOLI and C. VERNESI.  
Cambridge University Press, Cambridge UK.
- BERG, J. J., and G. COOP, 2013 The Population Genetic Signature of Polygenic Local Adaptation. arXiv:1307.7759.
- BHATIA, G., N. J. PATTERSON, S. SANKARARAMAN and A. L. PRICE, 2013 Estimating and interpreting  $F_{st}$ : the impact of rare variants. *Genome Research*.
- BIERNE, N., J. WELCH, E. LOIRE, F. BONHOMME and P. DAVID, 2011 The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044-2072.
- BRANDON, R. N., 1990 Mechanism and Teleology, pp. 159-194 in *Adaptation and Environment*. Princeton University Press, Princeton, NJ.
- CAO, J., K. SCHNEEBERGER, S. OSSOWSKI, T. GUNTHER, S. BENDER *et al.*, 2011 Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* **43**: 956-U960.
- CATCHEN, J. M., A. AMORES, P. HOHENLOHE, W. CRESKO and J. H. POSTLETHWAIT, 2011 Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* **1**: 171-182.
- CHAPUIS, E., G. MARTIN and J. GOUDET, 2008 Effects of Selection and Drift on G Matrix Evolution in a Heterogeneous Environment: A Multivariate  $Q_{st}$ - $F_{st}$  Test With the Freshwater Snail *Galba truncatula*, pp. 2151-2161 in *Genetics*.
- CHARLESWORTH, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**: 538-543.

- CHARLESWORTH, D., and J. H. WILLIS, 2009 The genetics of inbreeding depression. *Nat Rev Genet* **10**: 783-796.
- CLARK, R. M., G. SCHWEIKERT, C. TOOMAJIAN, S. OSSOWSKI, G. ZELLER *et al.*, 2007 Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*. *Science* **317**: 338-342.
- COOP, G., D. WITONSKY, A. DI RIENZO and J. K. PRITCHARD, 2010 Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**: 1411-1423.
- CRAWFORD, N. G., 2010 smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**: 556-557.
- DAVEY, J. W., P. A. HOHENLOHE, P. D. ETTER, J. Q. BOONE, J. M. CATCHEN *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**: 499-510.
- DAVID, P., 1998 Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**: 531-537.
- DE CAUWER, I., M. DUFAY, J. CUGUEN and J.-F. ARNAUD, 2010 Effects of fine-scale genetic structure on male mating success in gynodioecious *Beta vulgaris* ssp. *maritima*, pp. 1540-1558 in *Molecular Ecology*.
- DE MITA, S., A. C. THUILLET, L. GAY, N. AHMADI, S. MANEL *et al.*, 2013 Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**: 1383-1399.



- DELPH, L. F., J. C. STEVEN, I. A. ANDERSON, C. R. HERLIHY and E. D. BRODIE, III, 2011  
Elimination of a genetic correlation between the sexes via artificial  
correlational selection. *Evolution* **65**: 2872-2880.
- DUDASH, M., D. CARR and C. FENSTER, 1997 Five generations of enforced selfing and  
outcrossing in *Mimulus guttatus*: Inbreeding depression variation at the  
population and family level. *Evolution* **51**: 54-65.
- EBERT, D., C. HAAG, M. KIRKPATRICK, M. RIEK, J. W. HOTTINGER *et al.*, 2002 A selective  
advantage to immigrant genes in a *Daphnia* metapopulation. *Science* **295**:  
485-488.
- ECKERT, A. J., J. VAN HEERWAARDEN, J. L. WEGRZYN, C. D. NELSON, J. ROSS-IBARRA *et al.*, 2010  
Patterns of population structure and environmental associations to aridity  
across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**:  
969-982.
- EDELAAR, P., P. BURRACO and I. GOMEZ-MESTRE, 2011 Comparisons between Q(ST) and  
F(ST): how wrong have we been? *Molecular Ecology* **20**: 4830-4839.
- ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO *et al.*, 2011 A Robust,  
Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity  
Species. *Plos One* **6**: e19379.
- ESCOBAR, J. S., A. NICOT and P. DAVID, 2008 The different sources of variation in  
inbreeding depression, heterosis and outbreeding depression in a  
metapopulation of *Physa acuta*. *Genetics* **180**: 1593-1608.
- ETTERSON, J. R., and R. G. SHAW, 2001 Constraint to adaptive evolution in response to  
global warming. *Science* **294**: 151-154.

- EXCOFFIER, L., T. HOFER and M. FOLL, 2009 Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285-298.
- FAUBET, P., and O. E. GAGGIOTTI, 2008 A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* **178**: 1491-1504.
- FOLL, M., and O. E. GAGGIOTTI, 2006 Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875-891.
- FOLL, M., and O. E. GAGGIOTTI, 2008 A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**: 977-993.
- FOURNIER-LEVEL, A., A. KORTE, M. D. COOPER, M. NORDBORG, J. SCHMITT *et al.*, 2011 A Map of Local Adaptation in *Arabidopsis thaliana*. *Science* **334**: 86-89.
- FREELAND, J. R., P. BISS, K. F. CONRAD and J. SILVERTOWN, 2010 Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *Journal of Evolutionary Biology* **23**: 776-782.
- FREELAND, J. R., P. BISS and J. SILVERTOWN, 2012 Contrasting Patterns of Pollen and Seed Flow Influence the Spatial Genetic Structure of Sweet Vernal Grass (*Anthoxanthum odoratum*) Populations. *Journal of Heredity* **103**: 28-35.
- FRICHOT, E., S. SCHOVILLE, G. BOUCHARD and O. FRANÇOIS, 2012 Landscape genomic tests for associations between loci and environmental gradients. arXiv:1205.3347.
- GAGGIOTTI, O. E., and M. FOLL, 2010 Quantifying population structure using the F-model. *Molecular Ecology Resources* **10**: 821-830.

- GARTSIDE, D. W., and T. MCNEILLY, 1974 Genetic studies in heavy metal tolerant plants. *Heredity* **32**: 287-297.
- GAUT, B., 2012 *Arabidopsis thaliana* as a model for the genetics of local adaptation. *Nat Genet* **44**: 115-116.
- GAUTIER, M., K. GHARBI, T. CEZARD, J. FOUCAUD, C. KERDELHUE *et al.*, 2013 The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**: 3165-3178.
- GILPIN, M., 1991 The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**: 165-176.
- GRIVET, D., F. SEBASTIANI, R. ALIA, T. BATAILLON, S. TORRE *et al.*, 2011 Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution* **28**: 101-116.
- GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LEGER *et al.*, 2011 Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591-611.
- GÜNTHER, T., and G. COOP, 2012 Robust identification of local adaptation from allele frequencies. [arXiv.org](https://arxiv.org/).
- GUO, F., D. K. DEY and K. E. HOLSINGER, 2009 A Bayesian hierarchical model for analysis of single-nucleotide polymorphisms diversity in multilocus, multipopulation samples. *Journal of the American Statistical Association* **104**: 142-154.
- HARTL, D. L., and A. G. CLARK, 1997 *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts.

- HEDRICK, P., 2005a A standardized genetic differentiation measure. *Evolution* **59**: 1633-1638.
- HEDRICK, P. W., 2005b A standardized genetic differentiation measure. *Evolution* **59**: 1633-1638.
- HEDRICK, P. W., M. E. GINEVAN and E. P. EWING, 1976 Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics* **7**: 1-32.
- HEIN, J., M. SCHIERUP and C. WIUF, 2005 *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, Oxford.
- HEWITT, G., 2000 The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.
- HORTON, M. W., A. M. HANCOCK, Y. S. HUANG, C. TOOMAJIAN, S. ATWELL *et al.*, 2012 Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* **44**: 212-216.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process, pp. 1-44 in *Oxford Surveys in Evolutionary Biology*, edited by D. J. FUTUYMA and J. ANTONOVICS. Oxford University Press, Oxford.
- JAKOBSSON, M., M. D. EDGE and N. A. ROSENBERG, 2013 The Relationship Between  $F_{ST}$  and the Frequency of the Most Frequent Allele. *Genetics* **193**: 515-528.
- JAY, F., S. MANEL, N. ALVAREZ, E. Y. DURAND, W. THUILLER *et al.*, 2012 Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology* **21**: 2354-2368.

- JOST, L., 2008a G(ST) and its relatives do not measure differentiation. *Molecular Ecology* **17**: 4015-4026.
- JOST, L. O. U., 2008b GST and its relatives do not measure differentiation. *Molecular Ecology* **17**: 4015-4026.
- KARHUNEN, M., and O. OVASKAINEN, 2012 Estimating population-level coancestry coefficients by an admixture F model. *Genetics* **192**: 609-617.
- KAUL, S., H. L. KOO, J. JENKINS, M. RIZZO, T. ROONEY *et al.*, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.
- KAWECKI, T. J., and D. EBERT, 2004 Conceptual issues in local adaptation. *Ecology Letters* **7**: 1225-1241.
- KELLER, S. R., N. LEVSEN, M. S. OLSON and P. TIFFIN, 2012 Local Adaptation in the Flowering-Time Gene Network of Balsam Poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* **29**: 3143-3152.
- LASCOUX, M., and R. J. PETIT, 2010 The 'New Wave' in plant demographic inference: more loci and more individuals. *Molecular Ecology* **19**: 1075-1078.
- LEWANDOWSKA-SABAT, A. M., S. FJELLHEIM and O. A. ROGNLI, 2010 Extremely low genetic variability and highly structured local populations of *Arabidopsis thaliana* at higher latitudes. *Molecular Ecology* **19**: 4753-4764.
- LEWONTIN, R., 1980 Theoretical population genetics in the evolutionary synthesis, pp. 58-68 in *The Evolutionary Synthesis: Perspectives on the Unification of Biology*, edited by E. MAYR and W. B. PROVINE. Harvard University Press, Cambridge, Massachusetts.

- LEWONTIN, R. C., and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics* **74**: 175-195.
- LOWRY, D. B., 2012 Local adaptation in The model plant. *New Phytologist* **194**: 888-890.
- LU, F., A. E. LIPKA, J. GLAUBITZ, R. ELSHIRE, J. H. CHERNEY *et al.*, 2013 Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genet* **9**: e1003215.
- LUIKART, G., P. R. ENGLAND, D. TALLMON, S. JORDAN and P. TABERLET, 2003 The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**: 981-994.
- LYNCH, M., and J. B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Assocs., Inc., Sunderland, MA.
- MARTIN, G., E. CHAPUIS and J. GOUDET, 2008 Multivariate Qst-Fst Comparisons: A Neutrality Test for the Evolution of the G Matrix in Structured Populations, pp. 2135-2149 in *Genetics*.
- MARUYAMA, T., and M. KIMURA, 1980 Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Science, USA* **77**: 6710-6714.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favourable gene. *Genetical Research, Cambridge* **219**: 23-35.
- MCCAULEY, D. E., 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for

- studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.
- MCCAULEY, D. E., 1998 The Genetic Structure of a Gynodioecious Plant: Nuclear and Cytoplasmic Genes. *Evolution* **52**: 255-260.
- MEIRMANS, P. G., 2012 AMOVA-Based Clustering of Population Genetic Data. *Journal of Heredity*.
- MEIRMANS, P. G., and P. W. HEDRICK, 2011 Assessing population structure: F-ST and related measures. *Molecular Ecology Resources* **11**: 5-18.
- NARUM, S. R., C. A. BUERKLE, J. W. DAVEY, M. R. MILLER and P. A. HOHENLOHE, 2013 Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**: 2841-2847.
- NEI, M., 1973 Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science, USA* **70**: 3321-3323.
- NEIGEL, J. E., 1997 A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics* **28**: 105-128.
- NICHOLSON, G., A. SMITH, F. JÓNSSON, Ó. GÚSTAFSSON, K. STEFÁNSSON *et al.*, 2002 Assessing Population Differentiation and Isolation from Single-Nucleotide Polymorphism Data, pp. 695-715 in *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*.

- NIELSEN, R., S. WILLIAMSON, Y. KIM, M. J. HUBISZ, A. G. CLARK *et al.*, 2005 Genomic scans for selective sweeps using SNP data. *Genome Research* **15**: 1566-1575.
- NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proceedings of the Royal Society of London, B* **163**: 1033-1039.
- NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005a The Pattern of Polymorphism in *Arabidopsis thaliana*. *Plos Biol* **3**: e196.
- NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005b The pattern of polymorphism in *Arabidopsis thaliana*. *Plos Biology* **3**: 1289-1299.
- NORDBORG, M., and D. WEIGEL, 2008 Next-generation genetics in plants. *Nature* **456**: 720-723.
- OAKLEY, C. G., and A. A. WINN, 2012 Effects of population size and isolation on heterosis, mean fitness, and inbreeding depression in a perennial plant. *New Phytologist* **196**: 261-270.
- OBARD, D. J., S. A. HARRIS and J. R. PANNELL, 2006 Sexual systems and population genetic structure in an annual plant: testing the metapopulation model. *American Naturalist* **167**: 354-366.
- OETJEN, K., and T. B. H. REUSCH, 2007 Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology* **16**: 5156-5167.
- OVASKAINEN, O., M. KARHUNEN, C. ZHENG, J. M. C. ARIAS and J. MERILA, 2011 A New Method to Uncover Signatures of Divergent and Stabilizing Selection in Quantitative Traits. *Genetics* **189**: 621-U729.



- PANNELL, J. R., 2003 Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* **57**: 949-961.
- PANNELL, J. R., and B. CHARLESWORTH, 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664-676.
- PANNELL, J. R., and B. CHARLESWORTH, 2000a Effects of Metapopulation Processes on Measures of Genetic Diversity. *Philosophical Transactions: Biological Sciences* **355**: 1851-1864.
- PANNELL, J. R., and B. CHARLESWORTH, 2000b Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **355**: 1851-1864.
- PETERSON, B. K., J. N. WEBER, E. H. KAY, H. S. FISHER and H. E. HOEKSTRA, 2012 Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *Plos One* **7**: e37135.
- PETIT, R. J., R. BIALOZYT, S. BREWER, R. CHEDDADI and B. COMPS, 1997 From spatial patterns of genetic diversity to postglacial migration processes in forest trees, pp. 295-318 in *Plant life histories : ecology, phylogeny, and evolution*, edited by J. W. SILVERTOWN, M. FRANCO and J. L. HARPER. Cambridge University Press, Cambridge, U.K.
- PICKRELL, J. K., and J. K. PRITCHARD, 2012 Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* **8**: e1002967.

- PICKUP, M., D. L. FIELD, D. M. ROWELL and A. G. YOUNG, 2013 Source population characteristics affect heterosis following genetic rescue of fragmented plant populations. *Proceedings of the Royal Society B-Biological Sciences* **280**.
- PLATT, A., M. HORTON, Y. S. HUANG, Y. LI, A. E. ANASTASIO *et al.*, 2010a The Scale of Population Structure in *Arabidopsis thaliana*. *Plos Genetics* **6**.
- PLATT, A., M. HORTON, Y. S. HUANG, Y. LI, A. E. ANASTASIO *et al.*, 2010b The Scale of Population Structure in *Arabidopsis thaliana*. *PLoS Genet* **6**: e1000843.
- PRUNIER, J., S. GERARDI, J. LAROCHE, J. BEAULIEU and J. BOUSQUET, 2012 Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology* **21**: 4270-4286.
- PUJOL, B., and J. R. PANNELL, 2008 Reduced responses to selection after species range expansion. *Science* **321**: 96.
- PUJOL, B., A. J. WILSON, R. I. C. ROSS and J. R. PANNELL, 2008 Are  $Q_{ST}$  -  $F_{ST}$  comparisons for natural populations meaningful? *Molecular Ecology* **17**: 4782-4785.
- PUJOL, B., S. R. ZHOU, J. SAHCHÉZ-VILAS and J. R. PANNELL, 2009 Reduced inbreeding depression after species range expansion. *Proceeding of the National Academy of Sciences of the United States of America* **106**: 15379-15383.
- R DEVELOPMENT CORE TEAM, 2011 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.
- RENAUT, S., C. J. GRASSA, S. YEAMAN, B. T. MOYERS, Z. LAI *et al.*, 2013 Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications* **4**: 1827.

- ROUSSET, F., 2013 Exegeses on maximum genetic differentiation. *Genetics* **194**: 557-559.
- RYMAN, N., and O. LEIMAR, 2009 G(ST) is still a useful measure of genetic differentiation - a comment on Jost's D. *Molecular Ecology* **18**: 2084-2087.
- SACCHERI, I., M. KUUSSAARI, M. KANKARE, P. VIKMAN, W. FORTÉLIUS *et al.*, 1998 Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491-494.
- SCHOVILLE, S. D., A. BONIN, O. FRANCOIS, S. LOBREAUX, C. MELODELIMA *et al.*, 2012 Adaptive genetic variation on the landscape: methods and cases, pp. 23-43 in *Annual Review of Ecology, Evolution, and Systematics, Vol 43*, edited by D. J. FUTUYMA. Annual Reviews, Palo Alto.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinction. *Theoretical Population Biology* **12**: 253-262.
- SLATKIN, M., 1985 Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**: 393-430.
- SLATKIN, M., 1987 Gene flow and the geographic structure of natural populations. *Science* **236**: 787-792.
- SLATKIN, M., 1991 Inbreeding coefficients and coalescent times. *Genetical Research* **58**: 167-175.
- SLOAN, D. B., S. R. KELLER, A. E. BERARDI, B. J. SANDERSON, J. F. KARPOVICH *et al.*, 2012 De novo transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **12**: 333-343.

- SPITZE, K., 1993 Population-structure in *Daphnia obtusa* - quantitative genetic and allozymic variation. *Genetics* **135**: 367-374.
- STÄDLER, T., B. HAUBOLD, C. MERINO, W. STEPHAN and P. PFAFFELHUBER, 2009 The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations. *Genetics* **182**: 205-216.
- STEVEN, J. C., L. F. DELPH and E. D. BRODIE, III, 2007 Sexual dimorphism in the quantitative-genetic architecture of floral, leaf, and allocation traits in *Silene latifolia*. *Evolution* **61**: 42-57.
- STRIMMER, K., and O. G. PYBUS, 2001 Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* **18**: 2298-2305.
- THORNTON, K. R., and J. D. JENSEN, 2007 Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* **175**: 737-750.
- THRALL, P., C. RICHARDS, D. E. MCCAULEY and J. ANTONOVICS, 1998 Metapopulation collapse: the consequences of limited gene flow in spatially structured populations, pp. 83-104 in *Modeling spatiotemporal dynamics in ecology*, edited by J. BASCOMPTE and R. V. SOLE. Springer-Verlag Berlin.
- TOOMAJIAN, C., T. T. HU, M. J. ARANZANA, C. LISTER, C. L. TANG *et al.*, 2006 A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *Plos Biology* **4**: 732-738.
- TURCHIN, M. C., C. W. K. CHIANG, C. D. PALMER, S. SANKARARAMAN, D. REICH *et al.*, 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**: 1015-1019.

- TURNER, T. L., E. C. BOURNE, E. J. VON WETTBERG, T. T. HU and S. V. NUZH DIN, 2010  
Population resequencing reveals local adaptation of *Arabidopsis lyrata* to  
serpentine soils. *Nature Genetics* **42**: 260-263.
- TURNER, T. L., E. J. VON WETTBERG and S. V. NUZH DIN, 2008 Genomic Analysis of  
Differentiation between Soil Types Reveals Candidate Genes for Local  
Adaptation in *Arabidopsis lyrata*. *Plos One* **3**: e3183.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on  
the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WAGNER, C. E., I. KELLER, S. WITTEW, O. M. SELZ, S. MWAIKO *et al.*, 2013 Genome-wide  
RAD sequence data provide unprecedented resolution of species boundaries  
and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular  
Ecology* **22**: 787-798.
- WAKELEY, J., 2000 The effects of subdivision on the genetic divergence of populations  
and species. *Evolution* **54**: 1092-1101.
- WAKELEY, J., 2001 The coalescent in an island model of population subdivision with  
variation among demes. *Theoretical Population Biology* **59**: 133-144.
- WAKELEY, J., 2009 *Coalescent Theory: An Introduction*. Roberts and Company  
Publishers, Green Village, Colorado.
- WAKELEY, J., and N. ALIACAR, 2001 Gene genealogies in a metapopulation. *Genetics*  
**159**: 893-905.
- WANG, J., 2012 On the measurements of genetic differentiation among populations.  
*Genetics research* **94**: 275-289.

- WANG, Y., and J. HEY, 2010 Estimating divergence parameters with small samples from a large number of loci. *Genetics* **184**: 363-U390.
- WARD, J., J. BHANGOO, F. FERNANDEZ-FERNANDEZ, P. MOORE, J. SWANSON *et al.*, 2013 Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* **14**: 2.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.
- WHITLOCK, M., 2011a  $G'st$  and  $D$  do not replace  $F_{st}$ . *Molecular Ecology* **20**: 1083-1091.
- WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191-1202.
- WHITLOCK, M. C., 2011b  $G'(ST)$  and  $D$  do not replace  $F_{ST}$ . *Molecular Ecology* **20**: 1083-1091.
- WHITLOCK, M. C., and N. H. BARTON, 1997 The effective size of a subdivided population. *Genetics* **146**: 427-441.
- WHITLOCK, M. C., and F. GUILLAUME, 2009 Testing for spatially divergent selection: comparing  $Q(ST)$  to  $F(ST)$ . *Genetics* **183**: 1055-1063.
- WHITLOCK, M. C., and D. E. McCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.
- WHITLOCK, M. C., and D. E. McCAULEY, 1999 Indirect measures of gene flow and migration:  $F_{ST}$  not equal  $1/(4Nm+1)$ . *Heredity* **82**: 117-125.

- WILLI, Y., and M. FISCHER, 2005 Genetic rescue in interconnected populations of small and large size of the self-incompatible *Ranunculus reptans*. *Heredity* **95**: 437-443.
- WILLI, Y., J. VAN BUSKIRK and M. FISCHER, 2005 A threefold genetic allee effect: population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. *Genetics* **169**: 2255-2265.
- WILSON, G. A., and B. RANNALA, 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177-1191.
- WINTER, D. J., 2012 MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* **12**: 1158-1160.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159.
- WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics* **1**: 356-366.
- WRIGHT, S., 1943 Isolation by distance. *Genetics* **28**: 114-138.
- WRIGHT, S. I., and B. S. GAUT, 2005 Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* **22**: 1158-1158.
- YU, Q., E. D. ELLEN, M. J. WADE and L. F. DELPH, 2011 Genetic differences among populations in sexual dimorphism: evidence for selection on males in a dioecious plant. *Journal of Evolutionary Biology* **24**: 1120-1127.
- Table 1. The key milestones that ideally need to be reached to understand the evolution of local adaptation by selection in a heterogeneous environment, as listed

by Brandon (1990), and their realization in the classic model system for evolution of heavy-metal tolerance in *Anthoxanthum odoratum*.

Milestones in general	Milestone in <i>A. odoratum</i>	References
1. Measurement of differential fitness of variable traits under the environmental conditions of interest	Evidence gathered by conducting reciprocal transplants of individuals sampled in locations known to either be contaminated by heavy metals or not. Selection coefficients between different soil types were as high as 0.7.	(ANTONOVICS 1966; ANTONOVICS <i>et al.</i> 1971)
2. Providing an ecological/physiological explanation of the selection and verifying its operation in nature	Measurement and manipulation of the metal content of soils and determining its effect on root growth.	(ANTONOVICS <i>et al.</i> 1971; BRANDON 1990)
3. Documenting the inheritance of the selected trait	Targeted crosses among individuals sampled from different soils suggested that metal tolerance was inherited with partial dominance, without maternal effects, and that the trait is likely to be polygenic.	(GARTSIDE and MCNEILLY 1974)
4. Characterization of patterns of gene flow among populations and spatiotemporal variation in selection	The description of clinal variation in a number of morphological characters along a fine-scale transect between soil types that limit gene flow (thus isolating populations on different soils) .	(ANTONOVICS and BRADSHAW 1970)
5. Inferring trait polarity on the basis of phylogenetic information to show that it is indeed derived in the habitat	Discovery of tolerant individuals in multiple species growing on contaminated soils that were likely derived from non-tolerant populations. Selection experiments revealed	(ANTONOVICS <i>et al.</i> 1971; BRANDON 1990)



of interest

rapid evolution of tolerance, but the origin of tolerance genes was not clear.

Figure 1. The coalescent for a subdivided population. (a) The diagram depicts the genealogy of lineages sampled from three different demes. In demes 1 and 3, all locally sampled lineages coalesce with each other at a rate given by the local inbreeding effective size of the deme. The simultaneous coalescence of multiple lineages in deme three would suggest a local population bottleneck caused, for example, by a colonisation event. In deme 2, three of the four lineages coalesce locally, but the fourth lineage migrated into the deme from elsewhere. The waiting time to the final coalescent event with this lineage is determined by the eigenvalue effective size of the whole metapopulation, in the collecting phase of the structured coalescent. (b) A 'skyline plot' of simulated coalescent events for a subdivided population, showing: first, the scattering phase, where coalescence occurs at a rate governed by the local inbreeding effective sizes of the demes from which more than one lineage was sampled; and second, the collecting phase, where coalescent events occur at a rate determined by the metapopulation effective size, which is strongly influenced by the migration rate. Under strong migration, the two phases become one. Graph modified from Pannell (2003).

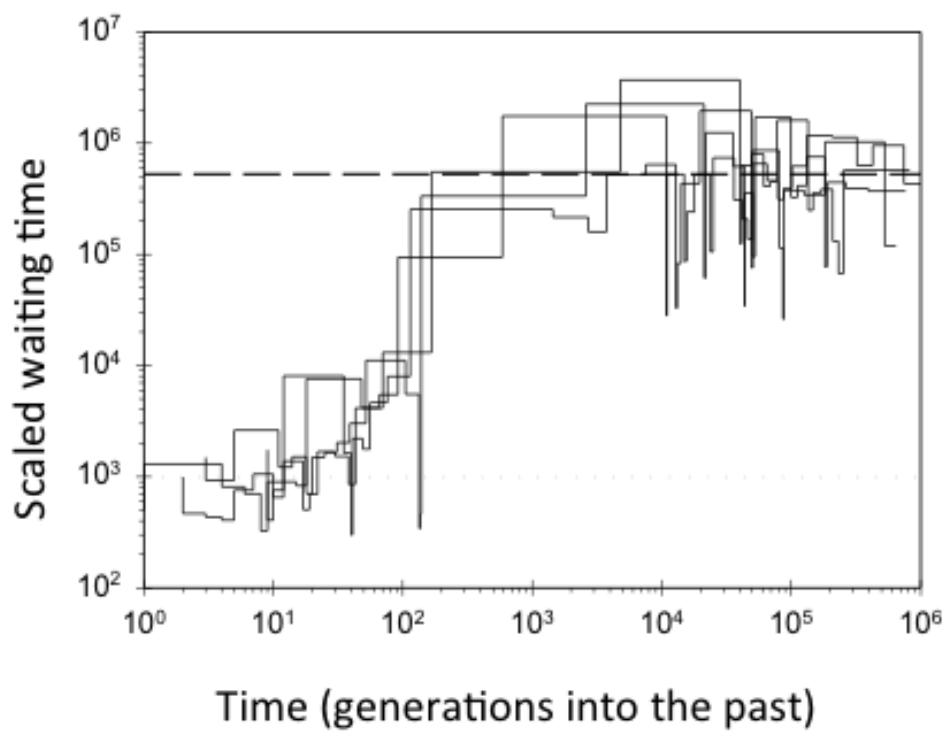
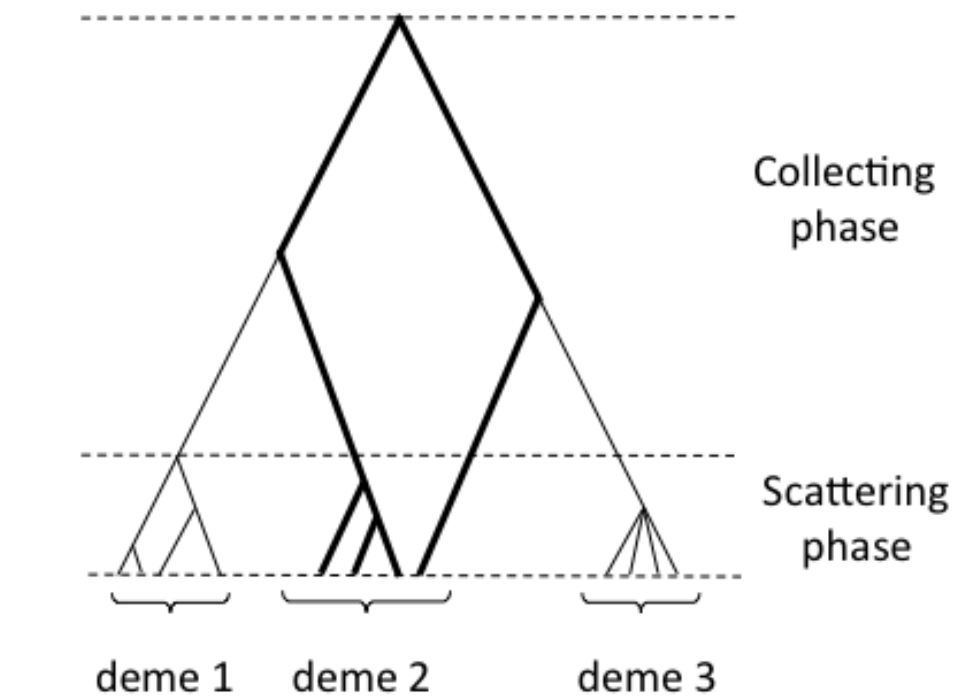


Figure 2. Coalescence in single populations that have gone through (a) a severe bottleneck and (b) a milder bottleneck. In a severe bottleneck, coalescence will be rapid during the bottleneck, with all coalescences coalescing in rapid succession (or even together, as shown). Here, we are likely to find a single peak in the site frequency spectrum for sequences from a sample (right). In a mild population bottleneck, some of the lineages will coalesce during the bottleneck, but those that do not may take much longer to coalesce. In this case, we expect to see two peaks in the site frequency spectrum, in a pattern resembling the coalescence events in a structured coalescent. Compare with Figure 2.

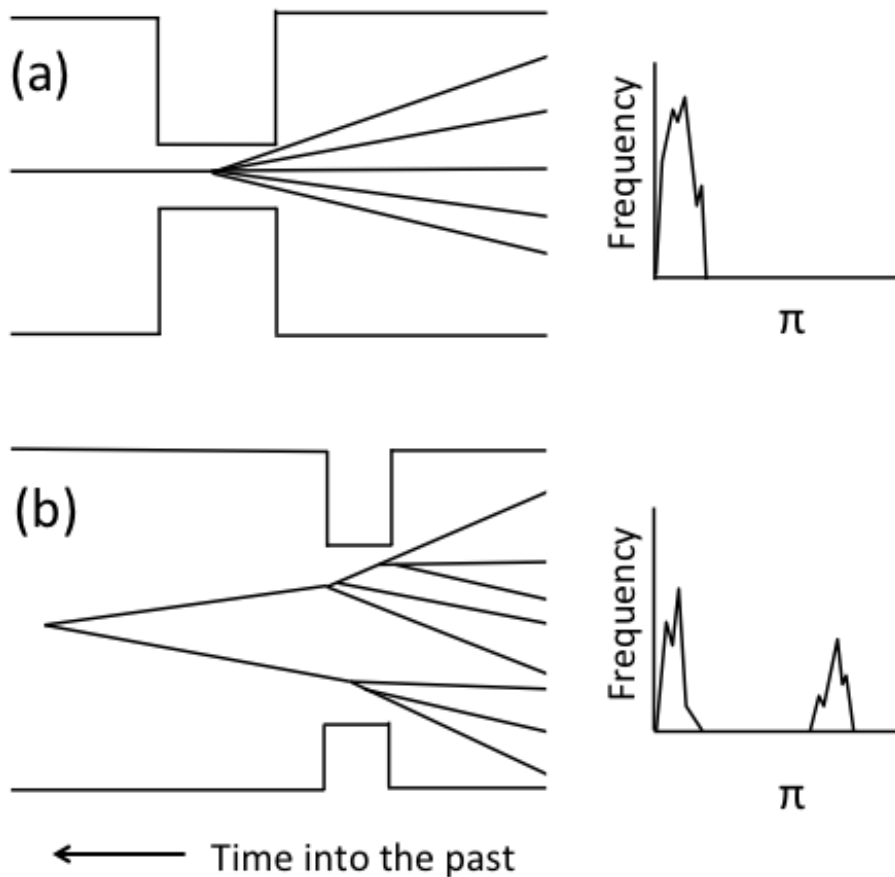


Figure 3. Among-population sexual dimorphism in calyx width in dioecious *Silene latifolia*, likely resulting from differential selection predominantly on males. The figure shows mean ( $\pm$ SE) calyx width of (a) female and (c) male flowers; (b) and (d) illustrate their differences. Additive genetic values were derived from controlled within- and among-population crosses in a greenhouse for genotypes from three populations (VIR = Giles County Virginia, USA; CRC = Cabo de Roca, Portugal; ZAG = Zagreb, Croatia). In (a) and (c), significant differences among means in calyx width are indicated with different letters (dams) or numbers (sires) above the means. Differences in  $F_{ST}$ - $Q_{ST}$  ratios between males and females suggest that males have been under stronger divergent selection for calyx width than females (ratio of 4.2, as opposed to 0.4 for females). In this study, the comparison between males and females helps to rule out the possibility that recent mutations might have influenced  $F_{ST}$  and  $Q_{ST}$  differently, because the two sexes provide a control for each other. Graphs from (Yu *et al.* 2011); images courtesy of L. Delph.



Figure 4. An application of the method of Ovaskainen et al. (2011) and Karhunen et al. (2012) to a simulated dataset of three populations (coloured symbols) measured for two traits undergoing directional selection. The inferred ancestral population mean for the simulated traits is located at 'A'. Neutral genetic data were simulated for 18 microsatellite markers, with a global  $F_{ST}$  set to 0.10. Simulations assumed that the red population was under strong directional selection, the green population under weak selection, and the black population under no selection. The ellipses depict the 50% probability sets for a given population under the effects of only drift. The results indicate that population-specific selection histories can be revealed by this approach, even for traits that are partially correlated. Simulations, as well as determination of evidence for divergent selection, were conducted using R v. 2.15.2 (R DEVELOPMENT CORE TEAM 2011).

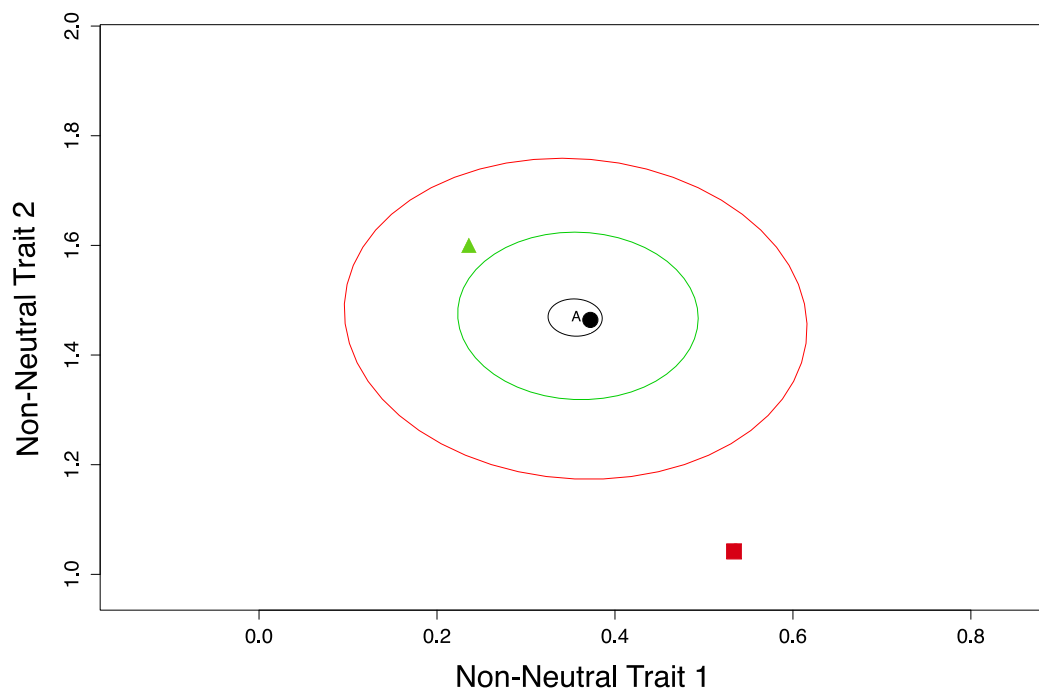


Figure 5. A comparison of three popular methods for detecting loci under divergent selection in *Populus balsamifera*. Two of the methods, BayeScan (FOLL and GAGGIOTTI 2008) and the Hierarchical Model in Arlequin (EXCOFFIER *et al.* 2009), attempt to detect signatures of local adaptation with  $F_{ST}$ -based outlier analysis, while a third method, Bayenv (COOP *et al.* 2010), tests for significant associations between particular alleles among loci and environmental variables. Here, 443 individuals were sampled from 31 populations across the species North American range and were genotyped (A) at 412 reference SNPs known to be neutral and generate a null distribution of comparison, and (B) for 339 candidate selected loci for geographically variable selection on 27 homologues of the *Arabidopsis* flowering-time network. There were a total of 43 SNPs, from 14 candidate genes, showing signatures of local adaptation, but only 10 were consistently identified by all three programs. The methods varied in their propensity to generate false positives, with the method instantiated in the program Arlequin showing proportionally larger rates. Modified line drawings courtesy of S. Keller.

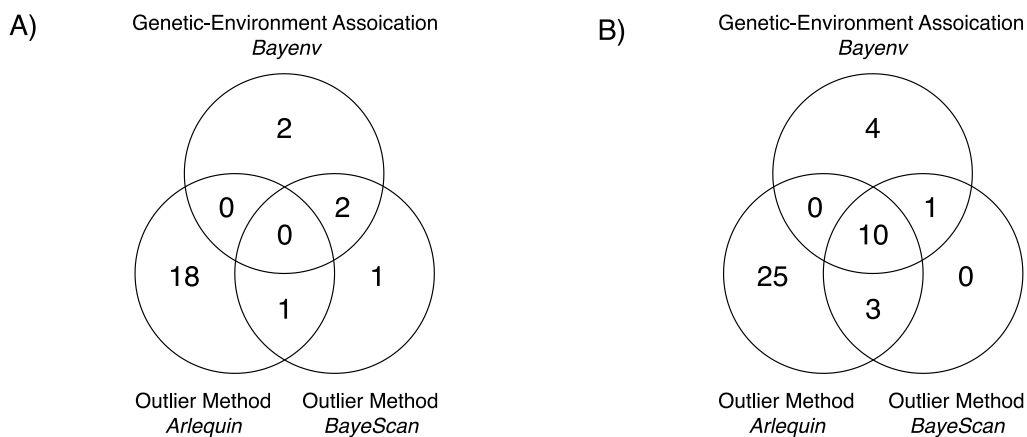
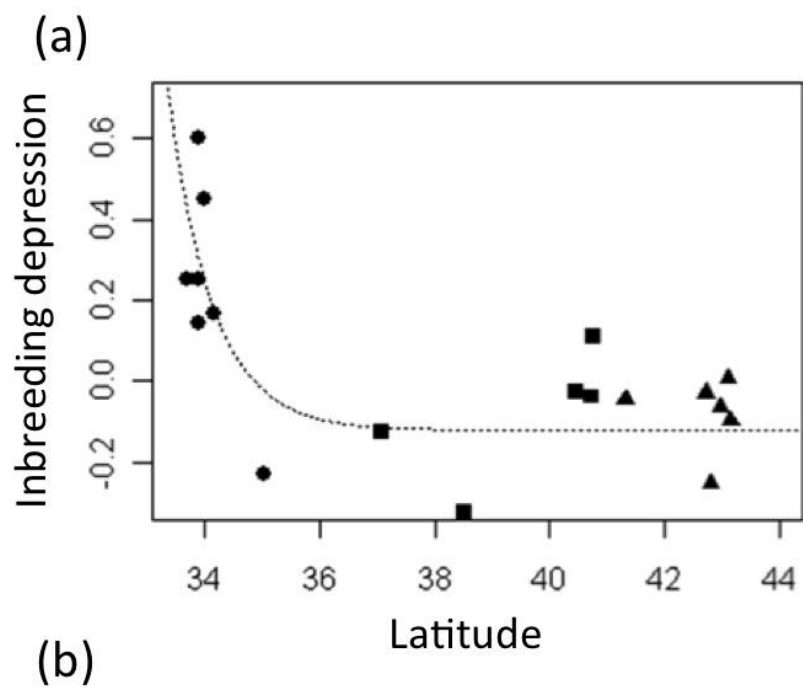


Figure 6. (a) Inbreeding depression in hexaploid Iberian and Moroccan populations of the European annual plant *Mercurialis annua* is highest in the south (Morocco), the species' putative refugium, and much reduced in northern populations towards the range boundary in the Iberian Peninsula. (b) An experiment to assay inbreeding depression in diploid populations of *M. annua* in northern Spain. Diploid *M. annua*, which expanded its range into the Iberian Peninsula from the north and east, expressed low but variable inbreeding depression at its range boundary, as did its hexaploid counterparts. Graph modified from Pujol et al. (2009). Image in (b) courtesy of S. Eppley.





(b)



Figure 7. Inbreeding depression expected in subdivided populations expressed as a ratio of that expected for an undivided species with the same parameters. Here, inbreeding depression is expressed as the the fitness of inbred individuals relative to other members of the same local populations. The relative role of population subdivision in lowering inbreeding depression depends on whether selection is soft (broken line) or hard selection (solid line). When most deleterious alleles are assumed to be recessive, population subdivision will significantly reduce the total genetic load. After Whitlock (2002).

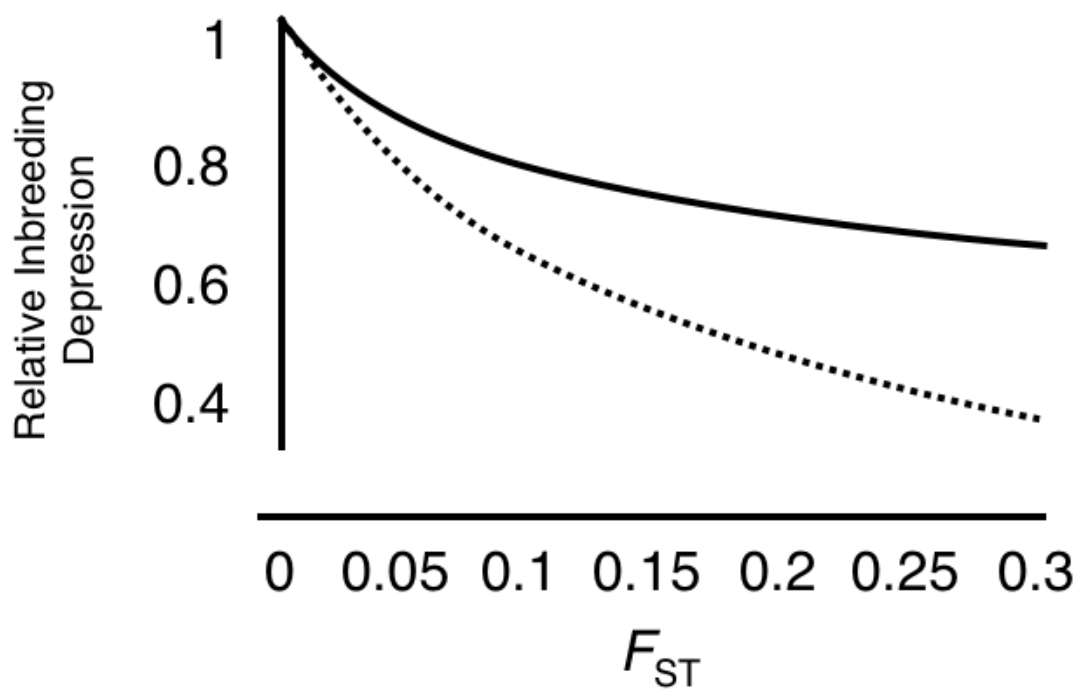


Figure 8. The change in fitness between control and  $F_2$  individuals of the perennial herb *Rutidosia leptorrhynchoides* from crosses, with pollen donor populations that varied in their effective number of alleles at microsatellite loci. Crossing target individuals with those from large populations with high genetic diversity produced greater increases in fitness than crossing them with individuals from small, genetically depauperate populations. Fitness was measured as the mean number of inflorescences, or flower heads.  $R^2 = 0.43$ ,  $P = 0.012$ . Figure modified from Pickup et al. (2013); image courtesy of A. Young.

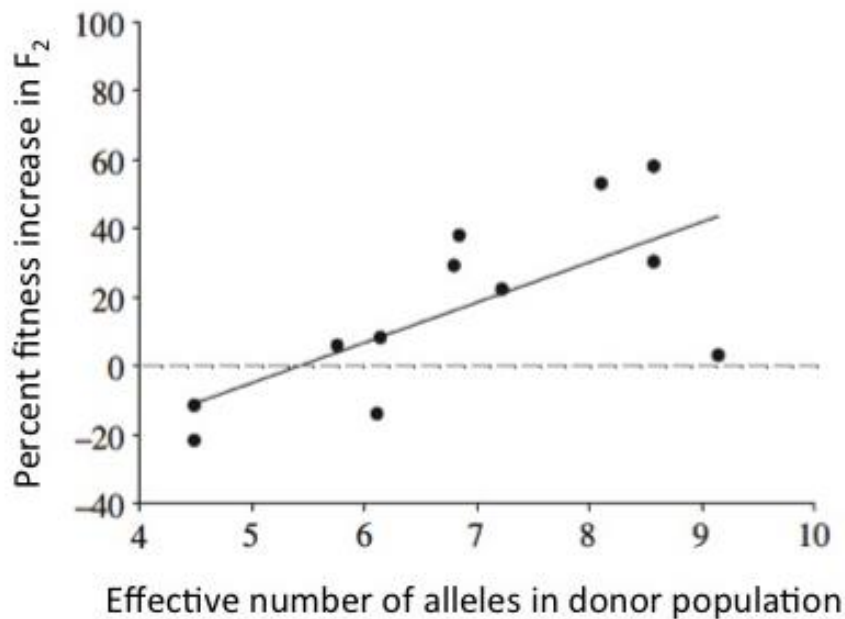


Figure 9. Magnitude of inbreeding and heterotic effects in natural populations of *Hypericum cumulicola*. (a) Spatial arrangement of 16 natural populations located in southern Highlands County, FL, USA (distance in m), were censused in 2007 and 2011. Populations containing 11-25 individuals were characterized as small, while those with populations with 124 to > 1000 individuals were considered large. Population size is indicated by column height, and column labels are population identifiers. Arrows between populations indicate estimates of gene flow, wherein the weight of the arrows are indicative of the quantity of gene flow, estimate using microsatellite markers and the software Migrate-n v.3.2.6 (BEERLI 2009). (b) Family mean cumulative fitness ( $\pm$  sem) over two years resulting from self-pollination, pollination by a different individual within the same population, and pollination by an individual from one large and one small population.

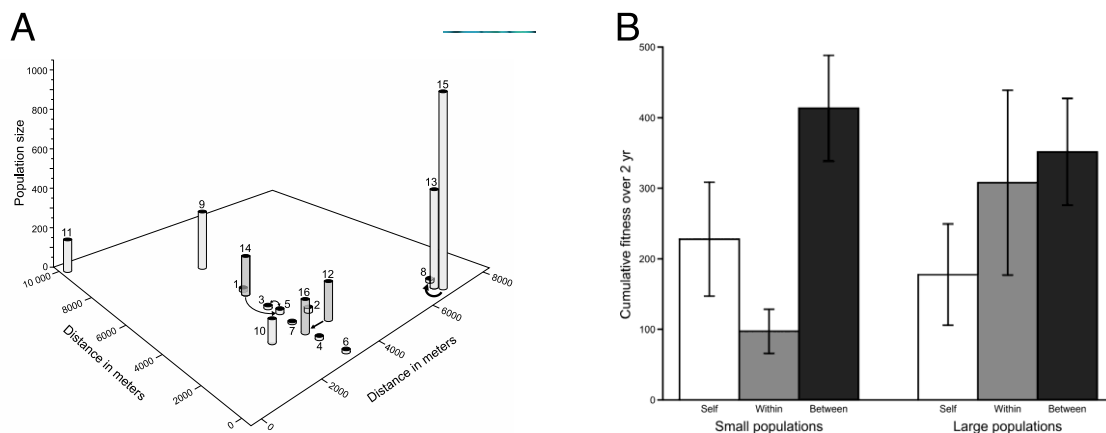




Figure 10. Phenotypic variation among populations ('accessions') of the annual herb *Arabidopsis thaliana*, grown under uniform conditions. The array shows individuals photographed at the rosette stage. Image courtesy of D. Weigel.





Chapter 2: Determinants of genetic structure in a nonequilibrium metapopulation of the plant *Silene latifolia* <sup>2</sup>

---

<sup>2</sup> Formatted for submission as a coauthored manuscript: Fields, P.D. and D. R. Taylor



**ABSTRACT**

Population genetic differentiation will be influenced by the demographic history of populations, opportunities for migration among neighboring demes and founder effects associated with repeated extinction and recolonization. In natural populations, these factors are expected to interact with each other and their magnitudes will vary depending on the spatial distribution and age structure of local demes. Although each of these effects has been individually identified as important in structuring genetic variance, their relative magnitude has never been estimated in a natural system. We conducted a population genetic analysis in a metapopulation of the angiosperm, *Silene latifolia*, from which we had more than 20 years of data on the spatial distribution, demographic history, and extinction and colonization of demes. Focusing on one subset of the metapopulation, we sampled intensively and genotyped individuals at eight highly variable microsatellite loci. We used hierarchical Bayesian methods to disentangle which features of those populations contributed to population genetic differentiation, including the magnitude and direction of their effects. We show that population age, long-term size and degree of isolation all combine to affect the distribution of genetic variance, with small, recently-founded, isolated populations contributing most to increase  $F_{ST}$  in the metapopulation. However, the effects of population size and population age are best understood as being modulated through the effects of connectivity to other extant populations, e.g.  $F_{ST}$  diminishes as populations age, but at a rate that depends how isolated the population is. These spatial and temporal correlates of population structure give insight into how migration, founder effect and within-deme genetic

drift have combined to enhance and restrict genetic divergence in a natural metapopulation.

## INTRODUCTION

Sewall Wright introduced the study of geographic population structure and its consequences with the Island Model (WRIGHT 1931). In the Island Model, populations are assumed to be stable and interconnected by migration, with population differentiation ( $F_{ST}$ ) generated by drift and/or divergent selection and diminished by gene flow and/or homogenizing selection among populations. A contrast to this equilibrium model is a metapopulation (WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990), where demes experience extinction and recolonization. Under this non-equilibrium model, founder effects can be a powerful structuring mechanism (SLATKIN 1977), and population differentiation can occur despite high levels of gene flow that acts to minimize genetic differentiation as populations age. Theoretically, the importance of these founder effects depends on several parameters including migration, extinction and colonization rates, the number of founding propagules, as well as the fraction of demes from which the colonists come (WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990).

In recent years, population genetic studies of single and multiple populations have benefitted from using an increasing number of individuals and loci (LASCoux and PETIT 2010). On the basis of a comprehensive literature search, Guichoux *et al.* (2011) recently identified approximately 8000 published population genetic analyses utilizing simple sequence repeats (SSRs) and single nucleotide

polymorphism (SNPs) in 2009 alone, many, if not most, focusing on identifying the presence and consequences of population subdivision. However, though population subdivision is often measured, less attention is given to understanding what factors determine the quantity of population genetic structure, and the magnitude (either positive or negative) of their effects.

Estimates of population structure formalized around the so-called F-model (GAGGIOTTI and FOLL 2010) provide an opportunity to reveal the underlying factors driving the magnitude of population genetic differentiation, or  $F_{ST}$ . The F-model is a likelihood-based approach that defines  $F_{ST}$  as a parameter of the full distribution of allele frequencies (BALDING and NICHOLS 1995; BHATIA *et al.* 2013; GAGGIOTTI and FOLL 2010; KARHUNEN and OVASKAINEN 2012; NICHOLSON *et al.* 2002). It accommodates differences in population size and migration rates across a species range (GAGGIOTTI and FOLL 2010), and thus has advantages over  $F_{ST}$  that estimates a single “global” value of differentiation. Foll and Gaggiotti (2006) introduced a hierarchical formulation of the F model that uses population-specific measurements to obtain priors for  $F_{ST}$ , then estimates the proportional contribution of population-specific drift and migration to the overall population genetic structure. Foll and Gaggiotti (2006) describe a hierarchical formulation that uses population specific measurements of biotic or abiotic factors in obtaining priors for  $F_{ST}$ , and in so doing, estimate the proportional contribution of population specific processes such as drift and migration in generating population genetic structure.

We used a hierarchical Bayesian approach to estimate the extent that population structure results from equilibrium process such as those in the island

model (e.g. drift, migration) versus non-equilibrium factors that operate in metapopulations (e.g. extinction and recolonization with founder effects). To accomplish this, we applied the F-model formulation to a long-term data set of the plant, *Silene latifolia*, where the size and spatial distribution of demes could be combined with information about their demographic history and age structure. We used fine-scale sampling and high-throughput genotyping methods, and applied an F-model approach to explore the spatial and temporal features of populations that affect their contributions to the observed distribution of  $F_{ST}$ , including simultaneous estimates of the magnitude and direction of these effects.

## **MATERIALS AND METHODS**

### **STUDY ORGANISM**

*Silene latifolia* Poir. (= *S. alba*, Caryophyllaceae) is a short-lived perennial plant that is broadly used as a model system for studying sex determination and sex chromosome evolution, host-pathogen dynamics, species invasion, organelle evolution, sexual dimorphism, sex ratio evolution, and evolution in structured populations (BERNASCONI *et al.* 2009).

We studied a *S. latifolia* metapopulation located in Giles and Craig Counties, Virginia, USA (Figure 1). This region has been the subject of a 20+ year study of population dynamics and genetic structure in more than 800 spatially distributed populations (ANTONOVICS *et al.* 1994; MCCAULEY 1994; MCCAULEY 1997; MCCAULEY *et al.* 1995; RICHARDS 2000; RICHARDS *et al.* 1999; RICHARDS *et al.* 2003). An annual census of an approximately 25x25-km area adjacent to Mountain Lake Biological

Station has been conducted since 1988, recording the location, gender, of *S. latifolia* individuals along ~150km of predominantly roadside habitat. The structure of the data and how it was collected is reported in Antonovics *et al.* (1994). Briefly, the roadside habitat is divided up into ~40m segments referred to as psilons. Which psilons are occupied and the number of plants in each occupied psilon are recorded annually. Important phase transitions, such as extinction and colonization events, are confirmed with a second census during the same season. Time since colonization (or population age) is based on the year plants were first observed in a given psilon. We identified extinction as the disappearance of plants from a psilon for a single year. Although a previous study in the focal metapopulation has indicated the presence of seed bank with a possible viability of approximately four years (PERONI and ARMSTRONG 2001), we interpret the recolonization of a population from a seed bank as roughly equivalent as recolonization from a nearby site with respect to the impact of founder effects on genetic diversity. *The census data provide the demographic data, spatial relationships among populations, and extinction/recolonization dynamics used in this study (Table 1).*

## **SPATIOTEMPORAL DATA**

We estimated characteristics of populations that are thought to play a significant role in genetic differentiation that occurs among demes in a metapopulation, population size, population isolation (a characteristic that will influence gene flow between populations, and population age (a factor that estimates the when founder effects occurred and opportunities for subsequent

migration).

We estimated harmonic mean population size based upon the number of plants (both flowering and vegetative) occupying a given grid-unit throughout the time that a population was extant. Detailed surveys on a subset of the overall metapopulation have been conducted to estimate the extent that our non-invasive census protocol sacrifices precision and accuracy. In the survey, we carefully searched the vegetation and when plants were found, the shoots were carefully traced down to the ground to distinguish individuals that were rooted close together. Non-flowering plants were also counted. Population sizes in the detailed survey were highly correlated with population size from the metapopulation census (0.74,  $P < 0.001$ ). Population size in the metapopulation census ( $\bar{m} = 14.32$ ) is lower than in the detailed survey ( $\bar{m} = 36.91$ ) because clustered plants may be counted as single individuals. The census, therefore, accurately measures relative population sizes, but likely underestimates absolute size.

To estimate population age, we assumed sites unoccupied for a single year were extinct. This reflects the simplifying assumption that recolonization from neighboring sites or from the seed bank involve similar bottlenecks of genetic diversity. Following extinction and re-colonization, age is calculated as the number of years a site has been occupied up until the time of collection. Given this operational definition of population age, individual sites could vary in age from one to 21 years (the extent of the demographic census started in 1988). Populations in the age class of 21 years might reasonably be considered as a heterogeneous

grouping of extant populations given the limited compass of the current long-term census.

There are a number of methods described in published literature that might be used to estimate population connectivity. The majority of ecological studies have utilized a nearest neighbor/patch approach, or distance to multiple neighbors within a limited neighborhood of a focal patch (or buffer) (MOILANEN and NIEMINEN 2002). However, these simple measures have been shown to be poor predictors of ecologically important metapopulation dynamics such as colonization potential. To measure population isolation, we generated two different composite variables, each of which includes a negative exponential dispersal kernel and accounts for distance to all other potential extant populations for gene flow (COSENTINO *et al.* 2012). Each variable has the following structure:

$$I_i = \sum_{j \neq i} \hat{a} p_j \exp(-a d_{ij}), \quad (1)$$

where  $p_j$  is in  $I_{Age}$  the proportion of the censuses for which a given site was occupied ( $1/21 - 1$ ) or in  $I_{\bar{N}}$ , the harmonic mean population size over the life span of the population, of some population  $j$ ;  $a$  is a parameter scaling the effect of distance on dispersal, and  $d_{ij}$  is the road network based distance between a target population  $i$  and source population  $j$  (COSENTINO *et al.* 2012; MOILANEN and NIEMINEN 2002). Pair-wise distances were calculated using a network constructed based upon the public roadway system, using ArcGIS (ESRI) Network Analyst tool. Given the mountain-valley geographic topology of the area, this network-based approach is

more appropriate than standard Euclidean distances in order to predict likely routes of the predominant pollinators, noctuid moths. As such, larger scores for either  $I_{Age}$  or  $I_N$  are indicative of a decrease in an individual population's probability of receiving migrants, whether through seeds or pollen, though these are modulated by the opportunity for gene flow over-time or through apparency to pollinators.

## **SAMPLING**

We sampled plants from 33 spatially distinct populations during peak flowering in the summer of 2008. Populations occur directly adjacent to public roads, and so no privately owned or protected lands were accessed. *S. latifolia* is identified as an introduced weed in the study area, and does not require specific permits to collect tissue samples. We collected leaf tissue from every plant in the population, or up to 50 individuals in the larger populations, and stored the leaves with silica gel (Sigma, USA). Genomic DNA was extracted and amplified following established microsatellite techniques for *S. latifolia* (JOLIVET and BERNASCONI 2007). We genotyped each plant at 8 microsatellite loci. Microsatellites were derived from multiple sources (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; TEIXEIRA and BERNASCONI 2007). PCR amplification was conducted using the published methods (KELLER *et al.* 2012). PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied



Biosystems, USA). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems, USA). Allele binning was accomplished using the software TANDEM (MATSCHINER and SALZBURGER 2009).

## STATISTICAL METHODS

Overall, the data involved 730 plants each with a multi-locus genotype derived from eight microsatellite markers. The plants were associated with one of 33 populations that had size, age and isolation data collected from the long-term demographic dataset.

We used the program GenoDive (MEIRMANS and VAN TIENDEREN 2004) to estimate global summary statistics of population structure for our molecular markers. We used the hierarchical Bayesian method of Foll and Gaggiotti (2006), implemented in the program GESTE, to evaluate the effect of spatial and temporal characteristics of individual populations on the magnitude of allelic variation distributed among populations within the metapopulation. We considered four factors, thus generating 16 alternative regression models. The software GESTE does not allow for testing a full model that included all pairwise interactions. The selected method provides posterior probabilities for each of the alternative models using a RJ-MCMC approach (FOLL and GAGGIOTTI 2006). Given these considerations, the model with the highest posterior probability is the one that best explains the data (FOLL and GAGGIOTTI 2006). We followed the method of (GAGGIOTTI *et al.* 2009), using

10 pilot runs of 1000 iterations to obtain parameters of the proposal distributions used by the MCMC, followed by additional burn-in of  $5 \times 10^6$  iterations and a thinning interval of 50, with a final iteration sample size of 60,000 on which the model fit probability was based. Using this method, we identified the model that best explained the observed genetic structuring. The magnitude and direction of a spatiotemporal character on genetic structuring was inferred from estimates of the regression coefficients from the model with the highest posterior probability.

## RESULTS

Global summary statistics (Table 1) revealed a high degree of population substructure, which is consistent with other published datasets on plant metapopulations, including other *Silene* species (TERO and SCHLOTTERER 2005). Previously published research has shown that the observed sub-structuring of alleles is not the consequence of PCR artifacts such as null alleles (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; TEIXEIRA and BERNASCONI 2007). The mean metapopulation-wide  $F_{ST}$  (0.103) was similar to previous estimates based upon allozymes in the same metapopulation of *S. latifolia* (0.134, McCAULEY 1994)

The magnitude and direction of the effect of each of the spatiotemporal characters on genetic differentiation was inferred from the estimates of the regression coefficients for the most probable model (GAGGIOTTI *et al.* 2009). The four individual parameters affected genetic differentiation in directions that were consistent with both theoretical expectations and previous research on the system (Table 3), i.e. founder effects and small population size enhanced  $F_{ST}$ . However,

individual, non-composite parameters for population age and harmonic mean population size did not contribute to models with the highest posterior probabilities. Instead, a composite parameter that combined population age and the degree of isolation,  $I_{Age}$ , provided the highest posterior probability (0.752) model (Table 4), followed by the model that included  $I_{Age}$  along with the composite parameter that combined population size with the degree of isolation,  $I_{\bar{N}}$  (0.104). The 14 other models attained  $< 0.1$  posterior model probability. Overall, population age was the most important determinant of population genetic structure, followed by population size, though each of these were influenced by the degree of isolation.

Posterior estimates of the regression parameters of the model that only included  $I_{Age}$  were highly significantly negative ( $\bar{m} = -0.439$ , mode = -0.447, 95% HPDI [-0.680; -0.200]), indicating that the initial increase in  $F_{ST}$  resulting from founder effects is reduced overtime. Posterior estimates of the regression parameters of the next best model indicated a consistent effect of  $I_{Age}$  ( $\bar{m} = -0.575$ , mode = -0.600, 95% HPDI [-0.857; -0.265]), and a positive effect of  $I_{\bar{N}}$  ( $\bar{m} = 0.207$ , mode = 0.201, 95% HPDI [-0.079; 0.472]) indicating that younger and smaller populations contributed proportionately greater to the overall magnitude of genetic differentiation, and that proportional effects were modulated through population connectivity.

## DISCUSSION

The goal of the present study was to simultaneously identify several important

spatiotemporal parameters that could contribute to population genetic differentiation, and to disentangle each effect with respect to its magnitude and direction. While previous research on other plant metapopulations has measured whether founder effects contribute to population differentiation (GILES and GOUDET 1997; GILES *et al.* 1998; INGVARSSON and GILES 1999; MCCAULEY 1994; MCCAULEY 1995; MCCAULEY 1997; MCCAULEY *et al.* 1995; MCCAULEY *et al.* 1996), none of these previous studies have been capable of quantifying these effects simultaneously while measuring their magnitudes. In a metapopulation of the angiosperm, *Silene latifolia*, population size, degree of isolation and population age were each important determinants of population structure, but composite characters (so-called connectivity scores; MOILANEN and NIEMINEN 2002) accounting for population age/long term size modulated by population isolation were most important. This implicates the extinction and recolonization of demes, followed by subsequent opportunities for gene flow as a population ages, as the most important driving force in the genetic differentiation among populations in this system. Importantly, measures of long term population size and age are ineffective for determining how populations will become differentiated at neutral loci without information concerning the opportunities for gene flow from extant source populations. The findings show that genetic drift within demes is not as powerful as founder effect, though both processes are influenced by the degree of isolation and hence the opportunities for gene flow.

The *S. latifolia* metapopulation has been ecologically characterized for more than 20 years. The system cannot be characterized as a true Island Model, nor an

idealized metapopulation (LEVINS 1969); populations are characterized by frequent colonizations and extinctions, on the order of 5-20% per year (ANTONOVICS 2004; ANTONOVICS *et al.* 1998; ANTONOVICS *et al.* 1994; THRALL and ANTONOVICS 1995), but populations vary in size, dispersal is limited, and within population dynamics are important relative to the time scale of the study. Colonizations are likely to be source-size and distance dependent, as is typical in other metapopulation systems (ALEXANDER *et al.* 2012; HANSKI and GAGGIOTTI 2004; HANSKI and SIMBERLOFF 1997). Our findings support the general notion that founder effects during colonization can enhance genetic differentiation among populations (MCCAULEY *et al.* 1995), but we additionally show that the magnitude of these effects are large (Figure 2) relative to the structuring mechanism assumed in most models, genetic drift among extant demes.

Theoretical explorations of metapopulation dynamics have shown that spatiotemporal characters should have direct effects on each other. We found empirical evidence for this. For example, we found a significant positive pairwise correlation (0.478,  $P < 0.01$ ) between population size and population age; older populations also tended to be larger. As such, although we show that both are potentially important in generating genetic divergence, our ability to fully disentangle the effects of age and population size is somewhat limited. Neither population size nor age was significantly correlated with population isolation ( $P > 0.05$ ).

It is nearly ubiquitous for species to have populations that are patchily distributed in space, and there are strong quantitative similarities in the

extinction/colonization and incidence parameters in other model systems; e.g. the Glanville fritillary (*Melitea cinxia*) metapopulation in Finland (HANSKI 2001). Moreover, spatial patterns of colonization and extinction respond similarly to scaling in regional studies of sunflowers in the mid-west (*Helianthus annuus*; MOODY-WEIS *et al.* 2008). While the direct quantitative values will obviously vary from system to system, the general principles to emerge is that non-equilibrium dynamics that result from the extinction and recolonization of local demes are likely to have broad biological significance.

The present discussion focuses primarily on evolutionary processes that are assumed to be driven by non-selective, drift related processes. However, previous studies have indicated the potential for metapopulation structure to have a significant outcome on selective dynamics and vice versa (WHITLOCK 2002). Being derived anonymously, the utilized markers are assumed to be neutral and to be unlinked to functionally important genomic regions (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; TEIXEIRA and BERNASCONI 2007). Although our panel of markers does not exhibit discontinuities in their distribution characteristic of outliers, given our limited panel size and the inherent complexity of detecting outliers in hierarchically structured populations (NARUM and HESS 2011), we have only limited power to test this assumption.

Natural selection could have a powerful effect on population structure, even when the alleles under selection are not closely linked to marker loci. Studies in this *S. latifolia* metapopulation have confirmed that population genetic structure, where individuals in closer proximity tend to be more related than expected from chance,

can have a significant negative average effect on individual fitness through the expression of deleterious recessive alleles (RICHARDS 2000). Because inbreeding depression will reduce average fitness and therefore population size, drift processes could be enhanced through selective reductions in population size. Because many of the young, recently colonized populations will experience inbreeding depression, gene flow from other populations may be enhanced beyond neutral expectations due to heterosis (INGVARSSON and WHITLOCK 2000; WHITLOCK *et al.* 2000). This initial enhancement in gene flow could, combined with stochastic dynamics associated with non-equilibrium metapopulation conditions, enhance the observed reductions in population specific  $F_{ST}$  over time. Thus, our analyses are unable to disentangle neutral and the selective effects of spatiotemporal metapopulation structure in generating variation in allele frequencies among populations.

The present study demonstrates that the appropriate combination of long term ecological data and population genetic analyses may be a powerful tool for studying the mechanisms that generate population structure. Spatiotemporal character has been hypothesized to have a significant effect in the generation of population genetic differentiation and its selective consequences. This fact seems to be particularly true in the *S. latifolia* metapopulation, where population connectivity, age and recent population size show significant evidence of driving population genetic differentiation.

## ACKNOWLEDGMENTS

PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT.

## LITERATURE CITED

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010 Permanent genetic resources added to Molecular Ecology resources database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ALEXANDER, H. M., B. L. FOSTER, F. BALLANTYNE, C. D. COLLINS, J. ANTONOVICS *et al.*, 2012 Metapopulations and metacommunities: combining spatial and temporal perspectives in plant ecology. *Journal of Ecology* **100**: 88-103.
- ANTONOVICS, J., 2004 Long term study of a plant-pathogen metapopulation, pp. 471-488 in *Ecology, genetics and evolution of metapopulations*, edited by I. HANSKI and O. GAGGIOTTI. Academic Press.
- ANTONOVICS, J., P. THRALL and A. JAROSZ, 1998 Genetics and the spatial ecology of species interactions: the *Silene-Ustilago* system, pp. 158-180 in *Spatial ecology: the role of space in population dynamics and interspecific interactions*, edited by D. TILMAN and P. KAREIVA. Princeton University Press.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in *Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.
- BALDING, D. J., and R. A. NICHOLS, 1995 A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**: 3-12.
- BERNASCONI, G., J. ANTONOVICS, A. BIERE, D. CHARLESWORTH, L. F. DELPH *et al.*, 2009 *Silene* as a model system in ecology and evolution. *Heredity* **103**: 5-14.



- BHATIA, G., N. J. PATTERSON, S. SANKARARAMAN and A. L. PRICE, 2013 Estimating and interpreting  $F_{st}$ : the impact of rare variants. *Genome Research*.
- COSENTINO, B. J., C. A. PHILLIPS, R. L. SCHOOLEY, W. H. LOWE and M. R. DOUGLAS, 2012 Linking extinction–colonization dynamics to genetic structure in a salamander metapopulation. *Proceedings of the Royal Society B: Biological Sciences* **279**: 1575-1582.
- FOLL, M., and O. E. GAGGIOTTI, 2006 Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875-891.
- GAGGIOTTI, O. E., D. BEKKEVOLD, H. B. H. JØRGENSEN, M. FOLL, G. R. CARVALHO *et al.*, 2009 Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic Herring as a case study. *Evolution* **63**: 2939-2951.
- GAGGIOTTI, O. E., and M. FOLL, 2010 Quantifying population structure using the F-model. *Molecular Ecology Resources* **10**: 821-830.
- GILES, B., and J. GOUDET, 1997 Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist* **149**: 507-526.
- GILES, B., E. LUNDQVIST and J. GOUDET, 1998 Restricted gene flow and subpopulation differentiation in *Silene dioica*. *Heredity* **80**: 715-723.
- GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LEGER *et al.*, 2011 Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591-611.
- HANSKI, I., 2001 Spatially realistic models of metapopulation dynamics and their implications for ecological, genetic and evolutionary processes, pp. 139-156 in

- Integrating ecology and evolution in a spatial context*, edited by J. SILVERTOWN and J. ANTONOVICS. Blackwell Science, Oxford.
- HANSKI, I., and O. GAGGIOTTI, 2004 Metapopulation biology: past, present and future, pp. 3-22 in *Metapopulation Ecology, Genetics and Evolution*, edited by I. HANSKI and O. GAGGIOTTI. San Diego Academic Press.
- HANSKI, I., and D. SIMBERLOFF, 1997 The metapopulation approach, its history, conceptual domain and application to conservation, pp. 5-26 in *Metapopulation Biology: Ecology, Genetics and Evolution*, edited by I. HANSKI and M. GILPIN. San Diego Academic Press.
- INGVARSSON, P. K., and B. E. GILES, 1999 Kin-structured colonization and small-scale genetic differentiation in *Silene dioica*. *Evolution* **53**: 605-611.
- INGVARSSON, P. K., and M. C. WHITLOCK, 2000 Heterosis increases the effective migration rate. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**: 1321-1326.
- JOLIVET, C., and G. BERNASCONI, 2007 Molecular and quantitative genetic differentiation in European populations of *Silene latifolia* (Caryophyllaceae). *Annals of Botany* **100**: 119-127.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.
- KARHUNEN, M., and O. OVASKAINEN, 2012 Estimating population-level coancestry coefficients by an admixture F model. *Genetics* **192**: 609-617.

- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- LASCOUX, M., and R. J. PETIT, 2010 The 'New Wave' in plant demographic inference: more loci and more individuals. *Molecular Ecology* **19**: 1075-1078.
- LEVINS, R., 1969 Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* **15**: 237-240.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MCCAULEY, D. E., 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1995 The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution* **10**: 198-202.
- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.
- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.

- MCCAULEY, D. E., J. STEVENS, P. PERONI and J. RAVEILL, 1996 The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* **83**: 727-731.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MOILANEN, A., and M. NIEMINEN, 2002 Simple connectivity measures in spatial ecology. *Ecology* **83**: 1131-1145.
- MOODY-WEIS, J., J. ANTONOVICS, H. ALEXANDER and D. PILSON, 2008 Predicting local colonization and extinction dynamics from coarser-scale surveys. *Ecography* **31**: 61.
- NARUM, S. R., and J. E. HESS, 2011 Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**: 184-194.
- NICHOLSON, G., A. SMITH, F. JÓNSSON, Ó. GÚSTAFSSON, K. STEFÁNSSON *et al.*, 2002 Assessing Population Differentiation and Isolation from Single-Nucleotide Polymorphism Data, pp. 695-715 in *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*.
- PERONI, P. A., and R. T. ARMSTRONG, 2001 Density, dispersion and population genetics of a *Silene latifolia* seed bank from southwestern Virginia. *Journal of the Torrey Botanical Society* **128**: 400-406.
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.

- RICHARDS, C., S. CHURCH and D. MCCAULEY, 1999 The Influence of Population Size and Isolation on Gene Flow by Pollen in *Silene alba*. *Evolution* **53**: 63-73.
- RICHARDS, C. M., S. N. EMERY and D. E. MCCAULEY, 2003 Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity* **90**: 181-186.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**: 253-262.
- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.
- TERO, N., and C. SCHLOTTERER, 2005 Isolation and characterization of microsatellite loci from *Silene tatarica*. *Molecular Ecology Notes* **5**: 517-518.
- THRALL, P. H., and J. ANTONOVICS, 1995 Theoretical and Empirical Studies of Metapopulations - Population and Genetic Dynamics of the *Silene-Ustilago* System. *Canadian Journal of Botany-Revue Canadienne De Botanique* **73**: S1249-S1258.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191-1202.
- WHITLOCK, M. C., P. K. INGVARSSON and T. HATFIELD, 2000 Local drift load and the heterosis of interconnected populations. *Heredity* **84**: 452-457.

WHITLOCK, M. C., and D. E. McCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups.

Evolution **44**: 1717-1724.

WRIGHT, S., 1931 Evolution in mendelian populations. Genetics **16**: 97-159.

**Table 1.** Global estimates of genetic diversity and variation in allele frequencies.

Overall, populations showed a high degree of substructure, as has been observed in other *Silene* metapopulations. Population structure was high for each marker. Our lowest  $F_{ST}$  corresponded to the only marker composed of a dinucleotide repeat. <sup>1</sup> (ABDOULLAYE *et al.* 2010), <sup>2</sup> (TEIXEIRA and BERNASCONI 2007), <sup>3</sup> (JUILLET *et al.* 2003).

Locus	$N$	$H_o$	$H_s$	$F_{ST}$
<i>slat_18</i> <sup>1</sup>	10	0.338	0.592	0.09
<i>slat_32</i> <sup>1</sup>	9	0.303	0.673	0.135
<i>slat_33</i> <sup>1</sup>	3	0.07	0.13	0.231
<i>slat_48</i> <sup>1</sup>	2	0.078	0.202	0.143
<i>slat_72</i> <sup>1</sup>	12	0.325	0.593	0.055
<i>slat_85</i> <sup>1</sup>	16	0.324	0.538	0.125
<i>SL_8</i> <sup>2</sup>	43	0.693	0.819	0.051
<i>SV_11</i> <sup>3</sup>	11	0.305	0.573	0.122
<b>Overall</b>	<b>13.25</b>	<b>0.304</b>	<b>0.515</b>	<b>0.103</b>

**Table 2.** *S. latifolia* populations used in the genetic analysis. Population ID = census ID of individual populations, Mean, Mode, and 95% HPDI of the RJ-MCMC derived posterior model estimates, Age = number of continuously occupied years of individual populations, Harmonic Mean Population Size = harmonic mean population size of each sampled population,  $I_{Age}$  = composite character composed of the sum distance between a focal population to all other extant populations and a focal populations age, and  $I_{\bar{N}}$  = composite character composed of the sum distance between a focal population to all other extant populations and a focal populations harmonic mean population size.

Population ID	<i>F<sub>ST</sub></i> Statistics			Age	Harmonic Mean Population Size	Factors	
	Mean	Mode	95% HPDI			$I_{Age}$	$I_{\bar{N}}$
Population 1	0.183	0.174	[0.128; 0.248]	21	22.038	2.150	44.543
Population 2	0.200	0.188	[0.111; 0.302]	3	1.246	0.143	1.246
Population 3	0.108	0.096	[0.041; 0.186]	3	4.485	0.510	9.304
Population 4	0.177	0.147	[0.0719; 0.296]	1	4.820	0.228	9.675
Population 5	0.064	0.058	[0.0298; 0.1]	21	13.477	6.192	54.116
Population 6	0.046	0.043	[0.0272; 0.0665]	21	9.023	4.370	18.625
Population 7	0.056	0.051	[0.0247; 0.0886]	18	5.956	5.308	23.917
Population 8	0.048	0.045	[0.0244; 0.0744]	21	10.309	5.218	21.658
Population 9	0.041	0.039	[0.0242; 0.06]	20	9.389	6.540	38.816
Population 10	0.044	0.039	[0.0197; 0.0718]	12	3.774	3.924	15.602
Population 11	0.043	0.038	[0.0199; 0.0715]	21	8.873	6.098	27.936
Population 12	0.048	0.045	[0.028; 0.0713]	18	1.733	5.362	7.004
Population 13	0.071	0.054	[0.014; 0.142]	21	7.827	6.256	31.624
Population 14	0.064	0.059	[0.0287; 0.105]	17	4.160	3.808	9.361
Population 15	0.024	0.022	[0.011; 0.0376]	10	6.521	1.663	8.719



Population 16	0.156	0.139	[0.0743; 0.248]	21	5.255	4.213	13.407
Population 17	0.099	0.088	[0.037; 0.165]	12	2.051	2.627	6.986
Population 18	0.139	0.129	[0.0784; 0.209]	4	3.840	0.876	13.077
Population 19	0.179	0.166	[0.0888; 0.275]	3	1.503	0.474	3.963
Population 20	0.106	0.098	[0.0533; 0.163]	1	1.392	0.139	3.147
Population 21	0.166	0.150	[0.0809; 0.269]	1	1.507	0.055	1.543
Population 22	0.155	0.137	[0.0669; 0.252]	12	5.771	1.175	11.573
Population 23	0.110	0.106	[0.0724; 0.151]	11	1.807	1.077	3.622
Population 24	0.156	0.148	[0.0962; 0.223]	3	2.345	0.190	2.408
Population 25	0.218	0.205	[0.119; 0.334]	1	1.486	0.066	1.506
Population 26	0.116	0.107	[0.0619; 0.173]	20	4.220	1.343	5.159
Population 27	0.124	0.111	[0.0572; 0.2]	21	7.679	2.696	15.767
Population 28	0.075	0.066	[0.0205; 0.139]	21	17.038	2.358	21.839
Population 29	0.124	0.113	[0.0649; 0.188]	21	19.110	2.283	21.082
Population 30	0.113	0.107	[0.0718; 0.161]	21	45.220	2.335	51.538
Population 31	0.143	0.133	[0.0885; 0.209]	4	1.535	0.614	3.454
Population 32	0.136	0.129	[0.0818; 0.193]	21	26.265	2.568	39.887
Population 33	0.104	0.101	[0.0677; 0.143]	1	1.161	0.128	2.166

---

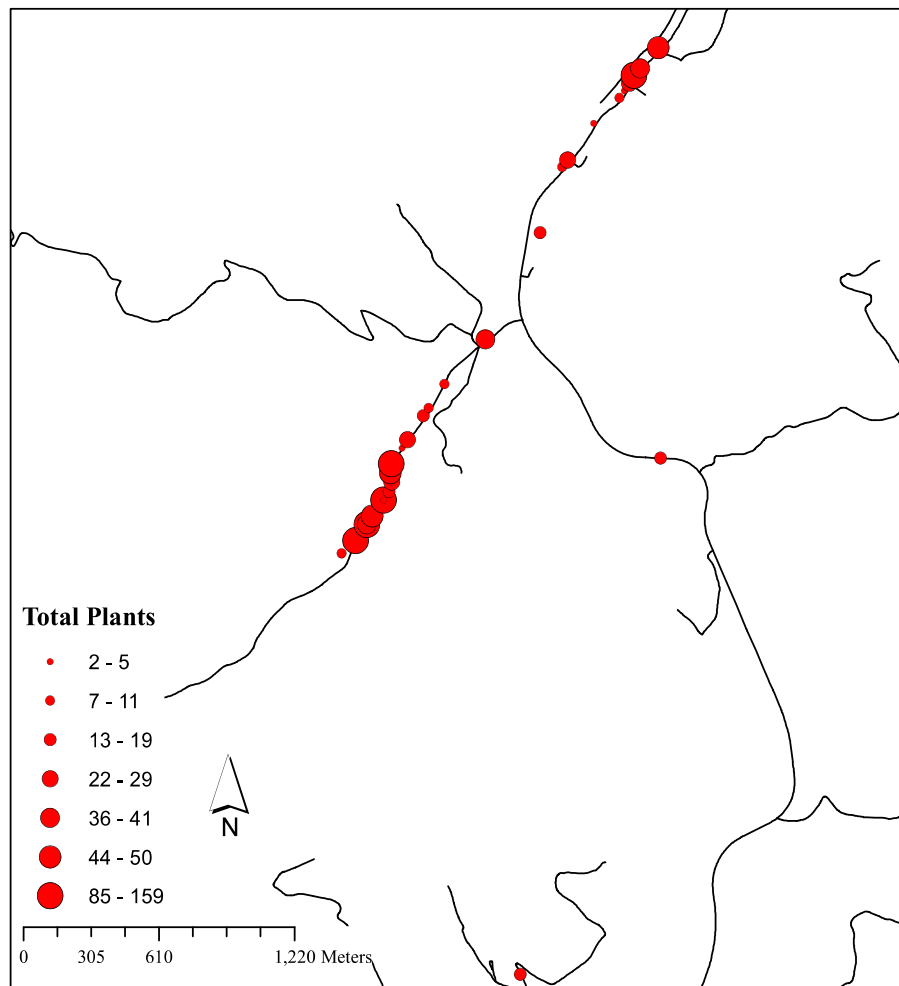
**Table 3.** Posterior probabilities of top two models. Our most probable model (bolded) included the composite variable of population age and population isolation. The second most probable model includes composite variables of population age and isolation, and harmonic mean population size and isolation.

Model	Pr	Factors included
<b>5</b>	<b>0.752</b>	$I_{Age}$
13	0.104	$I_{Age}, I_{\bar{N}}$

**Table 4.** Posterior estimates of regression parameters for the model with the highest posterior probability when only three factors are considered. Parameter estimates are consistent with theoretical expectations (e.g. older and larger populations contribute proportionately less to the global  $F_{ST}$ ).

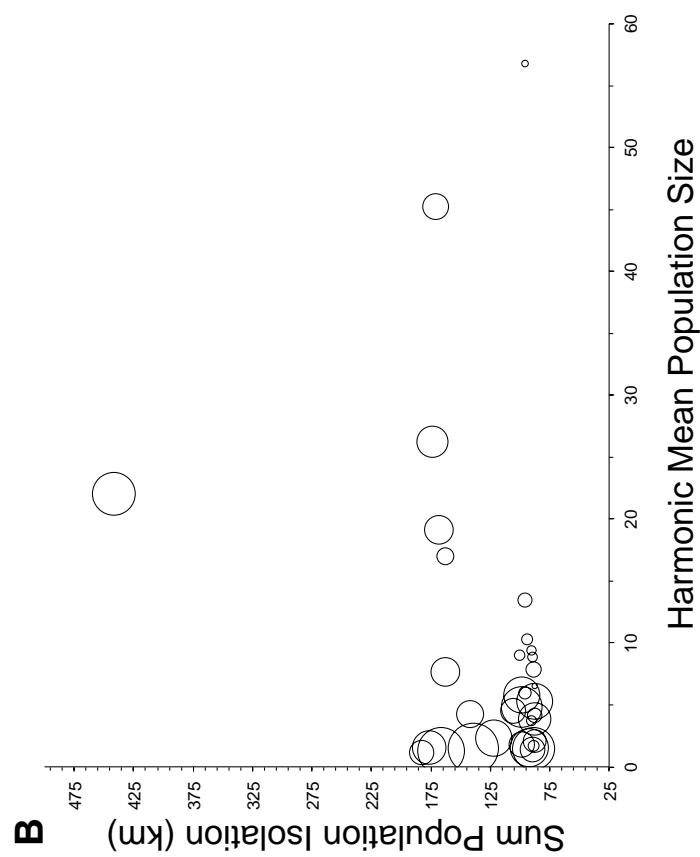
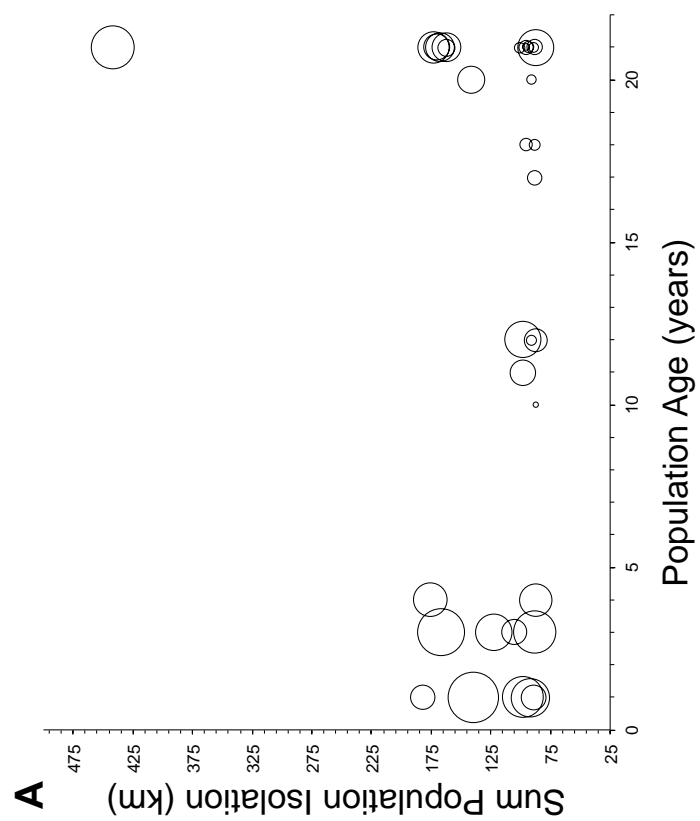
Regression coefficient	Factor	Mean	Mode	95% HPDI
$a_0$	Constant	-2.26	-2.25	[-2.51; -2.02]
$a_3$	$I_{Age}$	-0.439	-0.447	[-0.680; -0.200]
$S^2$	-	0.369	0.330	[0.169; 0.614]

**Figure 1.** Map of the focal populations of the *S. latifolia* metapopulation located in Giles and Craig County, VA, USA sampled in the presented analysis. Circles represent individual populations, where the size of the circle indicates the total number of plants located within our grid. Black lines represent both the assumed grid of population arrangement and small country roads. Due to the topology of the focal area (mountain, valley systems), it was assumed that pollinators move along the linear grid, rather than crossing over ridges.



**Figure 2.** The effect of population age, isolation, and harmonic mean size on  $F_{ST}$ .

Individual population  $F_{ST}$  is represented by the size of the circle, where larger circles represent larger  $F_{ST}$ . (A) Simultaneous effect of population isolation and population age effects on  $F_{ST}$ , and (B) of population isolation and harmonic mean population size on  $F_{ST}$ .



Chapter 3: Molecular-marker specific determinants of population genetic structure  
in a metapopulation of *Silene latifolia* <sup>3</sup>

---

<sup>3</sup> Formatted for submission as a coauthored manuscript: Fields, P.D. and D. R. Taylor

## ABSTRACT

Genomic heterogeneity in the degree of population structure is expected to arise due to competing processes that drive population divergence (e.g. directional selection and drift) and those that homogenise allele frequencies (e.g. gene flow and stabilizing selection). Population genetic software utilizes a number of simplifying assumptions in order to make estimation of summaries tractable. In particular, most commonly used estimators assume homogeneity of evolutionary processes of a given marker type. Consistent with these heterogeneous evolutionary processes, analysis may 1) result in different summaries from the same processes, or 2) be more or less effective in detecting certain types of evolutionary dynamics, or 3) both. Here we present a novel comparison of the consequences of spatiotemporal factors generating population genetic structure in anonymous vs. EST derived microsatellites or SSRs in a well characterized metapopulation of the angiosperm, *Silene latifolia*, where the proportional contribution of metapopulation effects toward generating population structure could be identified with a long-term data set. We show that the allelic structuring of different genomic regions, one thought to be more generally neutral (anonymous) and the other functional (EST), will be differently affected by metapopulation processes. Specifically, population structure of anonymous markers was enhanced by founder effects and population isolation, and decreased by population size. In contrast, the population structure of genomic regions deriving from functional regions was structured by population isolation alone. We also provide a potential explanation for the differences in structuring based upon the biologically relevant dynamics of selection against inbreeding



depression affecting functional portions of the genome disproportionately to non-functional regions of the genome.

Keywords: Metapopulation Genetics, *Silene latifolia*, Angiosperms, Population Structure

## INTRODUCTION

The extent to which plant populations are genetically divergent depends on the balance between processes that drive them apart and those that homogenize them (PANNELL and FIELDS In Press). For neutral loci, gene flow can prevent much genetic divergence between demographically stable populations (WRIGHT 1931). For loci under selection, genetic divergence can be restricted or enhanced, relative to the neutral case, depending on whether different phenotypes are selected in different populations or whether the same phenotype is selected globally. An important corollary is that because action of selection is expected to vary among loci, the magnitude of population divergence will be locus-specific.

Quantifying population subdivision through the estimation of  $F_{ST}$  remains a consistent feature of many population genetic studies (GUICHOUX *et al.* 2011; LASCoux and PETIT 2010). Considering the explosion of resources being developed with next-generation sequencing technologies, and the concomitant reduction in costs required for developing molecular markers such as microsatellites and SNPs, there will be a greater focus on determining the particular forces resulting in population structure (e.g. drift, local adaptation) and how this heterogeneous record of historical processes results in the observed species genome. This fact is evident from the range of methodologies currently in use for detecting particular loci undergoing selection, particularly those that utilize outlier identification or genetic-environment associations (BEAUMONT and NICHOLS 1996; COOP *et al.* 2010). In general, some loci will be affected more by neutral forces, others by selective forces. Markers within genes being the latter, markers within noncoding regions being the

former. Thus, there is insight to be gained by comparing these different classes of markers.

Different estimators of population substructure all have assumptions, many of which are inconsistent with specific details of a range of systems of biological interest, or natural populations generally (WHITLOCK and MCCAULEY 1999). Under Wright's (1931) model, or neutral alleles in an Island Model, an equilibrium balance between drift and migration generates the well-known expectation that  $F_{ST} \approx 1/(4Nm + 1)$ , where  $N$  is the effective population size of each population and  $m$  is the migration rate between populations. The Island Model predicts low values of  $F_{ST}$ , and therefore low differentiation, unless migration is very rare. Additional molecular marker specific considerations requires the inclusion of mutation rate parameters, where  $F_{ST}(\text{or } G_{ST}) \approx 1/(4Nm + 4N\mu + 1)$ , where  $\mu$  is the mutation rate, but can be simplified to its earlier, simpler formulation when  $m \gg \mu$  (WHITLOCK 2011). While this expectation may be reasonable for SNPs (OSSOWSKI *et al.* 2010), it has been found to be a poor reflection of reality for microsatellite markers, which have been shown empirically to have mutation rates ranging from  $10^{-6}$  to  $10^{-2}$  (ELLEGREN 2000; PAYSEUR *et al.* 2011; WHITLOCK 2011; WHITTAKER *et al.* 2003).

Consequently, 1) markers are subject to different processes (drift vs. selection), 2) markers are affected by the same processes but differ in our ability to detect them, or 3) both. While this outcome is generally understood, little empirical theoretical and empirical analysis has been dedicated to determining scenarios where marker specific evolutionary processes will aid or hinder inference, especially those based upon summaries such as  $F_{ST}$ . MARUKI *et al.* (2012) show that

regions of the genome experiencing higher purifying selection result in population genetic summaries of structure (e.g.  $F_{ST}$ ,  $\theta$ , and Jost's  $D_{EST}$ ) lower than regions experiencing less purifying selection, and thereby faster evolutionary rates. There are now a number of published studies comparing the level of population genetic structure of microsatellites derived from anonymous genomic regions using more traditional methods and those deriving from transcribed, potentially functional regions (expressed sequence tag, EST, libraries or transcriptomes). Microsatellites in functional regions might reasonably be assumed to be under selection, though direct comparisons with anonymously derived microsatellites are inconclusive at best. In the sea grass *Zostera* sp., authors compared the degree of differentiation between habitats of alleles at anonymous microsatellite loci with differentiation at EST-linked microsatellites, finding no clear difference between the two classes of marker (OETJEN and REUSCH 2007). In contrast, LAZREK *et al.* (2009) working with *Medicago truncatula* suggest that these respective binning of loci are actually affected by different biotic selective pressures.

Here we present a novel comparison of the consequences of spatiotemporal factors generating population genetic structure in anonymous vs. EST derived microsatellites in a well characterized metapopulation of the angiosperm, *Silene latifolia*, where the proportional contribution of metapopulation effects toward generating population structure could be identified with a long-term data set where the size and spatial distribution of demes could be combined with information about their demographic history and age structure. Rather than simply compare  $F_{ST}$  values among the two classes of markers, we use a series of hierarchical Bayesian model

tests to show that a subset of markers is contributing significantly more to the global population structure, as well as show that while the anonymous markers exhibit theoretical expectations for the mechanism of structuring, the EST-derived markers are primarily structured by spatial isolation. We also provide a potential explanation for the differences in structuring based upon the biologically relevant dynamics of selection against inbreeding depression affecting functional portions of the genome disproportionately to non-functional regions of the genome.

## **MATERIALS AND METHODS**

### **STUDY ORGANISM**

Our focal populations of *S. latifolia* derive from a metapopulation located in Giles and Craig Counties, Virginia, USA (Figure 1). This region has been the subject of a 20+ year study of population dynamics and genetic structure in more than 800 spatially distributed populations (ANTONOVICS *et al.* 1994; MCCAULEY 1994; MCCAULEY 1997; MCCAULEY *et al.* 1995; RICHARDS 2000; RICHARDS *et al.* 1999; RICHARDS *et al.* 2003). An annual census of an approximately 25x25-km area adjacent to Mountain Lake Biological Station has been conducted since 1988, recording the location, gender, of *S. latifolia* individuals along ~150km of predominantly roadside habitat. The structure of the data and how it was collected is reported in Antonovics *et al.* (1994). Briefly, the roadside habitat is divided up into ~40m segments referred to as psilons. Which psilons are occupied and the number of plants in each occupied psilon are recorded annually. Important phase transitions, such as extinction and

colonization events, are confirmed with a second census during the same season. Time since colonization (or population age) is based on the year plants were first observed in a given site. We identified extinction as the disappearance of plants from a psilon for a single year, though a previous study in the focal metapopulation has indicated the presence of seed bank with a possible viability of approximately four years (PERONI and ARMSTRONG 2001). The census data provide the demographic data, spatial relationships among populations, and extinction/recolonization dynamics used in this study.

### **SPATIOTEMPORAL DATA**

We estimated characteristics of populations that are known to play a significant role in theoretical and previously determined empirical studies of metapopulation dynamics: population size, population isolation (a characteristic that will influence gene flow between populations, as well as the magnitude of drift a population is subject to), and population age (a factor that estimates the recency of founder effects and opportunities for subsequent migration).

We estimated recent population size based upon the number of plants (both flowering and vegetative) occupying population site at the time of tissue collection (summer 2008). We also included a measure of long-term population size, specifically the harmonic mean population size of the full extent of a populations observed occupancy, though this measure showed very little support in posterior model comparisons (data not shown, see Chapter 2).

Detailed surveys on a subset of the overall metapopulation have been conducted to estimate the extent that our non-invasive census protocol sacrifices

precision and accuracy in the estimation of population size. In the survey, we carefully searched the vegetation and when plants were found, the shoots were traced down to the ground to distinguish individuals that were rooted close together. Non-flowering plants were also counted. Population sizes in the detailed survey were highly correlated with population size from the metapopulation census ( $r=0.74$ ,  $P<0.001$ ). Population size in the metapopulation census (mean = 14.32) is lower than in the detailed survey (mean = 36.91) because clustered plants many be counted as single individuals (Douglas Taylor, unpublished data). The census, therefore, accurately measures relative population sizes, but likely underestimates absolute size. This bias is probably not constant across the range of population sizes though (greater bias for larger pops; smaller pops estimated with greater precision).

We used a sum total of all pair-wise distances of a focal population to all other sampled populations as our measure of population isolation (MOILANEN and NIEMINEN 2002). Larger isolation scores are indicative of a decrease in an individual population's probability of receiving migrants, whether through seeds or pollen. We did not include size in our measure of isolation, as population size was being analyzed as a separate factor. Empirical tests have indicated substantial decreases in gene flow rate over the range of 20 to 80 meters in the focal *S. latifolia* metapopulation, though there was an interaction between population size and distance to source (RICHARDS *et al.* 1999). We did not identify a significant interaction between the effects of population size and our measure of connectivity (data not shown).

To estimate population age, we assumed sites unoccupied for a single year as extinct. This reflects the simplifying assumption that recolonization from neighboring sites or from the seed bank involve similar bottlenecks of genetic diversity. Following extinction and recolonization, age is calculated as the number of years a site has been occupied up until the time of collection. Given this operational definition of population age, individual sites could vary in age from one to 21 years (the extent of the demographic census started in 1988). Populations in the age class of 21 years might reasonably be considered as a heterogeneous grouping of extant populations given the limited compass of the current long-term census.

## **SAMPLING**

We sampled plants from 33 spatially distinct populations during peak flowering in the summer of 2008. We collected leaf tissue from every plant in the population, or up to 50 individuals in the largest populations, and stored the leaves in silica gel (Sigma). Our microsatellites were derived from multiple sources and are of two different types: 1) microsatellites derived whole genome enrichment (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; TEIXEIRA and BERNASCONI 2007), and 2) floral expressed sequence tag simple sequence repeats (EST-SSR) (MOCCIA *et al.* 2009). PCR amplification was conducted using published methodologies for the focal species (KELLER *et al.* 2012). PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size



standard (Applied Biosystems). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems). Allele binning was accomplished using the software TANDEM (MATSCHINER and SALZBURGER 2009).

## STATISTICAL METHODS

In order to initially validate our *a priori* expectation that our two different marker sets might be undergoing very different evolutionary processes, we used the program GenoDive (MEIRMANS and VAN TIENDEREN 2004) to estimate global summary statistics of population (sub)structure for our molecular markers, with 10,000 permutations and  $\alpha = 0.05$  for tests of significant  $F_{ST}$ .

In order to determine if either set of molecular markers were contributing significantly more to estimates of global population structure, we used the program BayeScan (FOLL and GAGGIOTTI 2008). Specifically, we used the method of Foll and Gaggiotti (2008) not to detect outlier loci, but rather to test for significant differences between the two marker types in generating the signatures of population structure. Using v. 2.1 of BayeScan, we utilized the default settings, with a total of 20 pilot runs, each with a length of 5000 iterations to tune the RJMCMC chain, a burn-in of 50,000 iterations, and a total of 10,000 retained iterations (100,000 iterations with a thinning interval of 10) of the RJMCMC run. Analyzing the MCMC trace of the model posteriors and parameter estimates identified model convergence.

We used the hierarchical Bayesian method of Foll and Gaggiotti (2006), implemented in the program GESTE, to evaluate the effect of spatiotemporal characteristics of individual populations on the magnitude of population genetic differentiation in our focal metapopulation of *S. latifolia*. We considered three factors, leading to a total of  $2^3$  (8) alternative models, starting with only a constant (or null), and running through each factor individually or in concert with one, two, or three factors in a given model. The selected method provides posterior probabilities for each one of the alternative models using a RJMCMC approach (FOLL and GAGGIOTTI 2006). Given these considerations, the model with the highest posterior probability is then one that best explains the data (FOLL and GAGGIOTTI 2006). We followed the method of Gaggiotti, Bekkevold et al. (2009), using 10 pilot runs of 1000 iterations to obtain parameters of the proposal distributions used by the MCMC, followed by additional burn-in of  $5 \times 10^6$  iterations and a thinning interval of 50, and final iteration sample size of 60,000 on which the model fit probability was based. This procedure was carried out for both sets of molecular markers individually, as well as combined. In order to determine if any single marker was contributing to the creation of the pool marker effect, we also applied the approach of (GAGGIOTTI *et al.* 2009), wherein GESTE models were run excluding one marker at a time and comparing the posterior probabilities of factors when a given marker was either included or excluded. No single marker contributed significantly more or less to the overall pattern (data not shown).

## RESULTS

Global summary statistics of population structure indicated that there did indeed seem to be two distinct clusters of markers, mostly though not completely delineated by molecular marker origin based upon their distribution of  $F_{ST}$ , as well as their deviation in observed and expected heterozygosity (Table 1)(ELLEGREN 2000; ELLEGREN 2004; WHITLOCK 2011; WHITTAKER *et al.* 2003).

Posterior probabilities for each marker contributing significantly more or less to overall population structure were consistent with the initial global summaries, with seven of the eight anonymously derived markers (and one of the EST markers) showing a posterior probability greater than 95% that  $\alpha \neq 0$  (i.e. the model parameter determining if a given marker does not significantly contribute to the increase or decrease of  $F_{ST}$ ; Table 2).

Most importantly, posterior model comparisons from GESTE indicated that each respective set of molecular markers, and their allelic variation, differing in genomic origin are being structured to different degrees by spatiotemporal factors. Analysis of the factors affecting allelic variation across the metapopulation in anonymously derived microsatellites were generally consistent with theoretical expectations (WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990), and previously published empirical study of the metapopulation using allozyme and microsatellite data (FIELDS and TAYLOR In Prep.; MCCAULEY 1994; MCCAULEY *et al.* 1995). That is, all three factors, population age, size, and isolation, occur in the highest posterior probability model (Table 3). Additionally, directions of regression coefficients are consistent with biological expectations (Table 4). However, previous empirical analyses were not capable fully disentangling the magnitude of effect of

each factor, and as can be seen in both the posterior model probabilities, and in the regression coefficients of the highest posterior model, population age has the strongest effect on allelic variation for anonymously derived microsatellite markers.

The highest posterior model for the EST derived markers diverges from those for anonymous markers in that only a single spatiotemporal factor seems to be most important for population structure, notably population isolation. Posterior estimates of regression coefficient for population isolation were in consistent direction with the anonymous markers, though the magnitude of effect is considerably larger.

Combining the two sets of markers results in a combination of the two most important factors across marker sets, population age and isolation, giving the highest posterior model probability. Importantly, the model including only a constant, or our null model, does not figure heavily in any of the model comparisons. Additionally, 95% high posterior density intervals, or HPDI, on regression values, though wide in some cases, never overlap with zero, and so can be considered significantly greater than zero effect.

## **DISCUSSION**

In the present study we describe divergence in the set of factors responsible for determining population genetic structure in a well-characterized plant metapopulation of the species *S. latifolia*. While previous studies have indicated different levels of structure in loci derived from anonymous vs. functional regions of the genome (LAZREK *et al.* 2009), ours is the first capable of directly determining the

metapopulation focused spatiotemporal dynamics responsible for this deviation.

For more traditional, anonymously derived microsatellite loci, we found that standard theoretical expectations were met, viz. population age, size, and isolation were all components of the highest posterior model. Significantly, age was the most important factor in determining population genetic structure, as can be seen by the magnitude of the regression component in the highest posterior model, as well as the distinct loss of posterior model probability when the factor is not present in the model. This observation is consistent with a number of previous empirical results, where it was found that the mode of population colonization, most closely likened to Slatkin's (1977) propagule model of population colonization and migration, initially enhances population differentiation which can be subsequently eroded by ongoing gene flow from older, extant populations (FIELDS and TAYLOR In Prep.; GILES and GOUDET 1997; MCCAULEY *et al.* 1995).

In contrast, markers derived from a floral transcriptome of the focal species (MOCCIA *et al.* 2009) suggest population genetic structure is primarily determining by population isolation. A number of potential explanations for these differences exist, though they are not mutually exclusive. Given that the quantity of population structure in EST based markers was lower than that of the anonymous markers, it is unlikely that the EST markers are undergoing divergent selection. Alternatively, anonymous markers might be undergoing less purifying selection; because EST markers are in transcribed regions they might be under less selective constraint. However, this does not seem to be the case either, as the EST based markers have a larger average number of alleles than the anonymous markers (Table 1; note the

exception of *SL\_8*, the only marker in our panel that is a di-nucleotide motif, which is suggested to have a higher mutation rate than either tri- or tetra-nucleotide motifs in our dataset).

A more likely mechanism driving the divergence in factors generating population structure across the two marker panels is the presence of selection against inbreeding. Within the *S. latifolia* metapopulation (RICHARDS *et al.* 1999), and more broadly within the species' range (AUSTERLITZ *et al.* 2011), male fitness has been shown to be affected by relatedness of the receiving female, implying selection against inbreeding. Further, studies in other biological systems have indicated that Heterozygosity Fitness Correlations, or HFCs (DAVID 1998), are more significant for molecular markers derived from genomic regions known to be functional (OLANO-MARIN *et al.* 2011). Taken together, one possibility is that selection against inbreeding will maintain higher genetic diversity, likely through low frequency alleles, at sites linked to the loci carrying deleterious recessive alleles (see Figure 2 for a cartoon of the described process). While the presence of low frequency alleles has been explored as a statistical mechanism that creates variation among estimates of  $F_{ST}$  (BHATIA *et al.* 2013; JAKOBSSON *et al.* 2013), much less effort has been dedicated to understanding what biological processes might be responsible for generating the allele frequencies that “bias” estimates of  $F_{ST}$ . Recent studies in maize (*Zea mays*) by MEZMOUK and ROSS-IBARRA (2013) have suggested that there is a predominance of deleterious recessive alleles in genomic regions with significant hits in genome wide association studies (GWAS). LAZREK *et al.* (2009) suggest that, in *M. truncatula*, differences in the EST and anonymous markers are the result of balancing selection

taking place on the EST markers. However, selection against homozygosity at functional loci would generate a similar pattern of population genetic summary statistics.

Further simulation of these dynamics is required to fully disentangle the role of spatiotemporal processes and inbreeding acting on individual loci (Fields et al. *In Prep*). Additionally, programs that take into account potential differences in molecular marker specific evolutionary processes will aid in disentangling locus and population specific effects (see GESTE v.3, Fields et al. *In Prep*; Appendix 4).

## **ACKNOWLEDGMENTS**

PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT and Janis Antonovics, and NSF-OISE# 1139716 to DT and PF.

## REFERENCES

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010 Permanent genetic resources added to Molecular Ecology resources database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in *Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.
- AUSTERLITZ, F. D. R., G. GLEISER, S. TEIXEIRA and G. BERNASCONI, 2011 The effects of inbreeding, genetic dissimilarity and phenotype on male reproductive success in a dioecious plant. *Proceedings of the Royal Society B: Biological Sciences*.
- BEAUMONT, M. A., and R. A. NICHOLS, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**: 1619-1626.
- BHATIA, G., N. J. PATTERSON, S. SANKARARAMAN and A. L. PRICE, 2013 Estimating and interpreting  $F_{st}$ : the impact of rare variants. *Genome Research*.
- COOP, G., D. WITONSKY, A. DI RIENZO and J. K. PRITCHARD, 2010 Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**: 1411-1423.
- DAVID, P., 1998 Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**: 531-537.
- ELLEGREN, H., 2000 Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**: 551-558.



- ELLEGREN, H., 2004 Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**: 435-445.
- FIELDS, P. D., and D. R. TAYLOR, In Prep. Spatiotemporal determinants of FST in a metapopulation of *Silene latifolia*.
- FOLL, M., and O. E. GAGGIOTTI, 2006 Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875-891.
- FOLL, M., and O. E. GAGGIOTTI, 2008 A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**: 977-993.
- GAGGIOTTI, O. E., D. BEKKEVOLD, H. B. H. JØRGENSEN, M. FOLL, G. R. CARVALHO *et al.*, 2009 Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic Herring as a case study. *Evolution* **63**: 2939-2951.
- GILES, B., and J. GOUDET, 1997 Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist* **149**: 507-526.
- GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LEGER *et al.*, 2011 Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591-611.
- JAKOBSSON, M., M. D. EDGE and N. A. ROSENBERG, 2013 The Relationship Between FST and the Frequency of the Most Frequent Allele. *Genetics* **193**: 515-528.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.

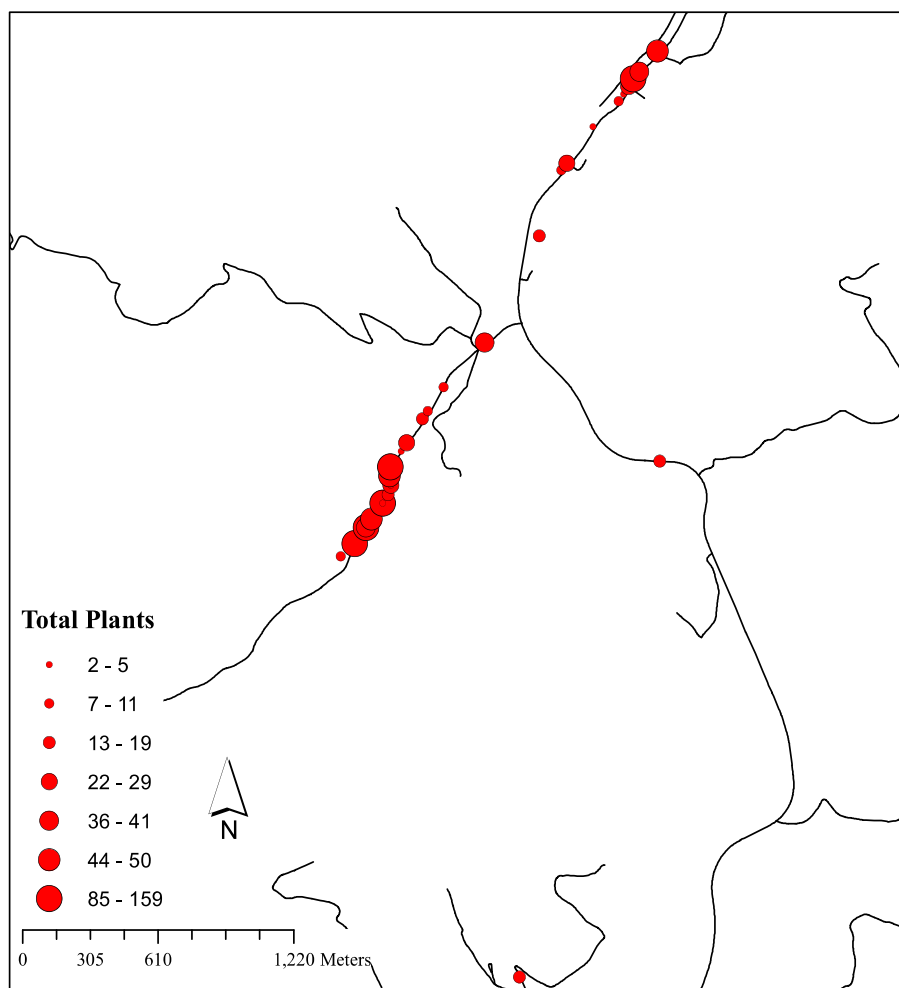
- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- LASCOUX, M., and R. J. PETIT, 2010 The 'New Wave' in plant demographic inference: more loci and more individuals. *Molecular Ecology* **19**: 1075-1078.
- LAZREK, F., V. ROUSSEL, J. RONFORT, G. CARDINET, F. CHARDON *et al.*, 2009 The use of neutral and non-neutral SSRs to analyse the genetic structure of a Tunisian collection of *Medicago truncatula* lines and to reveal associations with eco-environmental variables. *Genetica* **135**: 391-402.
- MARUKI, T., S. KUMAR and Y. KIM, 2012 Purifying Selection Modulates the Estimates of Population Differentiation and Confounds Genome-Wide Comparisons across Single-Nucleotide Polymorphisms. *Molecular Biology and Evolution* **29**: 3617-3623.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MCCAULEY, D. E., 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.

- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MEZMOUK, S., and J. ROSS-IBARRA, 2013 The pattern and distribution of deleterious mutations in maize. arXiv:1308.0380.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- MOILANEN, A., and M. NIEMINEN, 2002 Simple connectivity measures in spatial ecology. *Ecology* **83**: 1131-1145.
- OETJEN, K., and T. B. H. REUSCH, 2007 Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology* **16**: 5156-5167.
- OLANO-MARIN, J., J. C. MUELLER and B. KEMPENAERS, 2011 Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci. *Molecular Ecology* **20**: 4028-4041.
- OSSOWSKI, S., K. SCHNEEBERGER, J. I. LUCAS-LLEDO, N. WARTHMAN, R. M. CLARK *et al.*, 2010 The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science* **327**: 92-94.

- PANNELL, J. R., and P. D. FIELDS, In Press Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist*.
- PAYSEUR, B. A., P. JING and R. J. HAASL, 2011 A Genomic Portrait of Human Microsatellite Variation. *Molecular Biology and Evolution* **28**: 303-312.
- PERONI, P. A., and R. T. ARMSTRONG, 2001 Density, dispersion and population genetics of a *Silene latifolia* seed bank from southwestern Virginia. *Journal of the Torrey Botanical Society* **128**: 400-406.
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.
- RICHARDS, C., S. CHURCH and D. MCCAULEY, 1999 The Influence of Population Size and Isolation on Gene Flow by Pollen in *Silene alba*. *Evolution* **53**: 63-73.
- RICHARDS, C. M., S. N. EMERY and D. E. MCCAULEY, 2003 Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity* **90**: 181-186.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**: 253-262.
- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WHITLOCK, M., 2011  $G'st$  and  $D$  do not replace  $F_{st}$ . *Molecular Ecology* **20**: 1083-1091.

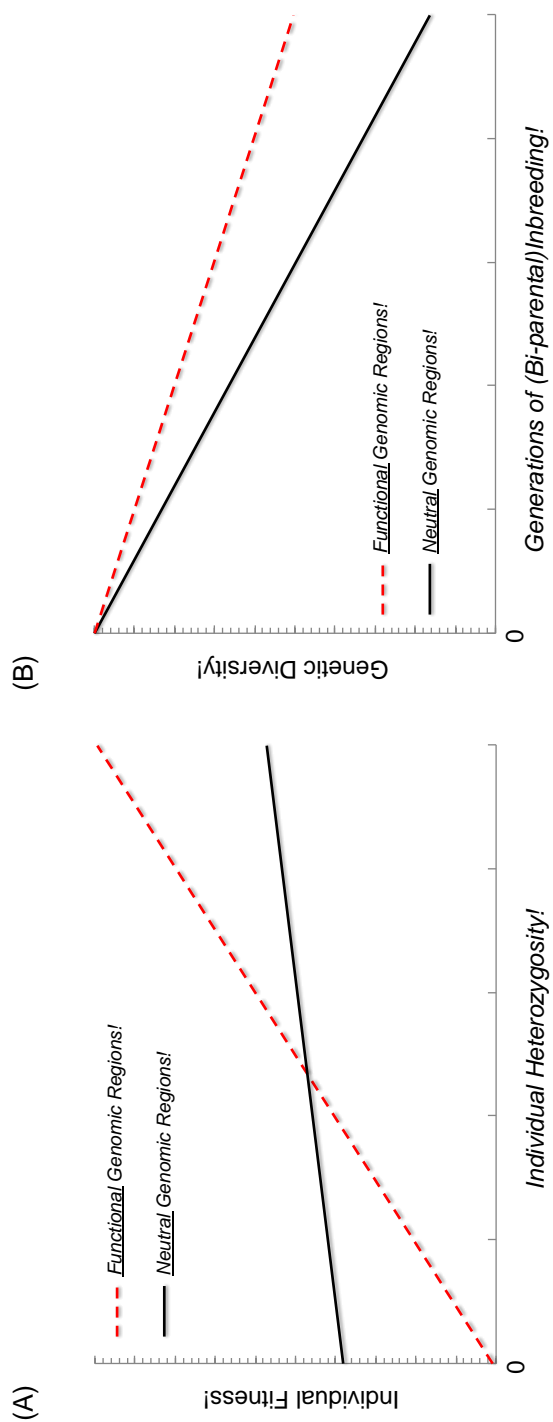
- WHITLOCK, M. C., and D. E. McCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.
- WHITLOCK, M. C., and D. E. McCAULEY, 1999 Indirect measures of gene flow and migration:  $F_{ST}$  not equal  $1/(4Nm+1)$ . *Heredity* **82**: 117-125.
- WHITTAKER, J. C., R. M. HARBORD, N. BOXALL, I. MACKAY, G. DAWSON *et al.*, 2003 Likelihood-based estimation of microsatellite mutation rates. *Genetics* **164**: 781-787.
- WRIGHT, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97-159.

**Figure 1.** Map of the focal populations of the *S. latifolia* metapopulation located in Giles and Craig County, VA, USA utilized in the presented analysis. Circles represent individual populations, where the size of the circle indicates the total number of plants located within our grid. Black lines represent both the assumed grid of population arrangement and small country roads. Due to the topology of the focal area (mountain, valley systems), it was assumed that pollinators move along the linear grid, rather than crossing over ridges.



**Figure 2.** Linkage, selection against inbreeding, and the maintenance genetic diversity. Within the focal *S. latifolia* metapopulation (RICHARDS *et al.* 1999), and more broadly within its species range (AUSTERLITZ *et al.* 2011), male fitness has been shown to be affected by relatedness of the receiving female, implying selection against inbreeding. Further, studies in other systems have indicated that Heterozygosity Fitness Correlations, or HFCs (DAVID 1998), are more significant for markers derived from genomic areas known to be functional. Selection against inbreeding will maintain higher genetic diversity within functional regions. (A) Functional regions are more likely to harbor deleterious alleles, and will maintain greater heterozygosity than neutral genomic regions, and thus (B) over multiple generations of inbreeding, diversity will be lost at different rates in functional and neutral regions.





**Table 1.** Global Summary Statistics. Locus is the name of the molecular marker, which can be either anonymously derived (unknown genomic origin) or EST (derived from a floral transcriptomes in the present case);  $N$  = number of observed alleles in the sample;  $H_0$  = observed genetic diversity;  $H_E$  = expected genetic diversity;  $F_{ST}$  = global estimate of allelic differentiation amongst populations.

<b>Locus</b>	<b><math>N</math></b>	<b><math>H_0</math></b>	<b><math>H_E</math></b>	<b><math>F_{ST}</math></b>
<b><i>Anonymous</i></b>				
<i>slat_18</i> <sup>1</sup>	10	0.338	0.592	0.09
<i>slat_32</i> <sup>1</sup>	9	0.303	0.673	0.135
<i>slat_33</i> <sup>1</sup>	3	0.07	0.13	0.231
<i>slat_48</i> <sup>1</sup>	2	0.078	0.202	0.143
<i>slat_72</i> <sup>1</sup>	12	0.325	0.593	0.055
<i>slat_85</i> <sup>1</sup>	16	0.324	0.538	0.125
<i>SL_8</i> <sup>2</sup>	43	0.693	0.819	0.051
<i>SV_11</i> <sup>3</sup>	11	0.305	0.573	0.122
<b>Overall</b>	<b>13.25</b>	<b>0.304</b>	<b>0.515</b>	<b>0.103</b>
<b><i>EST</i></b>				
<i>SL_SSR04</i> <sup>4</sup>	11	0.61	0.591	0.065
<i>SL_SSR12</i> <sup>4</sup>	20	0.71	0.726	0.057
<i>SL_SSR16</i> <sup>4</sup>	7	0.376	0.38	0.076
<i>SL_SSR29</i> <sup>4</sup>	25	0.832	0.855	0.049
<i>SL_SSR06</i> <sup>4</sup>	18	0.695	0.77	0.038
<i>SL_SSR09</i> <sup>4</sup>	10	0.294	0.299	0.042
<i>SL_SSR20</i> <sup>4</sup>	12	0.115	0.135	0.07
<i>SL_SSR30</i> <sup>4</sup>	21	0.649	0.739	0.068
<b>Overall</b>	<b>15.5</b>	<b>0.535</b>	<b>0.562</b>	<b>0.056</b>

**Table 2.** Analysis with BayeScan to determine significant differences in contribution of  $F_{ST}$ .  $\alpha$  is a regression coefficient indicating the effect a particular marker has on the overall increase/decrease in  $F_{ST}$ ;  $\Pr(\alpha \neq 0)$  is an RJMCMC based estimate determining the posterior probability that a given locus contributes significantly to the decrease or increase in  $F_{ST}$ .  $F_{ST}$ , as estimated through BayeScan, is an  $F$ -model based indication of an individual population's divergence from the metapopulation as a whole.

<b>Locus</b>	<b><math>\alpha</math></b>	<b><math>\Pr(\alpha \neq 0)</math></b>	<b><math>F_{ST}</math></b>
<b><i>Anonymous</i></b>			
<i>slat_18</i>	0.664	<b>0.999</b>	0.103
<i>slat_32</i>	0.903	<b>1</b>	0.126
<i>slat_33</i>	1.48	<b>1</b>	0.2
<i>slat_48</i>	1.06	<b>1</b>	0.146
<i>slat_72</i>	0.484	<b>0.938</b>	0.0884
<i>slat_85</i>	1.31	<b>1</b>	0.174
<i>SL_8</i>	-0.0404	0.126	0.0559
<i>SV_11</i>	0.81	<b>1</b>	0.117
<b><i>EST</i></b>			
<i>SL_SSR04</i>	0.0953	0.1835	0.0625
<i>SL_SSR12</i>	-0.102	0.169	0.534
<i>SL_SSR16</i>	0.566	<b>0.932</b>	0.0956
<i>SL_SSR29</i>	-0.0329	0.118	0.0563
<i>SL_SSR06</i>	0.03	0.162	0.059
<i>SL_SSR09</i>	-0.452	0.592	0.039
<i>SL_SSR20</i>	0.308	0.344	0.0803
<i>SL_SSR30</i>	0.231	0.345	0.0688

**Table 3.** GESTE based model comparison to determine factors contributing to increased  $F_{ST}$  in a metapopulation of *S. latifolia*. For anonymous markers, the full model showed the highest posterior model probability. In contrast, for EST markers, the model containing only population isolation had the highest posterior probability. Finally, the combined marker set showed the highest posterior probability that contained the two parameters that dominated the models of the individual marker sets.

<b>Model</b>	<b>Pr</b>	<b>Factors Included</b>
<b><i>Anonymous Markers</i></b>		
<b>8</b>	<b>0.405</b>	<b>Age, Size, and Population Isolation</b>
7	0.325	Age and Population Isolation
4	0.132	Size and Population Isolation
2	0.0614	Size
<b><i>EST Markers</i></b>		
<b>5</b>	<b>0.57</b>	<b>Population Isolation</b>
7	0.14	Age and Population Isolation
1	0.134	Null
6	0.098	Size and Population Isolation
<b><i>Combined</i></b>		
<b>7</b>	<b>0.525</b>	<b>Age and Population Isolation</b>
6	0.23	Size and Population Isolation
8	0.146	Age, Size, and Population Isolation
2	0.0432	Size

**Table 4.** Regression coefficients, and their confidence intervals, from GESTE

estimated model with highest posterior probabilities for Anonymous, EST, and the full combination of the two, respectively. Anonymous markers show increased  $F_{ST}$  due to increased population isolation, and decrease as a result of population size and isolation. In contrast, only population isolation has a significant effect on  $F_{ST}$  in EST markers. The combined marker set showed a combination of these two effects from individual marker sets.

Regression Coefficient	Factor	Mean	Mode	95% HPDI
<b><i>Anonymous Markers</i></b>				
$\alpha_0$	Constant	-2.26	-2.27	[-2.47; -2.04]
$\alpha_3$	Population Isolation	0.246	0.243	[0.0682; 0.415]
$\alpha_2$	Age	-0.335	-0.326	[-0.553; -0.114]
$\alpha_1$	Size	-0.278	-0.265	[-0.514; -0.0695]
$\sigma^2$	-	0.261	0.229	[0.119; 0.428]
<b><i>EST Markers</i></b>				
$\alpha_0$	Constant	-3.02	-3.01	[-3.33; -2.72]
$\alpha_3$	Population Isolation	0.328	0.336	[0.0844; 0.580]
$\sigma^2$	-	0.561	0.496	[0.252; 0.914]
<b><i>Combined Markers</i></b>				
$\alpha_3$	Constant	-2.62	-2.65	[-2.844; -2.41]
$\alpha_2$	Population Isolation	0.321	0.316	[0.118; 0.506]
$\alpha_1$	Age	-0.352	-0.357	[-0.546; -0.137]
$\sigma^2$	-	0.33	0.303	[0.170; 0.528]

## Chapter 4: Population history and differential consequences of inbreeding and outcrossing in a plant metapopulation <sup>4</sup>

---

<sup>4</sup> Formatted for submission as a coauthored manuscript: Fields, P.D., Gonzalez, G., Kniskern, J., and D. R. Taylor

## ABSTRACT

The phenotypic consequences of inbreeding typically take the form of inbreeding depression, which can be defined as the decline in fitness (or some trait related to fitness) associated with an increase in the inbreeding coefficient,  $F$ . However, the basic assumption of a predictable, inverse relationship between fitness and  $F$  has been questioned by a number of recent empirical studies. This may be especially common in structured populations, where among population variance in the effects of inbreeding may be driven by local demes having experienced different histories of colonization, inbreeding, drift, gene flow and selection. We explored the relationship between population history and inbreeding in a metapopulation of the plant, *Silene latifolia*, for which long-term data are available for the historical size and spatial distribution of hundreds of local demes. We used a population genetic analysis to estimate gene flow and bi-parental inbreeding ( $F_{IS}$ ) in demes that had different histories of spatial isolation. A controlled crossing experiment examined whether the effect of inbreeding and outcrossing on fitness-related traits varied with different histories of population size and isolation. Historically isolated demes experienced less gene flow into populations and an increase in the quantity of bi-parental inbreeding. Individuals from historically small, isolated populations were phenotypically different than individuals from large populations, though they did not necessarily have a lower overall fitness. This phenotypic divergence was genetically based. Individuals from small, isolated populations showed no increase in progeny fitness when outcrossed with individuals from other populations, suggesting that they were not suffering significantly from inbreeding depression. Rather, they exhibited

significant inbreeding advantage and outbreeding depression for traits expressed early in life. The causes of variation in the  $F$ -fitness relationship among populations, or among families within populations, must include some variance in the distribution of recessive, or nearly recessive, mutations, driven by aspects of population history including population size, founder effects, gene flow, bi-parental inbreeding, and other past opportunities for the purging (or fixation) of deleterious recessive alleles. Our findings show that two of these factors, historical variation in population size and isolation, likely contribute substantial variation in past inbreeding and the consequences of future inbreeding across the metapopulation.



## INTRODUCTION

Population structure can alter the outcome of selection in metapopulations. Consider the fate of a deleterious recessive allele. In structured populations, recessive alleles will be more exposed to selection because they experience higher relative frequency and homozygosity in a subset of demes. This results in inbreeding depression over the short term, and perhaps a reduced genetic load (via more efficient purging of deleterious recessives) over the longer term (THRALL *et al.* 1998; WHITLOCK 2002). If population structure is severe, fixation of deleterious recessive alleles may occur, a process that can only be reversed by inter-demic processes such as genetic rescue via migration (WILLI and FISCHER 2005; WILLI *et al.* 2005). More generally, population structure influences the outcome of natural selection whenever individuals interact ecologically or behaviorally with a local subset of conspecifics (McCAULEY 1994).

The phenotypic consequences of inbreeding typically take the form of inbreeding depression, which can be defined as the decline in fitness (or some trait related to fitness) associated with an increase in the inbreeding coefficient,  $F$  (WRIGHT 1932). Inbreeding depression is caused either by the increased homozygosity of deleterious recessive alleles, or by decreasing relative frequency of heterozygotes at overdominant loci. The basic assumption of a predictable, inverse relationship between fitness and  $F$  has been questioned by a number of recent empirical studies. In the plant *Ranunculus reptans* (creeping spearwort), inbred offspring may be equally fit or more fit relative to other individuals (WILLI *et al.* 2005). This study revealed substantial among-population variance in how fitness

declines with increasing  $F$  (WILLI *et al.* 2005). A 5-generation serial inbreeding experiment in the angiosperm *Mimulus guttatus* showed that the relationship between total flower production (an assay for individual relative fitness) and degree of inbreeding varied significantly among populations and among families (DUDASH *et al.* 1997). In this experiment, the extinction of inbred lines that suffered higher levels of inbreeding depression, and the retention of lines that suffered less from inbreeding, meant that the purging on genetic load could be accomplished more readily by selection among lines, rather than selection among individuals within lines. In *Physa acuta*, a freshwater snail, there was significant among-population, among-family and among-population class (river vs. pond) variance in the inbreeding/fitness relationship (ESCOBAR *et al.* 2008).

The causes of variation in the  $F$ -fitness relationship among populations, or among families within populations, must include some variance in the distribution of recessive, or nearly recessive, mutations. In a metapopulation, this will be caused by variance in population age, demographic history, genetic drift, founder effect, historical gene flow, bi-parental inbreeding, and other past opportunities for the purging (or fixation) of deleterious recessive alleles. In a metapopulation, these and related processes can occur at very local scales, and may combine to affect the level of inbreeding, inbreeding depression and opportunities for purging in an otherwise random-mating population (WHITLOCK 2002).

Metapopulation dynamics may also generate local variation in phenotypic evolution that may in turn feed back to affect metapopulation processes such as gene flow, and population size and growth. Using the long-term study of a Finnish

Glanville butterfly (*Melitaea cinxia*) metapopulation, Wheat et al. (2011) showed that females collected from newly colonized sites exhibited higher expression of abdomen genes involved in egg provisioning and thorax genes involved in the maintenance of flight muscle proteins, traits that might affect future gene flow or population size. If populations are segregating combinations of epistatically interacting loci, then local colonization events may contribute to reproductive isolation. Matute (2013) conducted an experimental evolution experiment with *Drosophila yakuba*; one thousand replicate lines were forced through a genetic bottleneck, not unlike what might occur in spatially distributed populations in nature. While the most common outcome of the bottleneck and subsequent inbreeding was extinction, a number of isolates persisted, simultaneously exhibiting premating isolation when crosses were made between other similarly bottlenecked isolates. It is not clear how important these effects might be among newly colonized sites in nature, or the extent that such among population variance may be dissolved through the action of subsequent migration (GUILLAUME and WHITLOCK 2007). Thus, in a metapopulation, local demes can be thought of as having potentially distinct histories of inbreeding, genetic drift, gene flow and adaptive evolution. Just how distinct these histories are, and how long they persist, is an empirical question. Corbett-Detig et al. (2013), using a large-panel of *D. melanogaster* recombinant inbred lines (RILs), describe the genomic footprint of this very sort of epistatic interaction underlying within-species reproductive isolation, and suggest that the requisite genetic variation will likely segregate

contemporaneously in natural populations rather than requiring special conditions implicit to the Dobzhansky–Muller model.

In this paper, we examine how local variation in population history (population size, degree of spatial isolation) can effect the phenotypic consequences of inbreeding. Using the angiosperm *Silene latifolia* Richards (2000) identified recently-established small populations, analogous to the recently bottlenecked lines in Matute (2013). Richards showed that these newly established populations were suffering from inbreeding depression relative to larger more established populations, and found evidence of enhanced gene flow, or genetic rescue, from outside sources. Our focus was to project this process forward, to test whether the contrasting histories of local populations led to different evolutionary outcomes. We identified chronically small, isolated populations and used a fine-scale, molecular population genetic analysis to determine the consequences of population isolation on the quantity of inter-population gene flow and bi-parental inbreeding. We also use a combination of greenhouse experiments and crossing designs to test whether small, isolated populations show different responses to further inbreeding.

## **MATERIALS AND METHODS**

### **STUDY ORGANISM**

*Silene latifolia* is a dioecious perennial plant that was introduced to the United States from Europe. It occurs primarily in the northeast, but also inhabits higher elevations further south. The populations in this study are distributed in patches along the roadsides and farmland of southwestern Virginia, in the vicinity of

Mountain Lake Biological Station. This region has been the focus of a metapopulation census since 1988 (ANTONOVICS *et al.* 1994; RICHARDS 2000). Census records consist of approximate numbers of male and female plants along continuous stretches of roadside.

For our controlled crosses, we confined this study to two types of populations: those that had been isolated for the entire duration of the census and those that have been consistently centrally located. Three small populations were used, ranging from approximately 10 to 30 individuals. Using 10-years of census data and correcting for sex ratio and among year variation in population size, we calculated the average effective population size ( $N_E$ ) to be 8.7 for the small populations. Population sizes from our metapopulation census data are generally underestimates relative to more detailed demographic studies we have done, but the relative size of populations is generally accurate. Two of the small populations were more than 1800m from the nearest neighboring patch of plants. The third small population was 360m from the nearest neighbor. Pollen flow in *S. latifolia* typically does not exceed 100m (RICHARDS *et al.* 1999) and divergence in gene frequency occurs over a distance of approximately 150m or even less (McCAULEY *et al.* 1996). Thus, the small populations in this study were relatively isolated genetically. For large populations, five populations that had well over 100 individuals for the past 10 years were selected ( $\bar{N}_E=105.8$ ). One large population, L5, experienced a demographic bottleneck when numbers were reduced from 200 to 50 individuals. One large population (L1) is not within the metapopulation census, so nothing is known about its specific demographic history, except that it

has been a large field with hundreds of plants since the beginning of the census (1988) (D.R. Taylor, pers. obs.).

## GENOTYPE SAMPLING

In order to determine the proportion of selfing and gene flow across a large proportion of the metapopulation with different levels of spatial isolation, we sampled plants from 77 spatially distinct populations during peak flowering in the summer of 2008, spanning  $\sim 1/3$  of the focal metapopulation (Figure 1). We collected leaf tissue from every plant in the population, or up to 50 individuals in the largest populations, and stored the leaves in silica gel (Sigma). Genomic DNA was extracted and amplified following established microsatellite techniques for *S. latifolia*. DNA was extracted from leaf tissue using the method described in (KELLER *et al.* 2012). We genotyped each individual plant at between 16 and 19 microsatellite loci. Our microsatellites are derived from multiple sources (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; MOCCIA *et al.* 2009; TEIXEIRA and BERNASCONI 2007). PCR amplification was conducted using published methods for each marker. PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems). Alleles were binned using

the software TANDEM (MATSCHINER and SALZBURGER 2009).

## **NATURAL POPULATIONS**

To determine whether large, central and small, isolated populations differed in traits related to fitness, we recorded the number of seeds per capsule, seed mass, and germination percentage. We collected seed capsules from up to ten females in each of the small populations and three of the large populations (Figure 1). For each capsule, we counted the seeds and weighed them together to obtain the mean seed mass. From each capsule, we then planted groups of five seeds in each of five one-inch tubular pots that contained a standard, homogeneous soil mixture. Pots were arranged randomly in racks indoors and were watered daily. We recorded the number of days until the cotyledons emerged. To increase the reliability of our estimate of germination percentage, we also recorded the germination percentage of seeds from these six populations and two additional large populations. After the seeds had aged at least five weeks, we put 100 seeds from each capsule into petri plates lined with filter paper, and recorded the proportion that germinated.

## **CROSSING EXPERIMENTS**

To determine whether differences between small and large populations were genetically based, and to test the predictions of the purging and drift models for small populations, we carried out a series of crosses in the greenhouse. If small populations were suffering from inbreeding, we expect plants from those populations to have lower fitness, with relatively little reduction in fitness with

further inbreeding, but with higher fitness when plants are crossed among populations. If small populations have tended to purge deleterious recessive alleles, we expect no overall reduction in plant fitness, little or no inbreeding depression, and little or no advantage to outcrossing (relative to larger populations).

Three small populations and five large populations were used. We randomly selected a male and a female from up to ten families per population (some small populations had less than ten families). Each female was crossed with three males: 1) the male plant from her family (sib-mating), 2) a randomly selected non-sibling male from within her own population (random mating), and 3) a randomly selected male from another population used in the study (outcrossing). All pollinations were carried out with male and female flowers that had opened in the previous 24 hours.

## **DATA ANALYSIS**

### **POPULATION GENETIC ANALYSIS OF CONNECTIVITY AND BI-PARENTAL INBREEDING**

There are a number of methods described in published literature that might be used to estimate population isolation. The majority of ecological studies have utilized a nearest neighbor/patch approach, or distance to multiple neighbors within a limited neighborhood of a focal patch (or buffer) (MOILANEN and NIEMINEN 2002). However, these simple measures have been shown to be poor predictors of important metapopulation dynamics such as colonization potential (MOILANEN and NIEMINEN 2002). Instead, we used a sum total of all pair-wise distances of a focal population to all other extant populations within an individual metapopulation



section (each section is separated by >> 1 kilometer). (GAGGIOTTI *et al.* 2009; MORA *et al.* 2010). These pair-wise distances were calculated using a network constructed based upon the public roadway system, using ArcGIS (ESRI) Network Analyst tool. Given the mountain-valley geographic topology of the area, this network-based approach is more appropriate than standard Euclidean distances in order to predict likely routes of the predominant pollinators, noctuid moths. As such, larger isolation scores are indicative of a decrease in an individual population's probability of receiving migrants, whether through seeds or pollen.

We calculated the population genetic summaries of genetic diversity, as well as estimates of genetic (sub)-structure via hierarchical *F*-statistics using the software GenoDive version 2.0b21 (MEIRMANS and VAN TIENDEREN 2004), with significant deviations from 0 assessed with 10,000 permutations and  $\alpha = 0.05$ .

To assess how isolation affects migration among populations, we used the program BayesAss v. 3.03 (WILSON and RANNALA 2003). Like above, we analyzed each section separately, as the likelihood of migration from one section to the other is quite low. A total of three runs per section were done, each using 50,000,000 Markov Chain Monte Carlo (MCMC) iterations and a burn-in of 500,000 iterations, and a thinning interval of 100, each with a different starting seed. In order to obtain appropriate mixing conditions, as determined by acceptance rate, in the MCMC chain, we modified the allele frequency, inbreeding coefficient, and migration rate parameters as per the BayesAss v. 3.03 manual suggestion. Chain convergence was assessed using the program Tracer v. 1.5 (RAMBAUT and DRUMMOND 2009).

We tested for significant consequences of isolation on population genetic summaries and gene flow using Gaussian linear models. Linear modeling of the effects of isolation on these population genetic parameters were performed in R v. 2.15.3.(R DEVELOPMENT CORE TEAM 2012).

## **NATURAL POPULATIONS**

Each trait was analyzed to determine whether mean values differed significantly for large and small populations in nature. Seed counts were normalized with a square-root transformation, and germination percentage was normalized with an arcsine transformation. These two traits were analyzed by ANOVA with the populations treated as random effects and nested within the size class of the population. The data on mean seed mass (they were distinctly bi-modal) so they were analyzed using Wilcoxon Rank-Sum tests.

## **CROSSING EXPERIMENTS**

To determine whether there was a genetic basis to the differences between population size classes in the field, we used ANOVA to compare the trait means in the offspring of plants from large and small populations. The offspring were derived from our crossing scheme. Separate analyses were used for random within-population crosses, among population crosses and sib-matings. The ANOVA was a mixed model with population size class (large versus small) as a fixed effect and population (nested within size class) as a random effect. To meet the assumptions of ANOVA, seed count was square root transformed. Seed mass was log-

transformed. The response variable for germination was the generalized logit of the ratio of the number of seeds that did not germinate to the number of seeds that did germinate.

When evaluating the fitness advantages/disadvantages associated with inbreeding and outcrossing, we made certain assumptions about how each trait was related to fitness. Fewer days to germination, and faster growth were assumed to positively affect fitness. (Relative fitness from inbreeding was calculated as  $1 - (\text{value from sib cross} / \text{value from random within population cross})$ ). Relative fitness from outbreeding was calculated as  $1 - (\text{value from outcross} / \text{value from random within population cross})$ . All other traits were assumed to be positively correlated with fitness, where relative fitness from inbreeding =  $(\text{value from sib cross} / \text{value from random within population cross}) - 1$  and relative fitness from outbreeding =  $(\text{value from outcross} / \text{value from random within population cross}) - 1$ . One-way ANOVA was used to determine the significance of the differences between the two relevant cross types on the trait.

## RESULTS

### GENETIC ANALYSIS OF CONNECTIVITY AND BI-PARENTAL INBREEDING

Individual microsatellites varied in the quantity of population (sub)structure, though only two markers in two of the three sections showed non-significant  $F_{ST}$  values. The global  $F_{IS}$  values for the three separate sections of the metapopulation were 0.229, 0.272, 0.317, respectively, with a range of -0.008 and 0.64 for individual markers across metapopulation sections (Table 1; Suppl. Table 1). Populations

varied in their level of isolation. There was a significant level of gene flow among populations with a global average of ~28% individuals within populations being a recent migrant (Table 1; Suppl. Table 1).

Linear models were used to test the effect of population isolation on resident proportion and on multi-locus  $F_{IS}$  (in this case, an approximate estimate of bi-parental inbreeding), while controlling for population size as part of the model. The log of sum isolation had a highly significant effect on resident proportion (P-value < 0.001, adj.  $R^2$  = 0.4904) and multi-locus  $F_{IS}$  (P-value = 0.0114, adj.  $R^2$  0.1083) (Figure 3).

## NATURAL POPULATIONS

Measurements on plants in the field detected significant differences between large and small populations in the means of three traits (Table 2). Plants in large populations produced more seeds per capsule and larger seeds than plants in small populations.

## GENETICS OF FIELD POPULATIONS

The results from our hand-pollinations showed that two differences in fitness between large and small populations were maintained in within-population crosses in the greenhouse, indicating a genetic basis to these traits (Table 3). Seed number and germination showed no consistent variation among population size classes.

There was significant inbreeding advantage for percent germination, and outbreeding depression for seed number, but only in small populations (Figure 3).

The progeny of sib-matings had a higher germination percentage than the progeny of within population crosses, and between population crosses produced a lower seed set than within population crosses (Figure 4). In large populations, there was no inbreeding depression and some evidence of a small outbreeding advantage. In contrast to small populations, the progeny of between population crosses had a higher germination percentage than the progeny of within population crosses (Figure 4).

## DISCUSSION

Our population genetic analysis suggests that populations vary significantly in the amount isolation they are experiencing with spatial isolation having substantive effects on the total amount of gene flow a population is likely to receive, as well as the subsequent relatedness of individuals available to breed with. In our study, the most surprising result was from the phenotypic and crossing study, where there was significant outbreeding depression that was restricted to small, isolated populations.

Outbreeding depression is believed to be common in angiosperms (FRANKHAM 1995; WASER 1993) and is likely caused by the development of epistatic interactions within populations (FENSTER *et al.* 1997). The finding that smaller, isolated populations exhibited more outbreeding depression, therefore supports the prediction that founder effects will, in some cases, generate large shifts in allele frequencies (or gene combinations in the case of epistasis), potentially generating genetic isolation between these and larger, central populations. In particular, the

colonization process will aid in the generation of these complexes, consistent Wright's shifting balances process (WADE 2013; WRIGHT 1931; WRIGHT 1932; WRIGHT 1969; WRIGHT 1977)

Population isolation has been hypothesized to increase inbreeding depression in the short-term, but given opportunities for subsequent purging, the long-term consequences of inbreeding become less predictable (IVES and WHITLOCK 2002; WHITLOCK 2002). Depending upon the magnitude of isolation and complex interactions with effective population size, we can expect an increase in variance in the consequences of inbreeding, and mean population fitness in general.

Metapopulation dynamics have been theorized to have substantive effects on the structuring of allelic variation (WADE and MCCAULEY 1988; WHITLOCK and MCCAULEY 1990). Population structure, as measured by  $F_{ST}$ , has been theorized to have a significant effect on total genetic load (WHITLOCK 2002). Therefore, we can expect factors affecting  $F_{ST}$  to also affect the consequences inbreeding.

Other studies have found individuals in small, isolated populations to have lower values of some fitness-related traits than individuals in large, central populations (HESCHEL and PAIGE 1995; MENGES 1991), consistent with the results for seed count and seed mass in the present study. However, we also found that seedlings from small populations emerged faster than those from large populations. Both emergence time and seed mass have been shown to positively affect fitness in other plant species (HOWELL 1981; KALISZ 1986; MAZER and SCHICK 1991; STANTON 1985; STANTON 1984; WULFF 1986). These data therefore, do not necessarily suggest that plants in isolated populations (or their progeny) are genetically inferior to

plants in large, central populations. Plants from smaller populations appear to have a different life history, involving more rapid juvenile growth. When the descendants of these field-collected seeds were grown under controlled environmental conditions, plants from small, isolated populations also exhibited smaller seeds that emerged and grew faster, suggesting that the difference between large, central and small, isolated populations are genetically based.

In our study, the results from small populations do not conform precisely to any of the predictions generated by simple models of inbreeding depression and the purging of deleterious recessive alleles. Small, isolated populations had lower values for some traits related to fitness, as would be expected for populations experiencing inbreeding depression. However, three results from the crossing experiments contradict this interpretation: 1) progeny from among population crosses had lower seed production than progeny from within population crosses, 2) sib-mated progeny had a higher germination percentage than within population crosses, and 3) the only evidence of increased fitness of outcrossed progeny was in large populations, where outcrossing advantage was expected to be less pronounced. These results are reminiscent of a recent study using the perennial rosemary scrub (*Hypericum cumulicola*) where population size, age, and isolation had a significant effect on determining the fitness consequences of certain crosses (OAKLEY and WINN 2012). Oakley and Winn (2012) combined molecular marker-based estimators of migration and estimates of relative effective size for 16 natural populations of *H. cumulicola*. Fitness assays revealed that outcrossing advantage was significantly greater for effectively small populations relative to large ones,

though there was an indication of outbreeding depression when comparing self-crosses to random-within population crosses.

The finding that outbreeding depression was significant only in small, isolated populations has several potential explanations. First, outbreeding depression can be the result of local adaptation (FENSTER and GALLOWAY 2000) and it is possible that local adaptation occurred more rapidly in isolated populations that are less likely to be inundated by gene flow from plants adapted to different environments (ANTONOVICS 1968; ELLSTRAND and ELAM 1993). Second some beneficial alleles may be partially recessive, and exposed to positive selection in smaller populations because those populations experience more inbreeding (CHARLESWORTH 1992). Finally, if there is significant epistatic variation segregating in larger populations, then genetic bottlenecks may fix certain combinations in local demes, with deleterious fitness consequences result from breaking up those associations during outcrossing (CORBETT-DETIG *et al.* 2013)

The present results confirm that plants from large, central and small, isolated populations have differences in many traits related to fitness and that these differences have a genetic basis. Thus, the metapopulation can be viewed as a heterogeneous landscape, with local variation in evolutionary outcomes. Importantly, these different evolutionary trajectories may influence the metapopulation ecology and vice versa, most obviously by influencing gene flow and population persistence (INGVARSSON 2001; NEWMAN and PILSON 1997; SACCHERI *et al.* 1998). The connection between local evolutionary processes, and their effects on population growth and persistence, could also present a unique opportunity for



studying the importance of inter-family or inter-demic selection. From an applied perspective, our findings highlight the fact that inbreeding depression from deleterious recessive alleles is not the sole genetic mechanism affecting the long-term persistence of natural populations; genetic bottlenecks may have very different consequences across populations and through time

### **ACKNOWLEDGEMENTS**

PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT and Janis Antonovics, and NSF-OISE# 1139716 to DT and PF.

## LITERATURE CITED

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010 Permanent genetic resources added to Molecular Ecology resources database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ANTONOVICS, J., 1968 Evolution in closely adjacent plant populations. V. Evolution of self-fertility. *Heredity* **23**: 219-238.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in *Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.
- CHARLESWORTH, B., 1992 Evolutionary rates in partially self-fertilizing species. *American Naturalist* **140**: 126-148.
- CORBETT-DETIG, R. B., J. ZHOU, A. G. CLARK, D. L. HARTL and J. F. AYROLES, 2013 Genetic incompatibilities are widespread within species. *Nature* **advance online publication**.
- DUDASH, M., D. CARR and C. FENSTER, 1997 Five generations of enforced selfing and outcrossing in *Mimulus guttatus*: Inbreeding depression variation at the population and family level. *Evolution* **51**: 54-65.
- ELLSTRAND, N. C., and D. R. ELAM, 1993 Population genetic consequences of small population-size - implications for plant conservation. *Annual Review of Ecology and Systematics* **24**: 217-242.

- ESCOBAR, J. S., A. NICOT and P. DAVID, 2008 The Different Sources of Variation in Inbreeding Depression, Heterosis and Outbreeding Depression in a Metapopulation of *Physa acuta*. *Genetics* **180**: 1593-1608.
- FENSTER, C., L. GALLOWAY and L. CHAO, 1997 Epistasis and its consequences for the evolution of natural populations. *Trends Ecol Evol* **12**: 282-286.
- FENSTER, C. B., and L. F. GALLOWAY, 2000 Inbreeding and Outbreeding Depression in Natural Populations of *Chamaecrista fasciculata* (Fabaceae) Depresión por Endogamia y Exogamia en Poblaciones Naturales de *Chamaecrista fasciculata* ( Fabaceae). *Conservation Biology* **14**: 1406-1412.
- FRANKHAM, R., 1995 Conservation genetics. *Annual Review of Genetics* **29**: 305-327.
- GAGGIOTTI, O. E., D. BEKKEVOLD, H. B. H. JØRGENSEN, M. FOLL, G. R. CARVALHO *et al.*, 2009 Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic Herring as a case study. *Evolution* **63**: 2939-2951.
- GUILLAUME, F., and M. C. WHITLOCK, 2007 Effects of migration on the genetic covariance matrix. *Evolution* **61**: 2398-2409.
- HESCHEL, M. S., and K. N. PAIGE, 1995 Inbreeding depression, environmental stress, and population size variation in scarlet gilia (*Ipomopsis aggregata*). *Conservation Biology* **9**: 126-133.
- HOWELL, N., 1981 The effect of seed size and relative emergence time on fitness in a natural population of *Impatiens capensis* Meerb (Balsaminaceae). *Am Midl Nat* **105**: 312-320.

- INGVARSSON, P., 2001 Restoration of genetic variation lost - The genetic rescue hypothesis. *Trends Ecol Evol* **16**: 62-63.
- IVES, A., and M. WHITLOCK, 2002 Inbreeding and metapopulations. *Science* **295**: 454-455.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.
- KALISZ, S., 1986 Variable selection on the timing of germination in *Collinsia verna* (Scrophulariaceae). *Evolution* **40**: 479-491.
- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MATUTE, D. R., 2013 The role of founder effects on the evolution of reproductive isolation. *Journal of Evolutionary Biology* **26**: 2299-2311.
- MAZER, S. J., and C. T. SCHICK, 1991 Constancy of population parameters for life-history and floral traits in *Raphanus sativus* L. II. Effects of planting density on phenotype and heritability estimates. *Evolution* **45**: 1888-1907.
- MCCAULEY, D. E., 1994 Intrademic group selection imposed by a parasitoid-host interaction. *American Naturalist*: 1-13.

- MCCAULEY, D. E., J. STEVENS, P. PERONI and J. RAVEILL, 1996 The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* **83**: 727-731.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MENGES, E. S., 1991 Seed germination percentage increases with population size in a fragmented prairie species. *Conservation Biology* **5**: 158-164.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- MOILANEN, A., and M. NIEMINEN, 2002 Simple connectivity measures in spatial ecology. *Ecology* **83**: 1131-1145.
- MORA, M. S., F. J. MAPELLI, O. E. GAGGIOTTI, M. J. KITTLEIN and E. P. LESSA, 2010 Dispersal and population structure at different spatial scales in the subterranean rodent *Ctenomys australis*. *Bmc Genetics* **11**.
- NEWMAN, D., and D. PILSON, 1997 Increased probability of extinction due to decreased genetic effective population size: Experimental populations of *Clarkia pulchella*. *Evolution* **51**: 354-362.
- OAKLEY, C. G., and A. A. WINN, 2012 Effects of population size and isolation on heterosis, mean fitness, and inbreeding depression in a perennial plant. *New Phytologist* **196**: 261-270.

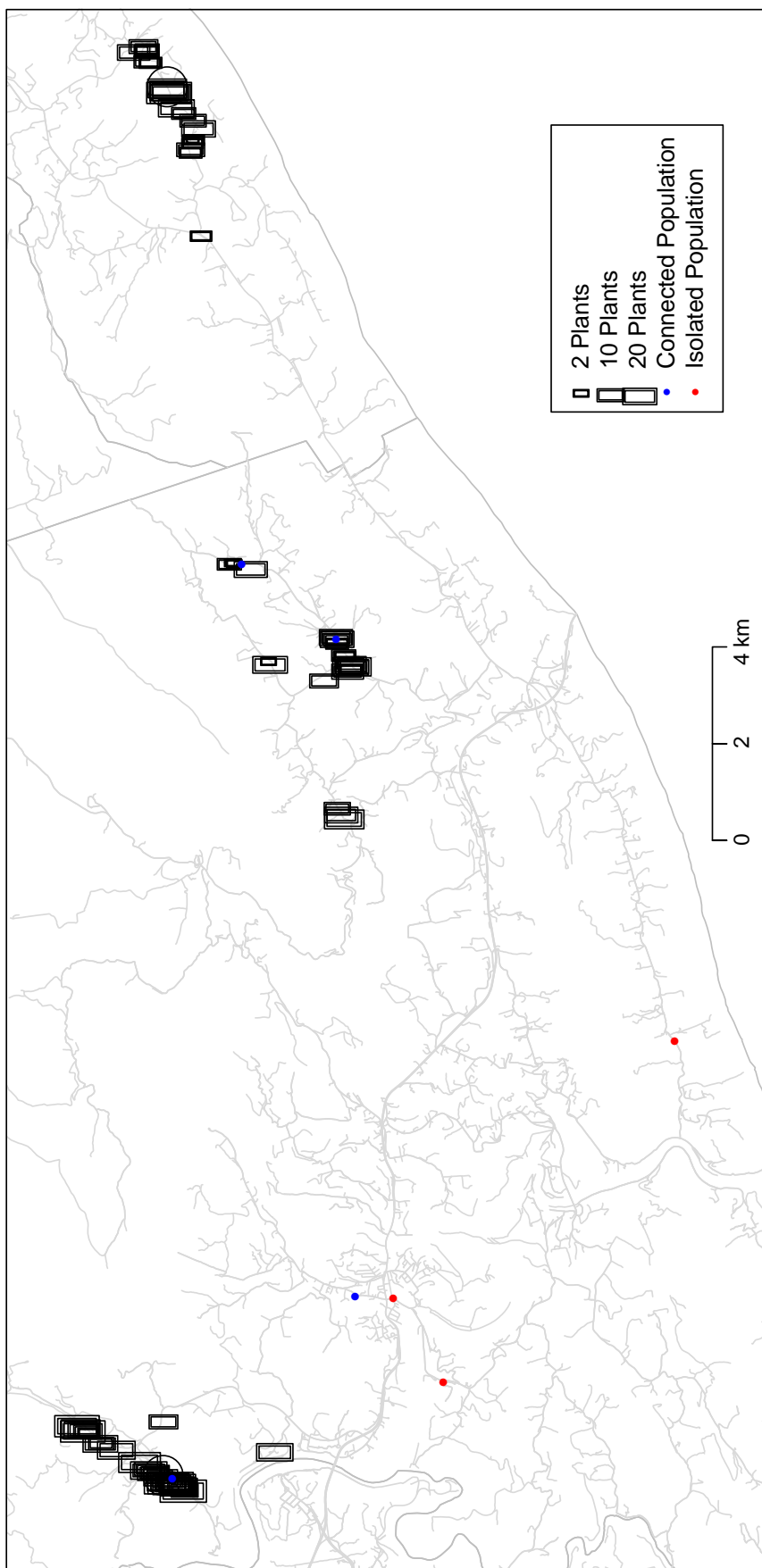
- R DEVELOPMENT CORE TEAM, 2012 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.
- RAMBAUT, A., and A. J. DRUMMOND, 2009 Tracer v1.5 [Internet]. **Available from: URL** <http://beast.bio.ed.ac.uk/Tracer>.
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.
- RICHARDS, C., S. CHURCH and D. MCCAULEY, 1999 The Influence of Population Size and Isolation on Gene Flow by Pollen in *Silene alba*. *Evolution* **53**: 63-73.
- SACCHERI, I., M. KUUSSAARI, M. KANKARE, P. VIKMAN, W. FORTELIUS *et al.*, 1998 Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491-494.
- STANTON, M., 1985 Seed size and emergence time within a stand of wild radish (*Raphanus raphanistrum* L.): The establishment of a fitness hierarchy. *Oecologia* **67**: 524-531.
- STANTON, M. L., 1984 Seed variation in wild radish: Effect of seed size on components of seedling and adult fitness. *Ecology* **65**: 1105-1112.
- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.
- THRALL, P., C. RICHARDS, D. E. MCCAULEY and J. ANTONOVICS, 1998 Metapopulation collapse: the consequences of limited gene flow in spatially structured populations, pp. 83-104 in *Modeling spatiotemporal dynamics in ecology*, edited by J. BASCOMPTE and R. V. SOLE. Springer-Verlag Berlin.

- WADE, M. J., 2013 Phase III of Wright's shifting balance process and the variance among demes in migration rate. *Evolution* **67**: 1591-1597.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WASER, N. M., 1993 Population structure, optimal outbreeding, and assortative mating in angiosperms in *The natural history of inbreeding and outbreeding, theoretical and empirical perspectives*, edited by N. W. THORNHILL. Chicago, USA, Univ. of Chicago Press.
- WHEAT, C. W., H. W. FESCEMYER, J. KVIST, E. TAS, J. C. VERA *et al.*, 2011 Functional genomics of life history variation in a butterfly metapopulation. *Molecular Ecology* **20**: 1813-1828.
- WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191-1202.
- WHITLOCK, M. C., and D. E. MCCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.
- WILLI, Y., and M. FISCHER, 2005 Genetic rescue in interconnected populations of small and large size of the self-incompatible *Ranunculus reptans*. *Heredity* **95**: 437-443.
- WILLI, Y., J. VAN BUSKIRK and M. FISCHER, 2005 A threefold genetic allee effect: population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. *Genetics* **169**: 2255-2265.

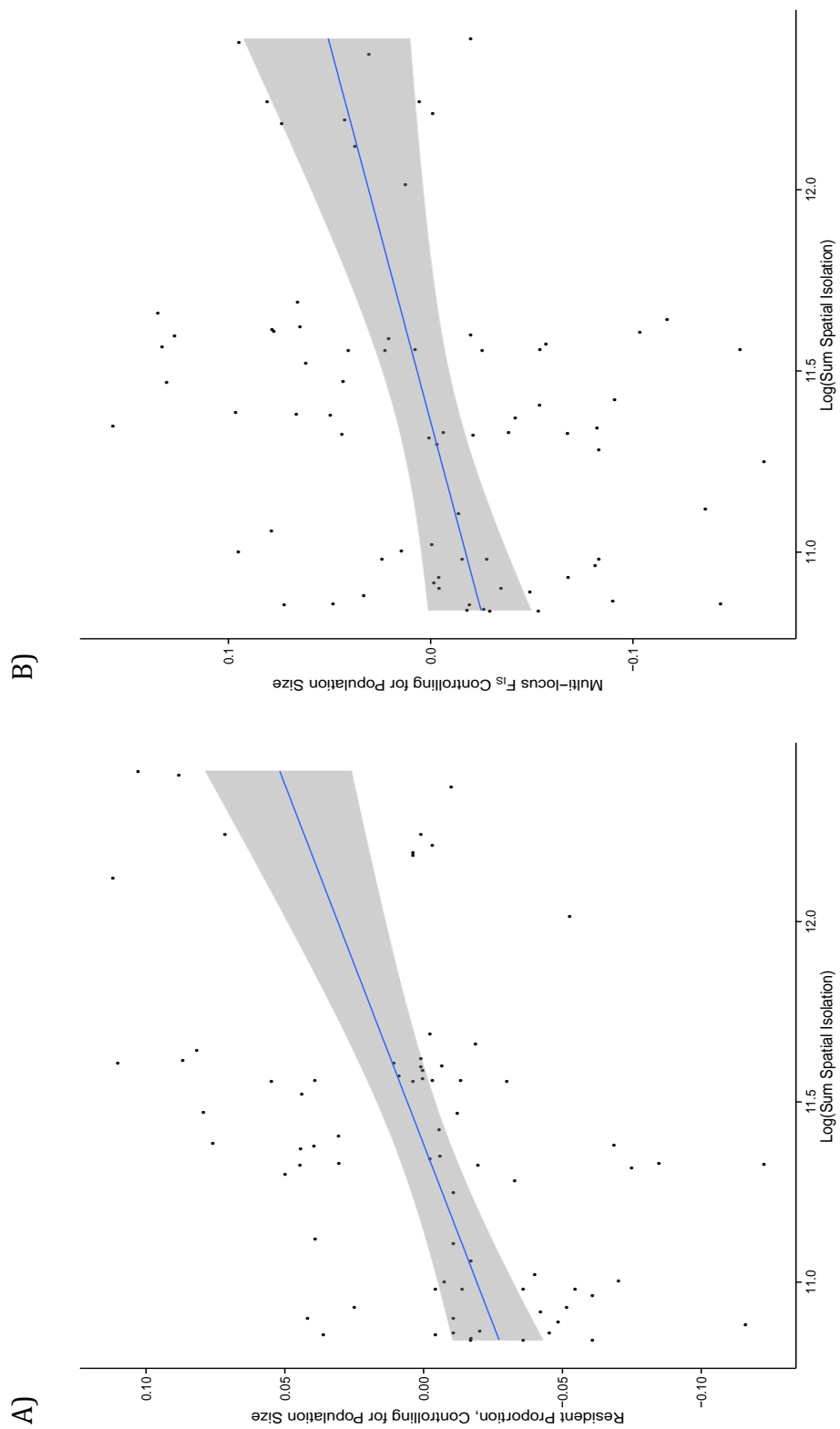
- WILSON, G. A., and B. RANNALA, 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177-1191.
- WRIGHT, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97-159.
- WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the 6th International Congress of Genetics* **1**: 356-366.
- WRIGHT, S., 1969 *Evolution and the genetics of populations. Vol 2*. University of Chicago Press, Chicago, IL.
- WRIGHT, S., 1977 *Evolution and the genetics of populations. Vol. 3*. University of Chicago Press, Chicago, IL.
- WULFF, R. D., 1986 Seed size variation in *Desmodium paniculatum*. III. Effects on reproductive yield and competitive ability. *J Ecology* **74**: 115-121.



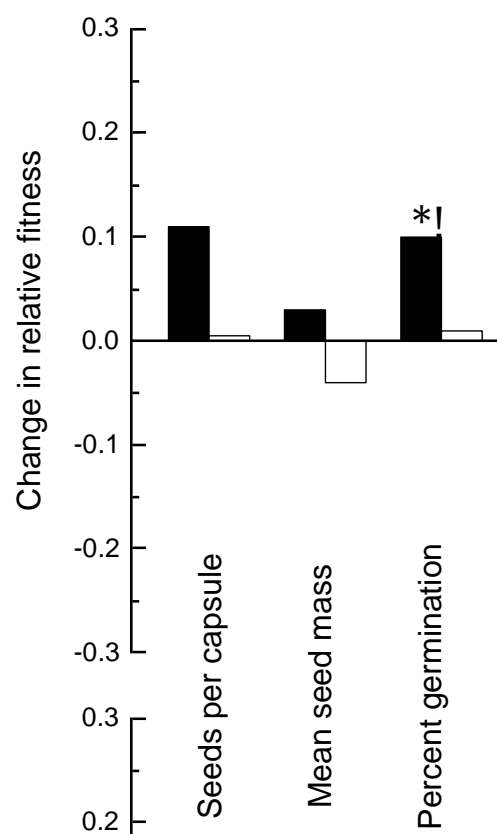
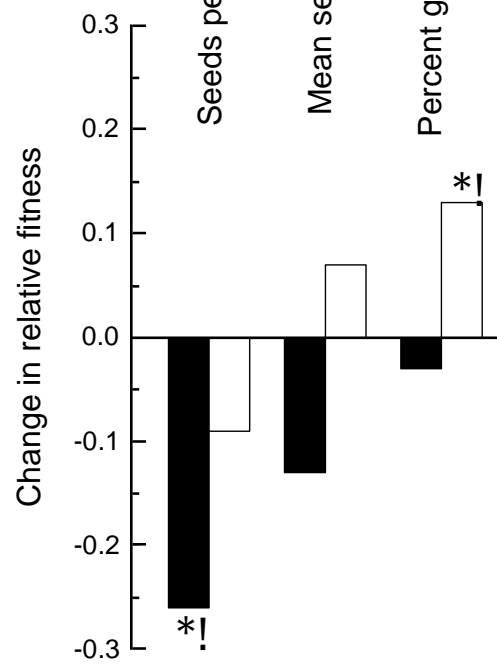
**Figure 1.** Map showing sampling locations for crossing design and population genetic analysis. Note, only populations used as part of the analysis are mapped. Other populations of *S. latifolia* exist in the sampled area.



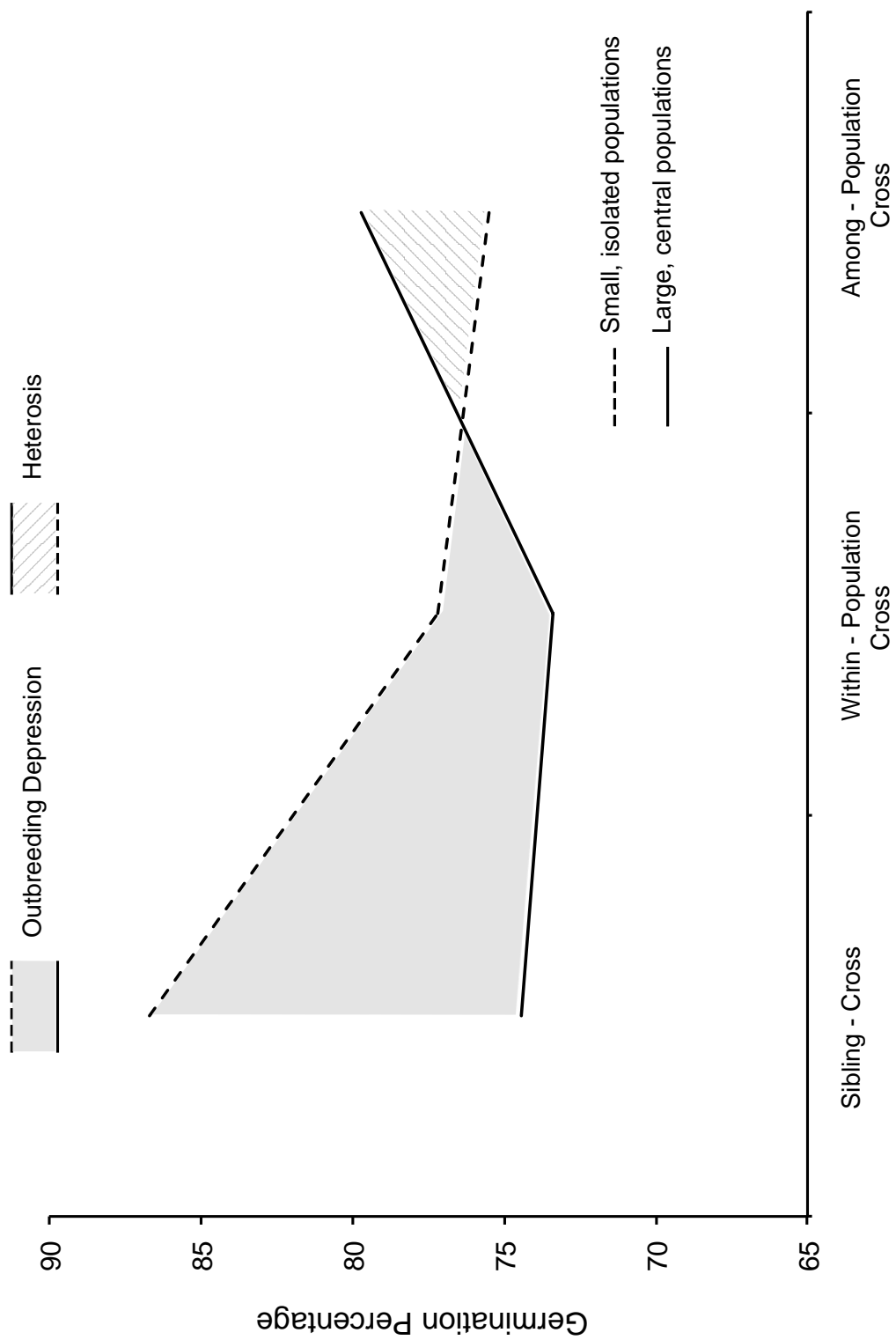
**Figure 2.** Consequences of spatial isolation on (A) resident proportion and (B) multi-locus  $F_{IS}$ . The log of sum isolation had a significant effect on multi-locus  $F_{IS}$  (P-value = 0.0114, adj.  $R^2$  0.1083) and resident proportion (P-value < 0.001, adj.  $R^2$  = 0.4904).



**Figure 3.** Differences in response to inbreeding and outcrossing for large (A) and small (B) populations. The horizontal baseline represents the average for each trait from the within population crosses. Open bars show the performance of progeny from crosses among full sibs relative to the within-population crosses; open bars above zero reflect an inbreeding advantage and open bars below zero reflect inbreeding depression. Solid bars show the performance of progeny from crosses between populations relative to the within-population crosses; Solid bars above zero indicate outcrossing advantage and solid bars below zero indicate outcrossing depression. Asterisks indicate that the two cross types used in that score were significantly different from each other ( $p < 0.05$ ).

**A****B**

**Figure 4.** Percentage germination rate among sites within a cross type. Solid lines represent focal crosses from large, central populations; dashed lines represent crosses from small, isolated populations. In this representation, differences between types of crosses are indicated by the slope of the line, and asymmetries between types of populations results for each cross by different shadings.





**Table 1.** Global population genetic summaries. Variables are the number of alleles ( $N$ ), the effective number of alleles ( $E_N$ ), observed heterozygosity ( $H_O$ ), Expected heterozygosity ( $H_E$ ), and the inbreeding coefficient ( $F_{IS}$ ), global among-population allelic variation ( $F_{IS}$ ), and Jost’s measure of population differentiation ( $D_{EST}$ ).

<i>Locus</i>	Section 2							Section 6							Section 9						
	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>
SL_eSSR01 <sup>1</sup>	-	-	-	-	-	-	-	7	1.818	0.224	0.488	0.542	<b>0.114</b>	0.129	6	1.722	0.161	0.452	0.644	<b>0.099</b>	0.094
SL_eSSR04 <sup>1</sup>	10	2.282	0.577	0.584	0.012	<b>0.05</b>	0.076	7	2.399	0.524	0.62	0.154	<b>0.05</b>	0.089	6	2.475	0.519	0.629	0.174	<b>0.105</b>	0.206
SL_eSSR06 <sup>1</sup>	18	3.762	0.699	0.765	0.086	<b>0.041</b>	0.145	14	3.333	0.503	0.751	0.33	<b>0.069</b>	0.233	13	3.108	0.39	0.725	0.462	<b>0.1</b>	0.304
SL_eSSR09 <sup>1</sup>	12	1.407	0.287	0.301	0.049	<b>0.037</b>	0.017	7	1.65	0.393	0.416	0.055	<b>0.073</b>	0.059	9	1.584	0.382	0.386	0.01	<b>0.091</b>	0.065
SL_eSSR12 <sup>1</sup>	20	3.296	0.724	0.724	-0.001	<b>0.044</b>	0.123	14	3.557	0.674	0.762	0.116	<b>0.117</b>	0.444	15	4.093	0.702	0.795	0.117	<b>0.087</b>	0.381
SL_eSSR16 <sup>1</sup>	8	1.569	0.334	0.378	0.117	<b>0.099</b>	0.068	4	2.245	0.583	0.584	0.002	<b>0.073</b>	0.116	8	2.248	0.537	0.583	0.079	<b>0.141</b>	0.238
SL_eSSR17 <sup>1</sup>	-	-	-	-	-	-	-	13	2.747	0.608	0.674	0.098	<b>0.106</b>	0.256	14	2.66	0.556	0.658	0.154	<b>0.069</b>	0.147
SL_eSSR20 <sup>1</sup>	11	1.161	0.121	0.145	0.168	<b>0.065</b>	0.012	3	1.057	0.056	0.057	0.016	0.028	0.002	3	1.048	0.038	0.049	0.218	0.039	0.002
SL_eSSR22 <sup>1</sup>	-	-	-	-	-	-	-	10	2.817	0.561	0.686	0.183	<b>0.149</b>	0.403	11	3.167	0.501	0.726	0.31	<b>0.084</b>	0.253
SL_eSSR27 <sup>1</sup>	-	-	-	-	-	-	-	9	2.462	0.58	0.628	0.077	<b>0.167</b>	0.355	11	2.568	0.595	0.641	0.071	<b>0.128</b>	0.271
SL_eSSR28 <sup>1</sup>	-	-	-	-	-	-	-	6	1.79	0.469	0.465	-0.008	<b>0.22</b>	0.257	6	1.353	0.24	0.275	0.126	<b>0.245</b>	0.127
SL_eSSR29 <sup>1</sup>	25	5.488	0.835	0.85	0.018	<b>0.054</b>	0.332	19	3.871	0.715	0.785	0.09	<b>0.103</b>	0.441	18	3.683	0.7	0.765	0.085	<b>0.083</b>	0.305
SL_eSSR30 <sup>1</sup>	21	3.407	0.65	0.737	0.119	<b>0.056</b>	0.17	21	2.683	0.345	0.679	0.492	<b>0.111</b>	0.278	18	2.705	0.281	0.678	0.585	<b>0.083</b>	0.197
slat_18 <sup>2</sup>	10	2.837	0.418	0.683	0.388	<b>0.058</b>	0.137	9	2.045	0.234	0.556	0.578	<b>0.208</b>	0.346	11	2.522	0.276	0.649	0.575	<b>0.119</b>	0.259
slat_32 <sup>2</sup>	9	2.709	0.29	0.67	0.568	<b>0.137</b>	0.331	7	2.271	0.364	0.602	0.396	<b>0.159</b>	0.301	9	2.384	0.34	0.62	0.451	<b>0.181</b>	0.375
slat_33 <sup>2</sup>	3	1.128	0.061	0.12	0.49	<b>0.237</b>	0.044	2	1.515	0.184	0.368	0.502	<b>0.265</b>	0.221	3	1.211	0.105	0.186	0.434	<b>0.597</b>	0.35
slat_48 <sup>2</sup>	6	1.253	0.087	0.214	0.593	<b>0.099</b>	0.031	5	1.766	0.194	0.472	0.589	<b>0.108</b>	0.113	5	1.666	0.233	0.427	0.455	<b>0.139</b>	0.125
slat_72 <sup>2</sup>	12	2.858	0.349	0.688	0.493	<b>0.048</b>	0.116	12	2.545	0.326	0.657	0.504	<b>0.097</b>	0.217	16	2.86	0.282	0.7	0.596	<b>0.171</b>	0.499
slat_85 <sup>2</sup>	15	2.085	0.311	0.55	0.434	<b>0.101</b>	0.142	16	1.829	0.284	0.489	0.418	<b>0.156</b>	0.185	16	1.618	0.232	0.408	0.432	<b>0.194</b>	0.172
SI_8 <sup>3</sup>	11	2.153	0.324	0.566	0.428	<b>0.132</b>	0.203	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A11 <sup>4</sup>	43	4.382	0.707	0.806	0.122	<b>0.062</b>	0.283	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Overall	14.625	2.611	0.423	0.549	0.229	<b>0.074</b>	0.099	9.737	2.337	0.412	0.565	0.272	<b>0.128</b>	0.2	10.421	2.351	0.372	0.545	0.317	<b>0.138</b>	0.198

**Table 2.** Means (and standard deviations) of seed and seedling traits of seeds collected from central and small, isolated populations in the field. Seed mass is the total weight of 50 seeds from a single capsule. P-values represent the significance of the difference in mean between large and small populations (see text for details).

Population Size	Seed Number	Seed Mass (g)	Germination Percentage
Large	277 (-28)	0.054 (-0.005)	69 (-15)
Small	134 (-64)	0.04 (-0.01)	71 (-15)
<b>P-values</b>	<b>&lt;0.0001</b>	<b>0.0041</b>	0.2749

**Table 3.** Means (and standard deviations) of seed, seedling, and vegetative traits of progeny from crosses in the greenhouse. Seed mass is the total weight of 50 seeds from a single capsule. Means for each population and P-values are provided in the following order: random within population cross, sib mating, and outcross. P-values represent the significance of the difference in mean between large and small populations for the separate cross types.

Population Size	Seed Number	Seed Mass (g)	Germination Percentage
<i>a) Sib-matings</i>			
Large populations	218 (-96)	0.0419 (-0.0101)	74.43 (-16.78)
Small populations	188 (-63)	0.0404 (-0.0085)	86.7 (-7.52)
<b>P-values</b>	<b>0.1355</b>	<b>0.3807</b>	<b>0.011</b>
<i>b) Within-population crosses</i>			
Large populations	218 (-105)	0.0437 (-0.0109)	73.39 (-17.4)
Small populations	206 (-72)	0.0377 (-0.0082)	77.17 (-14.05)
<b>P-values</b>	<b>0.8246</b>	<b>0.0296</b>	<b>0.4603</b>
<i>c) Among-population crosses</i>			
Large populations	240 (-92)	0.0444 (-0.01)	79.7 (-15.3)
Small populations	153 (-65)	0.0323 (-0.0125)	75.5 (-22.3)
<b>P-values</b>	<b>0.0011</b>	<b>0.0008</b>	<b>0.9029</b>

**Suppl. Table 1.** Individual population summaries of genetic diversity, recipient gene flow, and total isolation score. Variables are the number of alleles ( $N$ ), the effective number of alleles ( $E_N$ ), observed heterozygosity ( $H_O$ ), Expected heterozygosity ( $H_E$ ), and the inbreeding coefficient ( $F_{IS}$ ). Resident proportion is a BayesAss based estimate of the proportion of a population that is non-migrant. Isolation is the sum total distance (m) from a source population to all other extant populations within a focal metapopulation section.

<i>Population ID</i>	<i>Sample Size</i>	$N$	$E_N$	$H_O$	$H_E$	$H_T$	$F_{IS}$	<i>Proportion Resident</i>	<i>Isolation</i>
2_1	24	6.125	3.151	0.411	0.619	0.619	0.335	0.80	2.08E+05
2_2	12	3.062	2.314	0.521	0.472	0.472	-0.103	0.67	8.08E+04
2_3	10	4.5	2.739	0.362	0.558	0.558	0.35	0.67	6.36E+04
2_4	7	4.375	2.991	0.375	0.595	0.595	0.37	0.67	6.00E+04
2_5	16	5.375	3.018	0.398	0.56	0.56	0.288	0.67	5.89E+04
2_6	6	3.312	2.399	0.375	0.465	0.465	0.193	0.67	5.89E+04
2_7	9	4.438	2.982	0.389	0.523	0.523	0.257	0.67	5.89E+04
2_8	50	8.062	3.274	0.435	0.559	0.559	0.222	0.77	6.12E+04
2_9	22	5.875	3.365	0.44	0.571	0.571	0.229	0.67	5.89E+04
2_10	24	6.5	3.335	0.469	0.567	0.567	0.173	0.67	5.78E+04
2_11	39	6.938	3.361	0.431	0.562	0.562	0.232	0.80	5.59E+04
2_12	21	5.875	3.231	0.464	0.573	0.573	0.19	0.67	5.59E+04
2_13	18	5.812	3.457	0.438	0.591	0.591	0.26	0.67	5.52E+04
2_14	47	7.875	3.408	0.438	0.563	0.563	0.222	0.84	5.43E+04
2_15	8	3.688	2.494	0.367	0.483	0.483	0.239	0.67	5.43E+04
2_16	20	5.562	3.143	0.456	0.578	0.578	0.21	0.67	5.37E+04
2_17	44	8.25	3.523	0.415	0.563	0.563	0.263	0.68	5.32E+04
2_18	11	3.875	2.362	0.369	0.451	0.451	0.18	0.67	5.24E+04
2_19	8	4.438	2.86	0.383	0.565	0.565	0.322	0.67	5.20E+04
2_20	19	4.75	2.747	0.467	0.529	0.529	0.117	0.67	5.20E+04
2_21	10	3.688	2.302	0.369	0.488	0.488	0.245	0.67	5.13E+04
2_22	16	5.375	3.414	0.43	0.57	0.57	0.246	0.67	5.11E+04
2_23	10	4.312	2.678	0.413	0.528	0.528	0.218	0.67	5.10E+04
2_24	6	3.812	2.482	0.406	0.547	0.547	0.257	0.67	5.18E+04
2_25	36	6.625	2.938	0.415	0.603	0.603	0.312	0.81	5.18E+04
2_26	27	5.375	3.03	0.412	0.561	0.561	0.265	0.67	6.02E+04
2_27	8	3.188	1.941	0.32	0.433	0.433	0.26	0.67	6.67E+04
2_28	21	4.75	2.398	0.411	0.468	0.468	0.122	0.76	6.76E+04

2_30	8	3.812	2.592	0.523	0.588	0.588	0.109	0.67	7.69E+04
2_31	15	4.938	3.263	0.475	0.581	0.581	0.182	0.67	7.95E+04
2_32	43	6.688	3.119	0.479	0.62	0.62	0.228	0.84	8.08E+04
2_33	24	6.375	2.973	0.469	0.605	0.605	0.225	0.67	5.10E+04
2_34	29	5.938	2.971	0.429	0.571	0.571	0.249	0.67	8.23E+04
2_35	44	7	3.406	0.447	0.551	0.551	0.188	0.84	8.69E+04
6_1	31	5.211	2.792	0.423	0.546	0.546	0.226	0.86	2.47E+05
6_2	29	5.053	2.94	0.37	0.563	0.563	0.343	0.84	2.45E+05
6_3	9	3.421	2.422	0.363	0.52	0.52	0.303	0.68	2.37E+05
6_4	23	5.158	2.89	0.449	0.613	0.613	0.268	0.68	1.66E+05
6_5	7	3.211	2.381	0.383	0.528	0.528	0.274	0.68	2.01E+05
6_6	5	3	2.325	0.368	0.542	0.542	0.32	0.68	1.98E+05
6_7	5	3.053	2.257	0.358	0.551	0.551	0.351	0.68	1.96E+05
6_8	18	4.316	2.638	0.401	0.572	0.572	0.299	0.83	1.84E+05
6_9	18	4.579	2.826	0.389	0.589	0.589	0.34	0.80	1.11E+05
6_10	19	4.158	2.31	0.424	0.502	0.502	0.157	0.83	1.10E+05
6_11	8	4.158	2.641	0.447	0.6	0.6	0.254	0.68	1.09E+05
6_12	6	3.263	2.315	0.386	0.549	0.549	0.297	0.68	1.08E+05
6_13	15	4.895	3.104	0.442	0.637	0.637	0.306	0.68	1.05E+05
6_14	28	5.263	3.106	0.455	0.625	0.625	0.272	0.80	1.05E+05
6_15	10	4.474	2.841	0.426	0.591	0.591	0.279	0.68	1.05E+05
6_16	16	4.579	3.155	0.477	0.604	0.604	0.21	0.75	1.05E+05
6_17	7	2.947	2.203	0.436	0.497	0.497	0.122	0.68	1.05E+05
6_18	5	3.053	2.282	0.421	0.563	0.563	0.252	0.68	1.05E+05
6_19	6	3.526	2.683	0.386	0.654	0.654	0.409	0.68	1.06E+05
6_20	16	4.105	2.484	0.444	0.56	0.56	0.207	0.72	1.06E+05
6_21	12	3.526	2.209	0.395	0.465	0.465	0.152	0.78	1.14E+05
9_1	5	3	2.361	0.4	0.558	0.558	0.283	0.68	2.08E+05
9_2	6	2.684	2.15	0.316	0.48	0.48	0.342	0.68	1.19E+05
9_3	11	3.737	2.641	0.321	0.539	0.539	0.405	0.68	1.16E+05
9_4	5	3.316	2.593	0.379	0.576	0.576	0.342	0.68	1.12E+05
9_5	18	4.211	2.431	0.336	0.509	0.509	0.339	0.73	1.10E+05
9_6	5	3.421	2.686	0.379	0.636	0.636	0.404	0.68	1.09E+05
9_7	19	4.632	2.595	0.355	0.523	0.523	0.322	0.76	1.01E+05
9_8	9	3.737	2.655	0.351	0.587	0.587	0.403	0.68	9.58E+04
9_9	7	3.158	2.38	0.444	0.544	0.544	0.184	0.68	9.15E+04
9_10	23	3.895	2.482	0.366	0.526	0.526	0.305	0.77	8.75E+04
9_11	51	5.263	2.823	0.411	0.523	0.523	0.215	0.84	8.34E+04
9_12	32	4.789	2.728	0.433	0.545	0.545	0.206	0.67	8.34E+04
9_13	45	5.579	2.88	0.451	0.538	0.538	0.161	0.67	8.31E+04
9_14	25	4.211	2.458	0.331	0.47	0.47	0.297	0.78	8.30E+04
9_15	30	5	2.575	0.405	0.524	0.524	0.226	0.73	8.29E+04
9_16	6	3.211	2.509	0.482	0.598	0.598	0.194	0.68	8.44E+04
9_17	7	3.263	2.324	0.289	0.509	0.509	0.432	0.68	8.50E+04

9_18	32	5.474	2.92	0.401	0.583	0.583	0.311	0.69	8.77E+04
9_19	41	5.947	3.146	0.39	0.583	0.583	0.33	0.86	8.82E+04
9_20	12	3.947	2.605	0.456	0.581	0.581	0.215	0.73	8.99E+04
9_21	17	3.737	2.245	0.334	0.482	0.482	0.306	0.79	9.61E+04

---

Chapter 5: Variation in the phenotypic and population genetic consequences of co-evolution in a host and pathogen metapopulation <sup>5</sup>

---

<sup>5</sup> Formatted for submission as a coauthored manuscript: Fields, P.D., Antonovics, J., and D. R. Taylor

**ABSTRACT**

We apply statistical genetic methods and controlled inoculations to determine the consequences of metapopulation dynamics for the evolution of disease resistance in the *Silene latifolia* – *Microbotryum violaceum* host-pathogen system. We found that, while there was not a distinct genetic signature separating populations that had and had not been infected, populations in a subset of the metapopulation showed a distinct population genetic signature associated with their disease history. Specifically, populations that had been healthy for a longer period of time were more genetically differentiated than those that had recently recovered or were still infected. We estimated migration among these recovered populations, and found that populations that had been healthier for longer were composed of a significantly larger portion of residents than those that had been diseased. Controlled inoculations of individuals from these recovered populations showed that populations that had recently recovered from infection were more resistant than those that had been without the pathogen for a longer period of time. Our findings are consistent with a loss of resistance resulting from a trade off between resistance and fitness.



## INTRODUCTION

The theory of interconnected populations (or metapopulations) has shown that spatial structure can have profound effects on evolution. This assertion has been shown to be especially true in regards to the evolution of host-pathogen systems, where the consequences of spatial population structure can be seen in distinctly different evolutionary outcomes. For example, KERR *et al.* (2006) showed that, when populations of the bacterial host *Escherichia coli* and the viral pathogen T4 coliphage are more isolated from one another, host-pathogen populations are more likely to coexist. Additionally, under panmictic conditions, both host and pathogen tend towards mutual extinction. More generally, conclusions derived from assumptions of panmixia are likely to be incorrect under the real-world conditions of ubiquitous population structure in host-pathogen systems (THRALL and BURDON 2002).

The introduction of spatial population structure into a host-pathogen system, and spatial genetic structure, results in outcomes not necessarily consistent with an idealized, panmictic population (THRALL and BURDON 2002). Within a metapopulation context, the potential for colonization/extinction, whose rates may be similar or different for host and pathogen, exists (THRALL and BURDON 1997). Concomitantly, among-population asynchrony in disease incidence and prevalence will arise, generating spatial variation in adaptation of the host to the pathogen (e.g. among population variance in resistance) and vice versa (e.g. among population variance in virulence of the pathogen) (THRALL and BURDON 2002).

Much of our current empirical understanding of the evolution host-pathogen interactions in plants describe a classical gene-for-gene model (GFG) (THOMPSON and BURDON 1992). One system in particular that does not seem to conform to the GFG model, but has nonetheless been important for understanding ecological and

evolutionary dynamics is the *Silene latifolia* – *Microbotryum violaceum* host-pathogen system. In this particular system, resistance behaves as a quantitative trait with high heritability (ALEXANDER *et al.* 1993), while the pathogen shows very little genetic variation in pathogen infectiousness. In addition to showing high variation in resistance, *S. latifolia* individuals exhibit large fitness costs (20-30%) associated with resistance in the absence of the disease. Specifically, more resistant plants flower later in the season, produce fewer flowers, and generally have lower seed and pollen output (ALEXANDER 1989; BIERE and ANTONOVICS 1996). These aforementioned costs will likely lead to distinct molecular population genetic consequences, wherein populations that are highly resistant will become more differentiated than the metapopulation as a whole due to limited e- and immigration due to lowered individual, and potentially, population fitness.

In the present experiment we focus on understanding the causes and consequences of spatial structure in a host-pathogen metapopulation. We focus our experiments on a long-term metapopulation study of the *S. latifolia* – *M. violaceum* host-pathogen system located in Southwestern, Virginia, USA. The long-term nature of the study system allows for the identification of asynchronous histories of infection, resistance, and restoration to susceptibility, of the type that has been theoretical shown to stabilize host-pathogen interactions. In addition to identifying the effects of spatial population structure on host-pathogen interactions, we identify the feedbacks that take place on the metapopulation itself via changes in extinction, colonization, and migration.

We test the theoretical expectation that spatial population structure will have substantive effects on the co-evolution of host and pathogen using high-resolution sampling, recently derived statistical genetic methods, and controlled inoculations. What we find is that, while the hypothesized dynamics are detectable in a subset of the

metapopulation, stochastic elements as well as historical contingency are likely limiting our ability to detect these hypothesized dynamics throughout the sampled area.

## **MATERIALS AND METHODS**

### **HOST STUDY SYSTEM**

*S. latifolia* Poir. (= *S. alba*, Caryophyllaceae) is a short-lived perennial plant that is broadly used as a model system for studying sex determination and sex chromosome evolution, host-pathogen dynamics, species invasion, organelle evolution, sexual dimorphism, sex ratio evolution, and evolution in structured populations (BERNASCONI *et al.* 2009). The populations in this study are distributed in patches along the roadsides and farmland of southwestern Virginia, in the vicinity of Mountain Lake Biological Station (Figure 1). This region has been the focus of a metapopulation census since 1988 (ANTONOVICS *et al.* 1994; RICHARDS 2000). Census records consist of approximate numbers of male and female plants along continuous stretches of roadside, as well as the number of individuals that are diseased (no attempt is made to determine the sex of individual diseased plants as this would likely increase the risk of incidental transmission through the census).

### **PATHOGEN STUDY SYSTEM**

Infection by *M. violaceum* results in anther-smut disease, where anthers produce dark violet fungal spores in place of normal yellow pollen. Diseased plants are easily identified by their dark-smutted anthers. The disease has a large fitness effect on the host in that infected individuals are not longer capable of producing viable gametes (ovules or pollen). The pathogen induces female flowers to abort the ovaries and instead produce stamens that bear diseased anthers. Spores (=teliospores) are transmitted to

new hosts by pollinators where they germinate, producing a yeast-like haploid stage (=sporidia) that conjugates to produce infection hyphae. Because pollinators adjust flight distances to compensate for plant density, transmission depends on the frequency and not the density of infectious individuals. Although the pathogen is vector transmitted, the frequency- dependent nature of the transmission and the expression of the disease in the sexual organs of the adult plants result in strong parallels between the biology of this host-pathogen system and sexually transmitted diseases (ANTONOVICS 2004; KALTZ and SCHMID 1995; LOCKHART *et al.* 1996). Once delivered to the flower, the pathogen grows into developing flower buds that become smutted flowers. Anther smut is a relatively “slow” disease with a latent period of 3-6 weeks. Initially, infected plants may be partially diseased, but the disease soon becomes systemic and the disease persists between seasons inside the over-wintering rosette of the host plant. The pathogen is easy to culture on petri plates and can be stored indefinitely as sporidial cultures.

## **SAMPLING**

### **HOST**

We sampled plants from 77 spatially distinct populations during peak flowering in the summer of 2008, spanning  $\sim 1/3$  of the metapopulation (Figure 1). We collected leaf tissue from every plant in the population, or up to 50 individuals in the largest populations, and stored the leaves in silica gel (Sigma). Genomic DNA was extracted and amplified following established microsatellite techniques for *S. latifolia* DNA was extracted from leaf tissue using the method described in (KELLER *et al.* 2012). We genotyped each individual plant at between 16 and 19 microsatellite loci. Our microsatellites are derived from multiple sources (ABDOULLAYE *et al.* 2010; JUILLET *et al.*

2003; MOCCIA *et al.* 2009; TEIXEIRA and BERNASCONI 2007). PCR amplification was conducted using published methods for each marker. PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems). Alleles were binned using the software TANDEM (MATSCHINER and SALZBURGER 2009).

Seeds were sampled from 22 extant populations distributed across the same portion of the metapopulation from which the tissue samples were collected. Populations ranged from 1 to 20 years of time since their last exposure to the disease. For estimates of disease resistance we obtained 15-50 seeds from between 1-5 individual host families.

## **PATHOGEN**

Fungal spores were collected from all infected plants in each focal section in 2008. Individual strains were germinated on petri dishes. Strains were randomly assigned to families among populations, and among families within populations.

## **INOCULATION**

For *in vitro* inoculation, a 2 µl suspension containing 1000 spores (standardized by spore counts, spore germination percentage, and appropriate dilution) was applied to the apical meristem of seedlings. These were grown in petri dishes on agar to standardize the stage of plant development, humidity, and temperature. Seedlings were

incubated for five days at 20 C<sup>0</sup> prior to transplanting into sterile soil.

We grew plants to flowering to assess disease status, and to measure the fitness components. Plants were grown using standard potting mix in "conetainers" (plastic 1-inch diameter tube-like pots, 6" deep, held in a 30x60 cm racks that holds 98 individuals). Individual susceptibility was determined by assessing disease presence or absence at the time of first flower production.

## **STATISTICAL METHODS**

### **STATISTICAL GENETICS**

We calculated population genetic summaries of genetic diversity, as well as estimates of genetic (sub)-structure via hierarchical *F*-statistics using the software GenoDive version 2.0b21 (MEIRMANS and VAN TIENDEREN 2004), with significant deviations from 0 assessed with 10,000 permutations and  $\alpha = 0.05$ .

We used the hierarchical Bayesian method of Foll and Gaggiotti (2006), implemented in the program GESTE v.2.0, to evaluate the effect disease infection history of individual populations on the magnitude of population genetic differentiation, or  $F_{ST}$ . Testing for the importance of one factor, disease history, led to two alternative models, one with only a constant (or null) and a second with a constant and the effect of disease history. The selected method provides posterior probabilities for the two models using a RJMCMC approach (FOLL and GAGGIOTTI 2006); the model with the highest posterior probability was selected as the one that best explained the data (FOLL and GAGGIOTTI 2006). We followed the method of Gaggiotti, Bekkevold et al. (2009), using 10 pilot runs of 1000 iterations to obtain parameters of the proposal distributions used by the MCMC, followed by additional burn-in of  $5 \times 10^6$  iterations and a thinning interval of 50, and final iteration sample size of 60,000 on which the model fit probability was based.

To assess how population isolation affected migration amongst populations, we used the program BayesAss v. 3.03 (WILSON and RANNALA 2003). We subdivided the populations into three sections of the metapopulation, and analyzed the sections separately. These sections are isolated geographically, with migration between sections being very unlikely.

A total of three runs per section were done, each using 50,000,000 Markov Chain Monte Carlo (MCMC) iterations and a burn-in of 500,000 iterations, and a thinning interval of 100, each with a different starting seed. In order to obtain appropriate mixing conditions, as determined by acceptance rate, in the MCMC chain, we modified the allele frequency, inbreeding coefficient, and migration rate parameters as per the BayesAss v. 3.03 manual suggestion. Chain convergence was assessed using the program Tracer v. 1.5 (RAMBAUT and DRUMMOND 2009).

We tested for effects of disease history on migration using Gaussian linear models. Linear modeling of the effects of population disease history on population isolation was performed in R v. 2.15.2. (R DEVELOPMENT CORE TEAM 2012).

## **POPULATION SUSCEPTIBILITY**

We used Gaussian linear models to determine if historical presence of the pathogen in populations affected variation in susceptibility. Specifically, the proportion of susceptible individuals in a population was our response variable, while population time since recovery (weighted by the number of individuals that were tested per population) was our independent variable. Linear models of the effects of population disease history on population susceptibility were performed in R v. 2.15.2. (R DEVELOPMENT CORE TEAM 2012).

## RESULTS

Individual microsatellites varied in the quantity of population (sub)structure, though only two markers in two of the three sections showed non-significant  $F_{ST}$  values (Table 1). The incidence and prevalence of the disease varied across the three sections of the metapopulation (Figure 2). Of the three sections, only one showed a significant effect of disease history on the distribution of  $F_{ST}$  (Table 2.). In this section, populations that had recovered from infection at an earlier time in the census were more genetically differentiated from the metapopulation as a whole (Table 3).

In our analysis of migration rates we used proportion resident (ranging from 0 to 100%) as a measure of genetic isolation. After controlling for size of the population sampled (which will likely affect the overall apparency of a given population to pollinators, time since recovery showed a significant effect (p-value = 0.0174) on proportion resident. Populations that lost the pathogen in the more distant past were more genetically isolated than those that had the infection more recently (data for sections 2 and 9 not shown) (Figure 3).

To determine whether disease history also affected patterns of host resistance, we tested for the effect of time since recovery on the proportion of the population that was susceptible. We found that populations that had recently lost the pathogen were ~ 10% more susceptible (p-value = 0.0301; Table 4 and Figure 4).

## DISCUSSION

In the present study, we have shown that the presence and distribution of the host and pathogen in the *S. latifolia* – *M. violaceum* system has generated distinct patterns of among population genetic variance, as well variation in host susceptibility, consistent with theoretical expectations (ANTONOVICS *et al.* 1998). Specifically, spatially



distributed populations with asynchronous disease dynamics also have contrasting history and present distributions of disease resistance. The results are consistent with the progressive loss of costly resistance over time in the absence of the pathogen. Our result paints a picture of recurring epidemics and sweeps of disease resistance and renewed susceptibility once the disease is extirpated.

By understanding the particular cost associated with the focal system, we were able to generate distinct population genetic predictions for how historical presence of the pathogen in the metapopulation might structure allelic variation. By reducing the number and time of flowering (BIERE and ANTONOVICS 1996), resistant populations are likely to become more differentiated from the metapopulation as a whole. We were able to detect this particular pattern by utilizing recently derived statistical genetic methods. We found that, for section 6 of our focal metapopulation, which was observed to undergo an epidemic of cycle of the pathogen, populations that once had the pathogen but eventually lost it (potentially through the evolution of increased resistance) became more differentiated over time, as measure by an increase in  $F_{ST}$ .

Because variation  $F_{ST}$  can be caused by both statistical and biological processes, we directly estimated migration rates into these populations that recovered from infection of the pathogen. We were able to show that the proportion of individual populations that were resident (or non-migrant) has increased significantly with time since recovery, indicating that the increase in  $F_{ST}$  may result from reduced migration or enhanced drift (Figure 3). A heavy cost of resistance is also supported by the finding that more resistant populations tend to have a history of smaller population size (Antonovics, unpublished data).

Finally, to determine if the observed population genetic summaries were consistent with the assumed phenotype of resistance, we conducted controlled

inoculations on populations that had recovered at different times during the census period. We found that populations that had recently recovered showed higher resistance (or lower susceptibility) than populations that had recovered at an earlier period (Figure 4). Loss of resistance is expected in the absence of the pathogen given the observed tradeoff.

The presented experiments focus on the dynamics of host-pathogen interactions, and their distinct effect on individual populations. However, it should be understood that these particular dynamics will have a larger effect on the metapopulation as a whole. Field experiments and modeling studies have led to the hypothesis that both intra- and inter- demic selection influences resistance evolution and the disease dynamics in this system (ANTONOVICS 1999; ANTONOVICS 2004; ANTONOVICS *et al.* 1998; O'KEEFE and ANTONOVICS 2002).

In brief, there is evidence that the evolution of resistance drives the disease to low levels within populations, such that stochastic local extinction of the disease is commonplace (ANTONOVICS 2004). Evidence for rapid evolution of resistance within demes comes from contrasting dynamics seen in field experiments started from resistant or susceptible genotypes (THRALL and JAROSZ 1994a; THRALL and JAROSZ 1994b). Rapid changes in resistance also occurred in field experiments where both diseased and control healthy populations, started from identical seed sources, were maintained for 5 years (Antonovics and Ling, in preparation). The disease may be maintained by its establishment on newly colonized populations, and we hypothesize that these have a higher probability of becoming diseased because they are often founded by susceptible genotypes owing to the large cost of resistance (in the order of 20-30%) in the form of lower seed and pollen output of highly resistant individuals (Biere and Antonovics 1996).

The present study also points to the possibility of both intra- and inter-demic selection. Multilevel selection is often presented as different levels of selection in conflict with one another. In the present system, the scenario is one where the conflicts are not universal, but dependent on the presence or absence of a pathogen. In the absence of the pathogen, selection at the intra-demic level is concordant with selection at the inter-demic level, and populations should rapidly evolve reduced resistance. However, in the presence of the pathogen, there is conflicting selection for greater resistance at the intra-demic level but selection for lower resistance at the inter-demic level. This could have consequences for the maintenance of variation in the trait itself (i.e. disease resistance) but also ecological repercussions for pathogen occurrence on the host and its impact on population size and ecological distribution.

In the present study, we focus on a host-pathogen system where resistance in the host is quantitative, and where a tradeoff in fitness through flower and gamete production generates very distinct evolutionary predictions for how susceptibility will be maintained in the absence of the pathogen. Additionally, we have assumed little or no variation in infectivity exists in the pathogen based upon previous research in the focal metapopulation, which has been hypothesized to be the result of a strong bottleneck during the colonization of North America (Fontaine *et al.* 2013). However, recent studies of populations from central Italy (a glacial refuge where the host pathogen system has been long-standing) have shown a similar large variation in host resistance, but no evidence of host-genotype by pathogen-genotype interaction (Amos and Antonovics, unpublished).

## ACKNOWLEDGMENTS

PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT and JA, and NSF-OISE# 1139716 to DT and PF.

## LITERATURE CITED

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010 Permanent genetic resources added to Molecular Ecology resources database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ALEXANDER, H. M., 1989 An experimental field-study of anther-smut disease of *Silene alba* caused by *Ustilago-Violacea* - genotypic variation and disease incidence. *Evolution* **43**: 835-847.
- ALEXANDER, H. M., J. ANTONOVICS and A. W. KELLY, 1993 Genotypic variation in plant disease resistance--physiological resistance in relation to field disease transmission. *Journal of Ecology* **81**: 325-333.
- ANTONOVICS, J., 1999 Pathogens and plant population dynamics: the effects of resistance genes on numbers and distribution, pp. 49-55 in *Ecological effects of pest resistance genes in managed ecosystems*, edited by T. P.M. and W. H.H. Information Systems for Biotechnology, Blacksburg, VA.
- ANTONOVICS, J., 2004 Long term study of a plant-pathogen metapopulation, pp. 471-488 in *Ecology, genetics and evolution of metapopulations*, edited by I. HANSKI and O. GAGGIOTTI. Academic Press.
- ANTONOVICS, J., P. THRALL and A. JAROSZ, 1998 Genetics and the spatial ecology of species interactions: the *Silene-Ustilago* system, pp. 158-180 in *Spatial ecology: the role of space in population dynamics and interspecific interactions*, edited by D. TILMAN and P. KAREIVA. Princeton University Press.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in *Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.

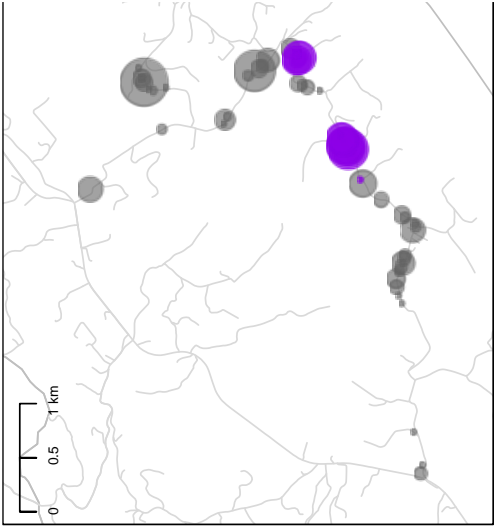
- BERNASCONI, G., J. ANTONOVICS, A. BIERE, D. CHARLESWORTH, L. F. DELPH *et al.*, 2009  
*Silene* as a model system in ecology and evolution. *Heredity* **103**: 5-14.
- BIERE, A., and J. ANTONOVICS, 1996 Sex-specific costs of resistance to the fungal pathogen  
*Ustilago violacea* (*Microbotryum violaceum*) in *Silene alba*. *Evolution* **50**: 1098-1110.
- FOLL, M., and O. E. GAGGIOTTI, 2006 Identifying the environmental factors that determine  
the genetic structure of populations. *Genetics* **174**: 875-891.
- FONTAINE, M. C., P. GLADIEUX, M. E. HOOD and T. GIRAUD, 2013 History of the invasion of  
the anther smut pathogen on *Silene latifolia* in North America. *New Phytologist* **198**:  
946-956.
- GAGGIOTTI, O. E., D. BEKKEVOLD, H. B. H. JØRGENSEN, M. FOLL, G. R. CARVALHO *et al.*,  
2009 Disentangling the effects of evolutionary, demographic, and environmental  
factors influencing genetic structure of natural populations: Atlantic Herring as a case  
study. *Evolution* **63**: 2939-2951.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of  
highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris*  
(Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.
- KALTZ, O., and B. SCHMID, 1995 Plant venereal-disease - a model for integrating genetics,  
ecology and epidemiology. *Trends in Ecology & Evolution* **10**: 221-222.
- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a  
complex invasion history revealed by nuclear and chloroplast genetic diversity in the  
colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- KERR, B., C. NEUHAUSER, B. J. M. BOHANNAN and A. M. DEAN, 2006 Local migration  
promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature*  
**442**: 75-78.

- LOCKHART, A. B., P. H. THRALL and J. ANTONOVICS, 1996 Sexually transmitted diseases in animals: ecological and evolutionary implications. *Biological Reviews* **71**: 415-471.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- O'KEEFE, K., and J. ANTONOVICS, 2002 Playing by different rules: the evolution of virulence in sterilizing pathogens. *American Naturalist* **159**: 597-605.
- R DEVELOPMENT CORE TEAM, 2012 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.
- RAMBAUT, A., and A. J. DRUMMOND, 2009 Tracer v1.5 [Internet]. **Available from: URL** <http://beast.bio.ed.ac.uk/Tracer>.
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.
- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.
- THOMPSON, J. N., and J. J. BURDON, 1992 Gene-for-gene coevolution between plants and parasites. *Nature* **360**: 121-125.

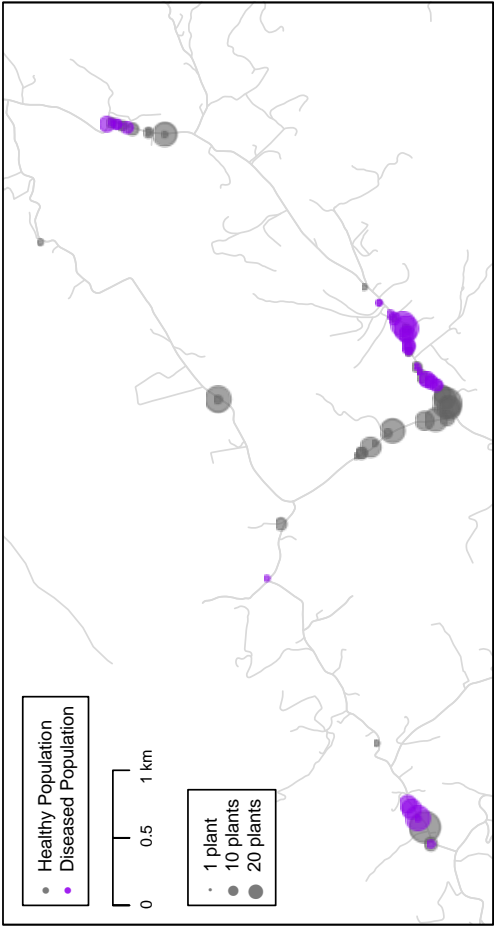
- THRALL, P. H., and J. J. BURDON, 1997 Host-pathogen dynamics in a metapopulation context: the ecological and evolutionary consequences of being spatial. *Journal of Ecology* **85**: 743-753.
- THRALL, P. H., and J. J. BURDON, 2002 Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathology* **51**: 169-184.
- THRALL, P. H., and A. M. JAROSZ, 1994a Host-pathogen dynamics in experimental populations of *Silene alba* and *Ustilago-violacea* .1. Ecological and genetic-determinants of disease spread *Journal of Ecology* **82**: 549-559.
- THRALL, P. H., and A. M. JAROSZ, 1994b Host-pathogen dynamics in experimental populations of *Silene alba* and *Ustilago-violacea* .2. Experimental tests of theoretical-models. *Journal of Ecology* **82**: 561-570.
- WILSON, G. A., and B. RANNALA, 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177-1191.



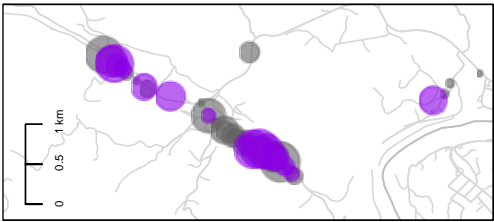
**Figure 1.** Spatial distribution of infected *S. latifolia* populations in the three focal metapopulation section. Populations colored grey were never infected by the pathogen, while those colored purple were infected at some point during the census history. Circle size is representative of the population size at the time of tissue collection in 2008.



Section 9



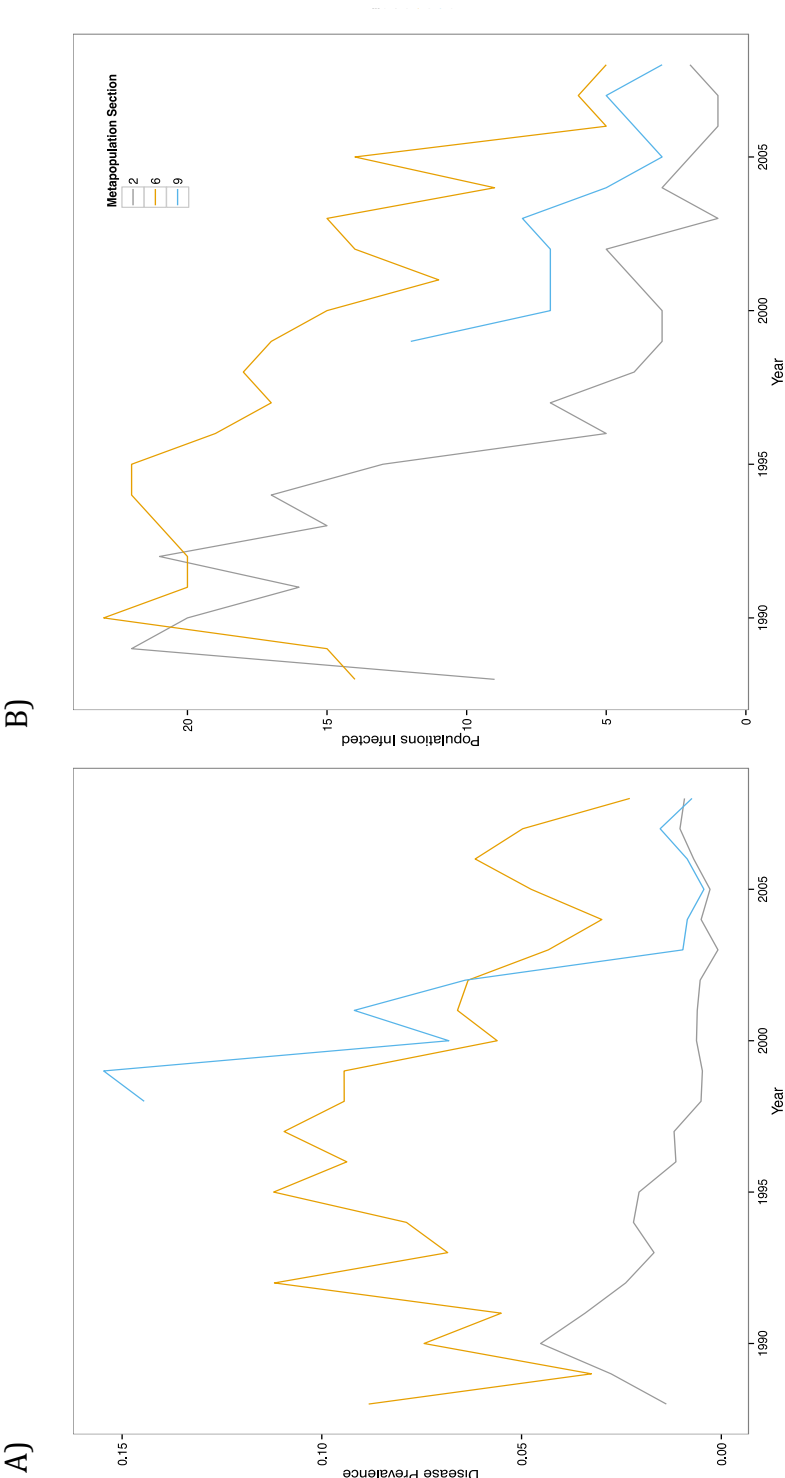
Section 6



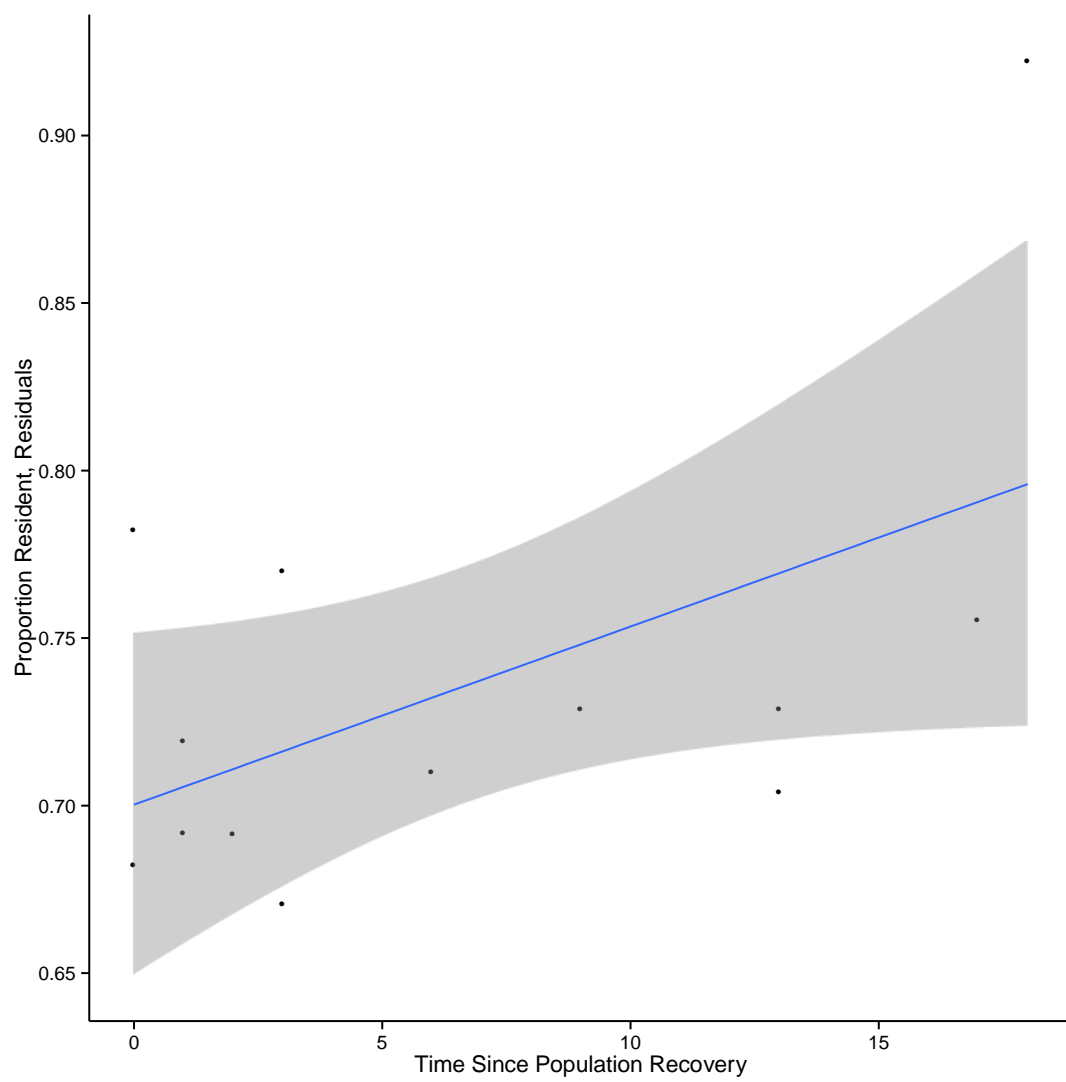
Section 2

**Figure 2.** Disease history of the focal *S. latifolia* – *M. violaceum* metapopulation. (A)

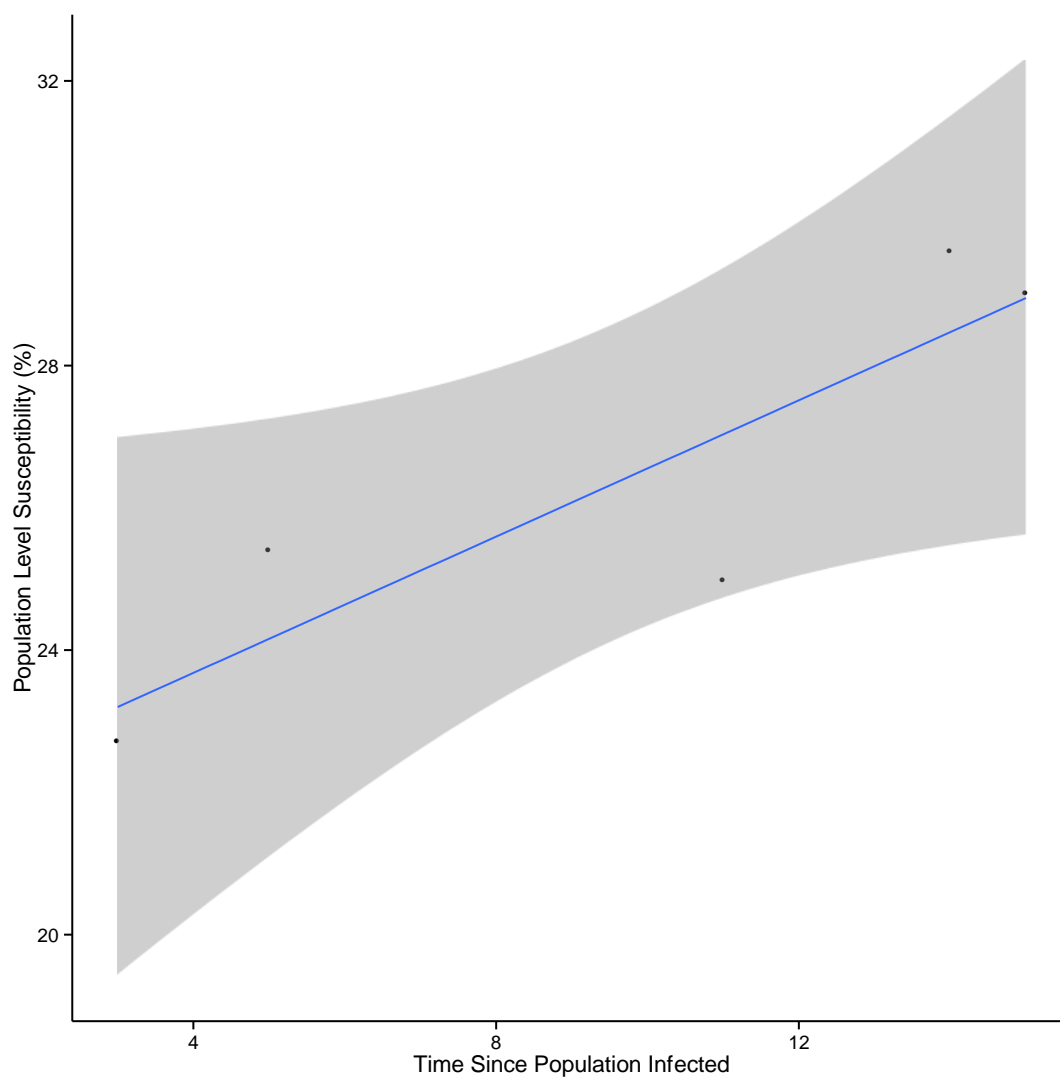
Disease prevalence is measured as the number of individuals diseased / (the number of individuals diseased + the number healthy individuals) in a focal metapopulation section. (B) Populations infected is a count of the number populations that are infected in a focal metapopulation. For section 2 and section 6, census data covers years 1988 - 2008, while for section 9 data covers the years 1998 – 2008. Each section's history is represented by a different color.



**Figure 3.** Proportion of population that is resident regressed against time since population recovery. Proportion resident is a BayeAss based estimate of the proportion of population that is not a recent migrant, or the result of a mating with a recent migrant, while time since population recovery is a measure of the time since an infected population became completely healthy. Populations with a value of 0 are still infected. Populations that have recovered at an earlier time have a lower proportion of migrants than those that have more recently recovered or are still infected (p-value = 0.0174). The points represent the residuals of the model after controlling for population size.



**Figure 4.** Loss of fungal resistance. Population susceptibility is a measure of the proportion of tested individuals that, after being exposed to fungal spores as seedlings, become infected, while time since population infected is a measure of time since an infected population recovered. Populations that have recently recovered from the presence of the pathogen are less susceptible than those that have recovered at an earlier time (p-value = 0.0301).



**Table 1.** Global population genetic summaries. Variables are the number of alleles ( $N$ ), the effective number of alleles ( $E_N$ ), observed heterozygosity ( $H_o$ ), Expected heterozygosity ( $H_E$ ), and the inbreeding coefficient ( $F_{IS}$ ), global among-population allelic variation ( $F_{IS}$ ), and Jost’s measure of population differentiation ( $D_{EST}$ ).

Locus	Section 2								Section 6								Section 9						
	N	E_N	H <sub>O</sub>	H <sub>E</sub>	F <sub>IS</sub>	F <sub>ST</sub>	D <sub>EST</sub>		N	E_N	H <sub>O</sub>	H <sub>E</sub>	F <sub>IS</sub>	F <sub>ST</sub>	D <sub>EST</sub>		N	E_N	H <sub>O</sub>	H <sub>E</sub>	F <sub>IS</sub>	F <sub>ST</sub>	D <sub>EST</sub>
SL_eSSR01 <sup>1</sup>	-	-	-	-	-	-	-	7	1.818	0.224	0.488	0.542	<b>0.114</b>	0.129	6	1.722	0.161	0.452	0.644	<b>0.099</b>	0.094		
SL_eSSR04 <sup>1</sup>	10	2.282	0.577	0.584	0.012	<b>0.05</b>	0.076	7	2.399	0.524	0.62	0.154	<b>0.05</b>	0.089	6	2.475	0.519	0.629	0.174	<b>0.105</b>	0.206		
SL_eSSR06 <sup>1</sup>	18	3.762	0.699	0.765	0.086	<b>0.041</b>	0.145	14	3.333	0.503	0.751	0.33	<b>0.069</b>	0.233	13	3.108	0.39	0.725	0.462	<b>0.1</b>	0.304		
SL_eSSR09 <sup>1</sup>	12	1.407	0.287	0.301	0.049	<b>0.037</b>	0.017	7	1.65	0.393	0.416	0.055	<b>0.073</b>	0.059	9	1.584	0.382	0.386	0.01	<b>0.091</b>	0.065		
SL_eSSR12 <sup>1</sup>	20	3.296	0.724	0.724	-0.001	<b>0.044</b>	0.123	14	3.557	0.674	0.762	0.116	<b>0.117</b>	0.444	15	4.093	0.702	0.795	0.117	<b>0.087</b>	0.381		
SL_eSSR16 <sup>1</sup>	8	1.569	0.334	0.378	0.117	<b>0.099</b>	0.068	4	2.245	0.583	0.584	0.002	<b>0.073</b>	0.116	8	2.248	0.537	0.583	0.079	<b>0.141</b>	0.238		
SL_eSSR17 <sup>1</sup>	-	-	-	-	-	-	-	13	2.747	0.608	0.674	0.098	<b>0.106</b>	0.256	14	2.66	0.556	0.658	0.154	<b>0.069</b>	0.147		
SL_eSSR20 <sup>1</sup>	11	1.161	0.121	0.145	0.168	<b>0.065</b>	0.012	3	1.057	0.056	0.057	0.016	0.028	0.002	3	1.048	0.038	0.049	0.218	0.039	0.002		
SL_eSSR22 <sup>1</sup>	-	-	-	-	-	-	-	10	2.817	0.561	0.686	0.183	<b>0.149</b>	0.403	11	3.167	0.501	0.726	0.31	<b>0.084</b>	0.253		
SL_eSSR27 <sup>1</sup>	-	-	-	-	-	-	-	9	2.462	0.58	0.628	0.077	<b>0.167</b>	0.355	11	2.568	0.595	0.641	0.071	<b>0.128</b>	0.271		
SL_eSSR28 <sup>1</sup>	-	-	-	-	-	-	-	6	1.79	0.469	0.465	-0.008	<b>0.22</b>	0.257	6	1.353	0.24	0.275	0.126	<b>0.245</b>	0.127		
SL_eSSR29 <sup>1</sup>	25	5.488	0.835	0.85	0.018	<b>0.054</b>	0.332	19	3.871	0.715	0.785	0.09	<b>0.103</b>	0.441	18	3.683	0.7	0.765	0.085	<b>0.083</b>	0.305		
SL_eSSR30 <sup>1</sup>	21	3.407	0.65	0.737	0.119	<b>0.056</b>	0.17	21	2.683	0.345	0.679	0.492	<b>0.111</b>	0.278	18	2.705	0.281	0.678	0.585	<b>0.083</b>	0.197		
slat_18 <sup>2</sup>	10	2.837	0.418	0.683	0.388	<b>0.058</b>	0.137	9	2.045	0.234	0.556	0.578	<b>0.208</b>	0.346	11	2.522	0.276	0.649	0.575	<b>0.119</b>	0.259		
slat_32 <sup>2</sup>	9	2.709	0.29	0.67	0.568	<b>0.137</b>	0.331	7	2.271	0.364	0.602	0.396	<b>0.159</b>	0.301	9	2.384	0.34	0.62	0.451	<b>0.181</b>	0.375		
slat_33 <sup>2</sup>	3	1.128	0.061	0.12	0.49	<b>0.237</b>	0.044	2	1.515	0.184	0.368	0.502	<b>0.265</b>	0.221	3	1.211	0.105	0.186	0.434	<b>0.597</b>	0.35		
slat_48 <sup>2</sup>	6	1.253	0.087	0.214	0.593	<b>0.099</b>	0.031	5	1.766	0.194	0.472	0.589	<b>0.108</b>	0.113	5	1.666	0.233	0.427	0.455	<b>0.139</b>	0.125		
slat_72 <sup>2</sup>	12	2.858	0.349	0.688	0.493	<b>0.048</b>	0.116	12	2.545	0.326	0.657	0.504	<b>0.097</b>	0.217	16	2.86	0.282	0.7	0.596	<b>0.171</b>	0.499		
slat_85 <sup>2</sup>	15	2.085	0.311	0.55	0.434	<b>0.101</b>	0.142	16	1.829	0.284	0.489	0.418	<b>0.156</b>	0.185	16	1.618	0.232	0.408	0.432	<b>0.194</b>	0.172		
SI_8 <sup>3</sup>	11	2.153	0.324	0.566	0.428	<b>0.132</b>	0.203	-	-	-	-	-	-	-	-	-	-	-	-	-			
A11 <sup>4</sup>	43	4.382	0.707	0.806	0.122	<b>0.062</b>	0.283	-	-	-	-	-	-	-	-	-	-	-	-	-			
Overall	14.625	2.611	0.423	0.549	0.229	<b>0.074</b>	0.099	9.737	2.337	0.412	0.565	0.272	<b>0.128</b>	0.2	10.421	2.351	0.372	0.545	0.317	<b>0.138</b>	0.198		



**Table 2.** GESTE based model comparison to determine if past presence *M. violaceum* contributes to  $F_{ST}$  in a metapopulation of *S. latifolia*. For sections 2 and 9, time since recovery was not part of the highest posterior model probability, which was instead the model that contained only a constant. For section 6, the model that contained time since population recovery was included in the highest posterior probability model, which is also the section that shows evidence of an extended epidemic.

Model	Factors Included	<b>2</b>	<b>6</b>	<b>9</b>
		Pr	Pr	Pr
1	Null	<b>0.925</b>	0.341	<b>0.937</b>
2	Time Since Population Recovery	0.0745	<b>0.659</b>	0.0634

**Table 3.** Regression coefficients, and their confidence intervals, from GESTE estimated model with highest posterior probabilities for section 6 (section 2 and 9 were excluded due to the model including only a constant showing the highest posterior model probability. Time since population recovery has a positive regression coefficient, indicating that populations that recovered recently or are still infected are less differentiated than those that recovered at an earlier period.

Regression Coefficient	Factor	Mean	Mode	95% HPDI
$\alpha_0$	Constant	-1.6	-1.58	[-1.89 ; -1.30]
$\alpha_1$	<b>Time Since Population Recovery</b>	<b>0.335</b>	<b>0.344</b>	<b>[0.0912; 0.582]</b>
$\sigma^2$	-	0.327	0.246	[0.117; 0.600 ]

**Table 4.** Summary of populations used for testing among population variation in population susceptibility. *N* is the number of individuals used from the population, susceptible is the number of individuals that become infected by the sterilizing pathogen and resistant the number of individuals that remained healthy, following exposure to fungal spores as young seedlings. Years since recovery is a measure of the time since an infected population become healthy.

Population	<i>N</i>	Suscpetible	Resistant	Years Since Recovery
6.1.a 0.275_R	59	15	44	3
6.1.a 0.3_R	88	20	68	1
6.1.f 0.625_R	27	8	19	12
6.1.g 0.275_R	62	18	44	13
6.1.h 0.2_R	40	10	30	9

Chapter 6: Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability<sup>6</sup>

---

<sup>6</sup> Formatted for submission as a coauthored manuscript: Fields, P.D., McCauley, D.E., McAssey, E. V., and D. R. Taylor

## ABSTRACT

Non-random association of alleles in the nucleus and cytoplasmic organelles, or cyto-nuclear linkage disequilibrium (LD), is both an important component of a number of evolutionary processes and a statistical indicator of others. The evolutionary significance of cyto-nuclear LD will depend on both its magnitude and how stable those associations are through time. Here, we use a longitudinal population genetic dataset to explore the magnitude and temporal dynamics of cytonuclear disequilibria through time. We genotyped 135 and 170 individuals from 16 and 17 patches of the plant species *Silene latifolia* in Southwestern, VA sampled in 1993 and 2008, respectively. Individuals were genotyped at 14 highly polymorphic microsatellite markers and a SNP in the mitochondrial gene, *atp1*. Normalized linkage disequilibrium ( $D'$ ) between nuclear and cytoplasmic loci varied considerably depending on which nuclear locus was considered (ranging from 0.005 to 0.632). Four of the 14 cyto-nuclear associations showed a statistically significant shift over approximately seven generations. However, the overall magnitude of this disequilibrium was largely stable over time. The observed origin and stability of cyto-nuclear LD is most likely caused by the slow admixture between anciently diverged lineages within the species' newly invaded range, and the local spatial structure and metapopulation dynamics that are known to structure genetic variation in this system.

**Keywords:** cyto-nuclear, linkage disequilibrium, metapopulations, spatiotemporal, *Silene*

## INTRODUCTION

Population genetic studies generally focus on either the nuclear or the cytoplasmic (mitochondria or chloroplast) genome. Additional insights can be gained by measuring joint allelic associations at loci derived from nuclear and cytoplasmic genomes, or cyto-nuclear linkage disequilibrium (LD). Analyses of cyto-nuclear LD can document evolutionary and demographic trends in population genetics. Processes like structure (NEI and LI 1973), gene flow (ASMUSSEN and SCHNABEL 1991), and non-random mating/inbreeding (ASMUSSEN *et al.* 1989; SLATKIN 2008) all influence cyto-nuclear linkage disequilibrium. Selection based treatments of cyto-nuclear LD focus on the importance of epistatic interactions between the nuclear and cytoplasmic genomes (BRANDVAIN and WADE 2009; WADE and GOODNIGHT 2006). Theoretical investigations of genetic drift predict an influence of population size on the variance of linkage disequilibrium measurements (DATTA *et al.* 1996). Additionally, demographic processes have been shown to greatly facilitate non-random associations between the nucleus and a cytoplasmic genome. For example, cyto-nuclear LD (or lack thereof) is an informative metric when considering admixture of two populations of the same species (ARNOLD 2006). Thus, cyto-nuclear LD will be driven by the interaction of spatiotemporal selective and demographic parameters specific to a system of study, or by non-equilibrium population dynamics associated with the metapopulation structure of many natural systems.

While the aforementioned investigations provide substantial information concerning the patterns resulting from cyto-nuclear disequilibrium, few incorporate a temporal component. A pioneering study in *Drosophila montana* used allozyme polymorphisms to track changes in nuclear linkage disequilibrium over a five year time

period (BAKER 1975). The author concluded that maintenance of LD was consistent with epistatic selection maintaining favorable gene – gene interactions (BAKER 1975).

Conversely, in *Drosophila melanogaster* it was found that there was an absence of linkage disequilibrium among allozyme loci when samples were taken at two time points within the same year (LANGLEY *et al.* 1977). More recent studies in *Caenorhabditis elegans* have suggested that the maintenance of significant linkage disequilibrium among nuclear-based microsatellite markers over short timescales could be generated by genetic drift in small populations or selection against hybrid progeny (outbreeding depression) (BARRIÈRE and FÉLIX 2007).

While the previously described animal oriented studies focused on linkage disequilibrium among nuclear markers (allozymes and microsattellites), the development of markers in cytoplasmic genomes has allowed for an in depth analysis of cyto-nuclear LD. Thus, tracking levels of cyto-nuclear LD on a microevolutionary timescale allow for an analysis of the tempo of the change in multi-locus interactions between the nuclear and organellar genomes. Of particular interest is determining whether or not a cyto-nuclear interaction is deteriorating, or strengthening over time. Deterioration of linkage disequilibrium can be achieved with sexual recombination although it should be noted that population structure (bi-parental inbreeding) can slow down the rate of deterioration due to outcrossing (BRANDVAIN and WADE 2009). Strengthening of linkage disequilibrium can be the result of selection for a particular set of alleles (epistasis), demographic expansion where population sizes grow rapidly, or due to a genetic bottleneck where only a small number of allelic associations remain after a stochastic event and thus particular multi-locus genotypes rise in number despite an absence of selection. It is essential that demographic and selective forces be separated when trying to infer the evolutionary dynamics of a particular system

(STINCHCOMBE and HOEKSTRA 2007). Since demographic processes generally act on the whole genome, locus specific patterns of maintenance or strengthening of linkage disequilibrium are suggestive of selection for a particular cyto-nuclear combination of alleles.

*Silene latifolia* is a dioecious, insect pollinated plant that has been the subject of a considerable number of studies that focus on local population genetic structure, and the effects of metapopulation dynamics on that structure (BERNASCONI *et al.* 2009; MCCAULEY 1994; MCCAULEY *et al.* 1995; MCCAULEY *et al.* 1996). Many of these studies were conducted more than fifteen years ago, before the introduction of most of the DNA-based genetic markers used in studies of population structure today. However, they did utilize PCR-based chloroplast DNA (cpDNA) markers for comparison with allozyme nuclear markers (the workhorse of empirical population genetics at that time). McCauley (1994) observed a contrast between the moderate local population structure of the nuclear gene markers ( $F_{ST} \sim 0.20$ ) and a much greater degree of structure of the chloroplast markers ( $F_{ST} \sim 0.60$ ). He attributed this to differences in the mode of dispersal of the two genomes (seed + pollen movement in the bi-parentally inherited nuclear genome vs. seed movement in the maternally inherited chloroplast genome) (MCCAULEY 1994). Further, founding events associated with the repeated establishment of new populations enhanced the magnitude of structure for both genomes (MCCAULEY *et al.* 1995). These metapopulation dynamics have since been shown to be important determinants of population genetic structure in several other systems (GILES and GOUDET 1997).

The earlier studies of *S. latifolia* did not focus on cyto-nuclear LD, but the DNA extractions were archived and can now be used to generate nuclear and cyto-nuclear genotypes using the more powerful DNA-based genotyping techniques that are currently available. A comparison of cyto-nuclear associations between archived samples and



current *S. latifolia* collections allow us to examine the temporal dynamics of cyto-nuclear LD in a natural metapopulation. Standard questions about the magnitude of cyto-nuclear LD, and locus-to-locus variation in LD, can be extended to include their temporal dynamics. Here we report on the association between variants found in a cytoplasmic (mitochondrial) SNP marker and variants found in fourteen nuclear genes in samples of natural populations of *S. latifolia* taken in 1993 and again in 2008, including observations of locus-to-locus heterogeneity of cyto-nuclear LD and the temporal stability of those associations.

We also include a full population genetic analysis of our nuclear markers. By doing so, we allow for a larger comparison to many other published studies of nuclear-nuclear LD, suggesting that the presence and stability of cyto-nuclear LD may be far more common than has previously been suggested.

## MATERIALS AND METHODS

Collections of *Silene latifolia* were first made from 16 patches of plants (with a mean, or  $\bar{N}$ , per site of 8.5, and a range of 1 to 29 individuals per patch) found along the roadsides of Giles County, Virginia in summer 1993. All patches were located <20 km from one another. These samples represent a subset of the individuals included in earlier studies of local population structure (McCAULEY 1994; McCAULEY *et al.* 1995). For those studies leaf tissue was used both to generate nuclear allozyme genotypes and chloroplast DNA (cpDNA) RFLP's. Genomic DNA used for cpDNA genotyping was extracted using standard methods. This DNA was then stored at -80C until genotyped recently for the mtDNA and microsatellite markers used in the current study.

The same population networks were sampled (17 in total, and  $\bar{N} = 10$  per patch) in the summer of 2008. *S. latifolia* has been described as a short lived perennial

(BERNASCONI *et al.* 2009) and it has been our experience from field experiments that the average lifespan of *S. latifolia* is ~2 years with the age at first reproduction being considerably less. Thus, we estimate that a minimum of seven generations separates the collections, and are confident that no individual plants were sampled twice. It should be noted that while the two collections were made from the same stretch of the metapopulation, they were not necessarily made from exactly the same places. In fact, because this weedy plant is known to undergo fine-scale episodes of extinction and colonization (MCCAULEY *et al.* 1995), some local populations collected in 1993 may have gone extinct by 2008. Similarly, some 2008 collections may have been from recent colonizations of localities that did not contain *S. latifolia* in 1993.

For the 2008 collections, DNA was extracted from leaf tissue using the method described in KELLER *et al.* (2012). All individuals from both collections were assayed for a SNP known to occur in the mitochondrial gene *atp1* by a PCR/RFLP method (MCCAULEY and ELLIS 2008). We prefer this mitochondrial SNP as our cytoplasmic marker, rather than the cpDNA cytoplasmic markers used in earlier studies (e.g. MCCAULEY 1994), because the cpDNA markers consisted of indel polymorphisms that could be more subject to homoplasy. The SNP in question determines the presence or absence of an Alu I restriction enzyme cut site. Genomic DNA was subject to a polymerase chain reaction (PCR) that utilized the *atp1* primers and cycle conditions used by MCCAULEY and ELLIS (2008). 10 µl of the resulting PCR product was digested with Alu I using the manufacturers recommended conditions (NEB). The resulting fragments were electrophoresed on a 4% Metaphor agarose gel, which was then stained with ethidium bromide for visualization.

Genomic DNA from the same individuals was used to generate multilocus genotypes at 14 unlinked microsatellite loci. Microsatellites were derived from multiple

sources (ABDOULLAYE *et al.* 2010; MOCCIA *et al.* 2009). PCR amplification was conducted using published methods for each marker. PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems). Alleles were binned using the software TANDEM (MATSCHINER and SALZBURGER 2009).

We calculated the observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity of our nuclear genetic markers using the software GenoDive version 2.0b21 (MEIRMANS and VAN TIENDEREN 2004). Estimates of genetic substructure using hierarchical  $F$ -statistics were calculated using the software FSTAT 2.9.3.2 (GOUDET 2002), with significant deviations from panmixia assessed by testing for Hardy-Weinberg Equilibrium with 10,000 permutations and  $\alpha = 0.05$ .

To estimate patterns of nuclear-nuclear LD (hereafter, nuclear LD) among microsatellite loci, we applied Hedrick's (1987) multi-allelic extension of Lewontin's (1964) normalized  $D'$ . This measure of LD is preferable to others as it is widely used, and given its normalization, can be compared to other studies (LI and MERILA 2010; SLATE and PEMBERTON 2007).  $D'$  ranges from zero (no allelic associations between loci) to one (complete allelic associations at two loci), though estimation of the statistic may be sensitive to allele frequencies and sample sizes (LI and MERILA 2010; SLATE and PEMBERTON 2007; but see ZAPATA 2011).  $D'$  between two multi-allelic markers was calculated following LI and MERILA (2010) equations (1) and (3), using Multiallelic interallelic disequilibrium analysis software (MIDAS) (GAUNT *et al.* 2006).

Statistical significance of nuclear LD between pairs of loci for a given sampling date was estimated, under the null hypothesis of random allelic assortment, using a Monte-Carlo approximation of Fisher's exact test implemented in the software Arlequin (EXCOFFIER and LISCHER 2010). Arlequin uses a Markov chain extension of Fisher's exact test for  $R \times C$  contingency tables (LI and MERILA 2010; SLATKIN 1994). 100,000 alternative tables were explored by the Markov chain (LI and MERILA 2010; SLATE and PEMBERTON 2007).

Cyto-nuclear LD was estimated between each nuclear microsatellite locus and the *atp1* mtDNA locus. Analysis followed the approach of Basten and Asmussen (1997), using the program CNDm to estimate a standardized estimate an allelic  $D'$  between each nuclear locus and *atp1* mtDNA locus. CNDm uses a Monte Carlo approach to approximate Fisher's exact test for  $R \times C$  contingency tables and tests for significant deviations from the null hypothesis of no allelic association (BASTEN and ASMUSSEN 1997). For this analysis, all nuclear loci were treated as bi-allelic by pooling all alleles other than the most common allele in 1993 into a single composite allele (LATTA *et al.* 2001). This approach is preferable for the present analysis, as it provides a single value for each locus-locus pair, while also generating the most intermediate allele frequencies, thereby maximizing the bounds on  $D'$ . Additionally, given finite sample sizes, and the propensity of microsatellite mutation rates to generate a large class of rare, private alleles, which will rarely be in linkage equilibrium, our binning procedure preserves statistical robustness.

Cyto-nuclear LD was calculated separately for the 1993 and 2008 collection. We also calculated cyto-nuclear LD of a pooled sample of both years, which allows one to detect long-term patterns of cyto-nuclear LD, and its consequences. For example, year-to-year consistency of individual  $D'$  values will reinforce one another, generating

increased significance of cyto-nuclear LD. Conversely, when  $D'$  reverses between years, the overall significance of cyto-nuclear LD would be canceled out in a single pooled value. Year to year stability of  $D'$  values was evaluated statistically by two methods. Temporal consistency in the relative rankings of locus-specific  $D'$  values was evaluated by estimating the between-year correlation of the 14 pairs of locus-specific  $D'$  values. Recall that each locus-specific  $D'$  was based on observations of four possible cyto-nuclear genotype combinations (two mitochondrial variants combined with the common or binned minor nuclear alleles). This yields a 2x2 table of  $D'$  values in which each of the four entries has the same absolute value (two positive and two negative). The CNDm program summarizes this by reporting an absolute  $D'$  value for each year, and pooled across years (Table 2). These absolute values are not suitable, however, for testing the consistency of year-to-year associations since such comparisons would not be sensitive to changes in the sign of an allele-specific value of  $D'$  between years. Thus, when calculating the correlation mentioned above we consider for each locus the sign of the  $D'$  value specific to the association of the most common nuclear allele with the most common mitochondrial allele in 1993 and that same combination in 2008.

Year-to-year heterogeneity was also tested for statistical significance on a nuclear locus-specific basis by noting that  $D'$  is mathematically similar to a product-moment correlation coefficient in that both consist of a covariance standardized to range from -1 to 1. We used the Z-transformation approach suggested by SOKAL and ROHLF (2012, p. 573) for testing for heterogeneity among pairs of correlation coefficients (i.e., 1993 vs. 2008 for each of the 14 loci). Heterogeneity among the 14 pooled  $D'$  values was also tested using this method. The pairwise test employs a  $t$ -statistic while the test across all 14 values employs a  $\chi^2$  statistic (SOKAL and ROHLF 2012).

The significance of multiple comparisons of cyto-nuclear LD amongst markers within years, and across years, was assessed using a Benjamini-Hochberg correction for false discovery (BENJAMINI and HOCHBERG 1995). This approach has been suggested as a powerful analogue Bonferroni correction, and has recently been used for nuclear-nuclear LD comparisons (ANDRAS and EBERT 2013).

## RESULTS

Cyto-nuclear genotypes were obtained for 305 individuals (135 from the 1993 collection and 170 from the 2008 collection). For the *atp1* SNP, the common variant occurred at a frequency of 0.53 in 1993 and 0.58 in 2008. Table 1 presents a summary of global estimates of nuclear genetic diversity, population substructure ( $F_{IT}$ ) and a summary of nuclear LD for the 1993 and 2008 samples. Individual microsatellite markers were consistently polymorphic between years, with an overall range of 3-22 alleles per locus. Additionally, individual loci show a large range of  $H_0$  (0.11 – 0.84) and expected  $H_E$  (0.08– 0.90). Average  $F_{IT, 1993} = 0.353$  (range -0.041 – 0.780) and  $F_{IT, 2008} = 0.322$  (range 0.101 – 0.792), with 13 and 14 markers, respectively, exhibiting statistically significant ( $p \leq 0.05$ ) deviations from panmixia, corroborating deviations observed in  $H_E$ . In the 1993 and 2008 samples, there was an average of eight statistically significant non-random associations between each nuclear marker locus and the other 13 nuclear loci. Specific  $D'$  values and statistical significance of the deviation from random association for each marker pair is described in Suppl. Tables 1 and 2, for 1993 and 2008, respectively.

Table 2 presents the 1993, 2008, and pooled year absolute  $D'$  values between each of the 14 nuclear genes and the mitochondrial *atp1* SNP. Inspection of Table 2 shows that there was little change in the overall levels of cyto-nuclear linkage

disequilibrium between 1993 and 2008. The absolute magnitude of  $D'$  increased for eight loci and decreased for six (average locus-specific  $|D'|$  in 1993 = 0.266, average  $|D'|$  in 2008 = 0.263). The mean absolute pooled  $D'$  value was 0.230. Locus-specific pooled values vary by more than an order of magnitude, ranging from 0.019 for locus *slat72* to 0.487 for locus *slat33*. The relative rankings of locus-specific  $D'$  values were consistent between years (see Fig. 1), yielding a between-year Spearman Rank Correlation value (SOKAL and ROHLF 2012) of  $r_s = 0.653$  ( $p=0.01$ ). This view of sample-to-sample temporal consistency is reinforced by the fact that only five of fourteen locus-specific year-to-year heterogeneity tests were significant (Table 2). Given the year-to-year consistency in the locus-specific  $D'$  values, it seems warranted to test whether the wide range of such values noted above does, in fact, represent statistically significant among-locus heterogeneity. Application of the method of Sokal and Rohlf (2012) to the 14 pooled absolute  $D'$  values indicates statistically significant heterogeneity in locus-specific values ( $\chi^2 = 78.59$ ,  $df = 13$ ,  $p < 0.001$ ).

## DISCUSSION

The results demonstrate that cyto-nuclear associations can be highly variable depending on which nuclear locus is considered, and can average approximately 25% of the theoretical maximum as defined by allele frequencies. Furthermore, this locus-specific pattern persists across independent samples separated by more than seven generations. This raises three questions. 1) What is the reason certain loci are associated with different cytoplasmic backgrounds while others are not? 2) Why is this heterogeneity in locus-specific cyto-nuclear LD temporally stable? 3) What are the evolutionary consequences of this cyto-nuclear disequilibria?

Given that the nuclear and cytoplasmic genomes are not physically linked, in purely outcrossing species like *S. latifolia*, the accumulation of LD must be due either to epistatic selection or neutral demographic processes associated with the history of population structure (e.g. founder effects). These founder effects could include ongoing local events that occur regularly in metapopulations (KELLER *et al.* 2012; McCAULEY 1994; McCAULEY 1997; McCAULEY *et al.* 1996), or historical processes including the species post glacial history of range expansion (TAYLOR and KELLER 2007).

*S. latifolia* has a history of post-glacial expansion in the species native range that would favour the persistence of allelic associations. *S. latifolia* was introduced to North America relatively recently, ca. 200 years ago (TAYLOR and KELLER 2007). Following multiple, likely separate, introductions to both the Eastern and Western coasts of North America, the introduced range of *S. latifolia* expanded rapidly (KELLER *et al.* 2009; TAYLOR and KELLER 2007). Studies have concluded that during the invasion process, *S. latifolia* has maintained genetic diversity, though the distribution of genetic variation at various hierarchical levels has become reorganized (KELLER *et al.* 2012). Particularly striking is the decrease in among-regional scale allelic differentiation ( $F_{RT}$ ) in North America, as compared to the species' native range. However, at the scale of local populations, the level of genetic structure ( $F_{PT}$ ) is much more comparable between ranges ( $F_{PT,Nuclear-EU} = 0.147$ ,  $F_{PT,Nuclear-NA} = 0.131$ ;  $F_{PT,cpDNA-EU} = 0.498$ ,  $F_{PT,cpDNA-NA} = 0.382$ ). This scenario presents the hypothesis that ancestral population structure in the native range, and the LD that would result, is incompletely dissolved by the incomplete admixture occurring in the introduced range. Accordingly, the observed cyto-nuclear associations seen in the species' native range remain significant, with some evidence of a significant shift in cyto-nuclear associations, suggesting the action of admixture process to mix nuclear and organelle constituents (KELLER *et al.* 2012). This admixture, however, is not a uniform



process, and the fact remains that many ancestral allelic associations may persist. Given the results of Keller et al. (2012), the shifting cyto-nuclear associations detected here could be a local reflection of the admixture among European lineages following relatively recent establishment in North America. Though the focal organellar genome differed between the prior studies and our study (cpDNA vs. mtDNA), the high likelihood that each are inherited maternally should make their histories congruent, or nearly so.

The present study focuses on nuclear- and cyto-nuclear LD at a finer spatial scale than in previous studies of *S. latifolia*. Because the microsatellite loci and organelle locus utilized in the present study are assumed to be unlinked and neutral with respect to fitness, the observed measures of  $F_{IT}$  (Table 1), nuclear-nuclear LD (Table 1; Suppl. Tables 1 and 2), and cyto-nuclear LD likely reflect random sampling processes, such as arise within spatio-temporally distributed metapopulations (HANSKI and GAGGIOTTI 2004). Further analyses of the sort described will likely corroborate the observed patterns given observed levels of population subdivision (EDELAAR *et al.* 2011), though further research will be required to pinpoint the effect of individual spatio-temporal characters in determining the magnitude of cyto-nuclear LD (see below).

The finding that the overall level of cyto-nuclear LD has persisted over more than seven generations, even as some of the underlying allelic associations have shifted, could result from local population structure. Under panmixis and selective neutrality, one would expect cyto-nuclear LD to decay by about 50% each generation. Any deviation from panmixis will slow this process since nonrandom mating would limit the opportunity for cyto-nuclear genotypic mixing. The collection of individuals contributing to this dataset showed a marked deviation from panmixia as evidenced by a  $F_{IT}$  value of 0.35. This is not surprising, given that prior studies of *S. latifolia* (= *S. alba*) in this region of Virginia detected significant population structure as well ( $F_{ST} = 0.20$ ) when local

patches were used to define populations (McCAULEY 1994; McCAULEY *et al.* 1995). Further studies revealed additional structuring within those patches (McCAULEY 1997; McCAULEY *et al.* 1996). Since the present data set does not account for the fine-scale arrangement of individuals within patches (the sample size per patch is too small for meaningful estimates of very local population structure), our high value of  $F_{IT}$  probably reflects within and among patch divergence. Given that *S. latifolia* is dioecious and hence an obligate outcrosser, spatial structuring rather than self-fertilization must be responsible for these patterns.

In contrast to the moderate level of nuclear  $F_{ST}$  found in previous studies, values of  $F_{ST}$  based on cpDNA markers were very high - approximately 0.65 when defined at the patch level (McCAULEY 1995). Spatial sub-structuring of cpDNA was also found within populations (McCAULEY *et al.* 1996). Thus, the movement of maternally inherited cytoplasmic genes (in seeds only) must be considerably more restricted than the movement of nuclear genes (in seeds and pollen) (McCAULEY 1997).

An additional feature of *S. latifolia* populations is that local demes in this region of Virginia undergo frequent turnover and that these metapopulation dynamics influence population genetic structure (McCAULEY *et al.* 1995). This could affect the persistence of cyto-nuclear LD depending on the degree of mixing that accompanies local colonization events. SLATKIN (1977) modelled two modes of colonization – the “propagule pool” mode in which all  $k$  individuals contributing to a given colonization event are drawn from the same source, and the “migrant pool” mode in which a group of  $k$  colonists are drawn from a genetically representative sample of sources. WHITLOCK and McCAULEY (1990) define  $\phi$  as the probability that two alleles in a newly formed population were drawn from the same source population ( $\phi=1$  for propagule pool colonization). While these models were developed to consider the effect of colonization on  $F_{ST}$  (for a given number

of colonists,  $k$ ,  $F_{ST}$  tends to increase as  $\phi$  increases), the same logic will apply to allelic associations. Colonization events in which  $\phi$  approaches zero (migrant pool) would provide genetic mixing that would enhance the decay of allelic associations, whereas ancestral disequilibria are expected to persist under a propagule pool model. MCCAULEY *et al.* (1995) showed that local population structure was a consequence of these founding events. Moreover, they estimated that  $\phi$  was on the order of 0.80, meaning that colonization events in this *S. latifolia* metapopulation provide relatively little opportunity for the breakup of allelic associations. Taken together, the local spatial structuring of the nuclear and cytoplasmic genomes could clearly slow the rate of decay of nuclear and cyto-nuclear LD relative to an expectation based on panmixis. Finally, it is important to point out that Slatkin's migrant pools and propagule pools are themselves oversimplifications of any natural situation. For example, even if the source of founders are diverse so that old associations may dissolve after admixture, if the number of founders are few then new associations may form. This dynamic would generate the type of result that we observed, where the specific associations are shifting, but statistical associations of some form persist.

As an example of this latter process, we present a simple model. Within this model, we assume two bi-allelic loci. Let A and B be the alleles at the first locus, 1 and 2 be the alleles at the second locus. Now consider four differentiated source populations, fixed for genotypes A1, A2, B1, and B2, respectively. Each contributes 20, 60, 60, and 10 colonists, respectively, to a newly colonized study site. Under this scenario, A and 1 would be the most common alleles in this newly mixed population, yet a negative LD for the A1 genotype would be detected. Importantly, under this entirely neutral drift-like process, significant negative or positive LD values will be generated.

A persistent association among nuclear and organellar loci could be important for understanding many evolutionary processes. Persistent cyto-nuclear associations would result from strong cyto-nuclear epistasis for fitness. A well-known example of this is the interaction between mitochondrial cytoplasmic male sterility (CMS) and nuclear restorer sex-determining loci seen in many cases of gynodioecy (MCCAULEY and BAILEY 2009). The accumulation of positive epistatic interactions among nuclear and cytoplasmic genomes could be responsible for cyto-nuclear incompatibilities when those interactions are disrupted in crosses among lineages (MOYLE *et al.* 2004). This could have implications for the process of speciation (SAMBATTI *et al.* 2008). Persistent cyto-nuclear associations have implications for the coevolution of the two genomes, which could favour the transfer of genes from the mitochondrial to nuclear genome (BRANDVAIN and WADE 2009).

Theoretical models developed to predict when selection can effectively act on allelic associations among interacting genomes have focused on the term  $\theta$ , the joint probability that a pair of genes on each respective genome is identical-by-descent (IBD) (WADE and GOODNIGHT 2006). As the magnitude of  $\theta$  increases, the degree to which cyto-nuclear gene combinations are inherited together in transmission from parents to offspring also increases (WADE and GOODNIGHT 2006). Many models have explored how hybridization and patterns of non-random mating lead to different patterns of  $\theta$  and the resulting cyto-nuclear disequilibria (BASTEN and ASMUSSEN 1997). Additionally, population structure and the resulting increase in bi-parental inbreeding will enhance  $\theta$ , resulting in increased cyto-nuclear disequilibria, and hence the potential for selection on cyto-nuclear interactions (BRANDVAIN and WADE 2009).

The results presented here suggest that the magnitude of cyto-nuclear LD necessary for selection to act on cyto-nuclear interactions may be found in these *S.*

*latifolia* populations, at least for some nuclear/organellar combinations. These associations likely accumulated from the combined influence of invasion history and current metapopulation structure. Importantly, some of our data suggest that allelic associations could be transient, reduced by admixture and regenerated by founder effect. If that were a general result, then an instantaneous estimate of cyto-nuclear LD may over estimate the potential for longer-term co-evolutionary interactions among organellar genomes.

Few empirical studies have quantified the magnitude of intra-species cyto-nuclear LD (Latta *et al.* 2001), and none have quantified how the magnitude and among-marker variance in cyto-nuclear associations have changed over time. Estimating the magnitude and variance in allelic associations over time, and identifying the processes (both neutral and selective) that generate these associations, would contribute to our understanding of the causes and consequences of coevolution among eukaryotic organelles.

## **ACKNOWLEDGEMENTS**

Financial support for this work was provided by NSF grants DEB # 1050331 and 9221175 to D. E. M. and NSF DEB # 0919335 to D.R.T..

## LITERATURE CITED

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010  
Permanent genetic resources added to Molecular Ecology resources database 1  
August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ANDRAS, J. P., and D. EBERT, 2013 A novel approach to parasite population genetics:  
Experimental infection reveals geographic differentiation, recombination and  
host-mediated population structure in *Pasteuria ramosa*, a bacterial parasite of  
*Daphnia*. *Molecular Ecology* **22**: 972-986.
- ARNOLD, M. L., 2006 *Evolution Through Genetic Exchange*. Oxford University Press, Oxford  
UK.
- ASMUSSEN, M. A., J. ARNOLD and J. C. AVISE, 1989 The effects of assortative mating and  
migration on cytonuclear associations in hybrid zones. *Genetics* **122**: 923-934.
- ASMUSSEN, M. A., and A. SCHNABEL, 1991 Comparative effects of pollen and seed migration  
on the cytonuclear structure of plant-populations .1. maternal cytoplasmic  
inheritance. *Genetics* **128**: 639-654.
- BAKER, W. K., 1975 Linkage disequilibrium over space and time in natural populations of  
*Drosophila montana*. *Proceedings of the National Academy of Sciences of the  
United States of America* **72**: 4095-4099.
- BARRIÈRE, A., and M. A. FÉLIX, 2007 Temporal dynamics and linkage disequilibrium in  
natural *Caenorhabditis elegans* populations. *Genetics* **176**: 999-1011.
- BASTEN, C. J., and M. A. ASMUSSEN, 1997 The exact test for cytonuclear disequilibria.  
*Genetics* **146**: 1165-1171.
- BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and  
powerful approach to multiple testing. *Journal of the Royal Statistical Society:  
Series B (Statistical Methodology)* **57**: 289-300.

- BERNASCONI, G., J. ANTONOVICS, A. BIERE, D. CHARLESWORTH, L. F. DELPH *et al.*, 2009 *Silene* as a model system in ecology and evolution. *Heredity* **103**: 5-14.
- BRANDVAIN, Y., and M. J. WADE, 2009 The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. *Genetics* **182**: 1129-1139.
- DATTA, S., Y. X. FU and J. ARNOLD, 1996 Dynamics and equilibrium behavior of cytonuclear disequilibria under genetic drift, mutation, and migration. *Theoretical Population Biology* **50**: 298-324.
- EDELAAR, P., P. BURRACO and I. GOMEZ-MESTRE, 2011 Comparisons between  $Q_{st}$  and  $F_{st}$ -how wrong have we been? *Molecular Ecology* **20**: 4830-4839.
- EXCOFFIER, L., and H. E. L. LISCHER, 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.
- GAUNT, T. R., S. RODRIGUEZ, C. ZAPATA and I. N. M. DAY, 2006 MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *Bmc Bioinformatics* **7**.
- GILES, B., and J. GOUDET, 1997 Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist* **149**: 507-526.
- GOUDET, J., 2002 FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- HANSKI, I., and O. GAGGIOTTI, 2004 Metapopulation biology: past, present and future, pp. 3-22 in *Metapopulation Ecology, Genetics and Evolution*, edited by I. HANSKI and O. GAGGIOTTI. San Diego Academic Press.

- HEDRICK, P. W., 1987 Gametic disequilibrium measures - proceed with caution. *Genetics* **117**: 331-341.
- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- KELLER, S. R., D. R. SOWELL, M. NEIMAN, L. M. WOLFE and D. R. TAYLOR, 2009 Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist* **183**: 678-690.
- LANGLEY, C. H., K. ITO and R. A. VOELKER, 1977 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. *Genetics* **86**: 447-454.
- LATTA, R. G., Y. B. LINHART and J. B. MITTON, 2001 Cytonuclear disequilibrium and genetic drift in a natural population of ponderosa pine. *Genetics* **158**: 843-850.
- LEWONTIN, R. C., 1964 The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**: 49-67.
- LI, M. H., and J. MERILA, 2010 Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**: 600-610.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MCCAULEY, D. E., 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1995 The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution* **10**: 198-202.

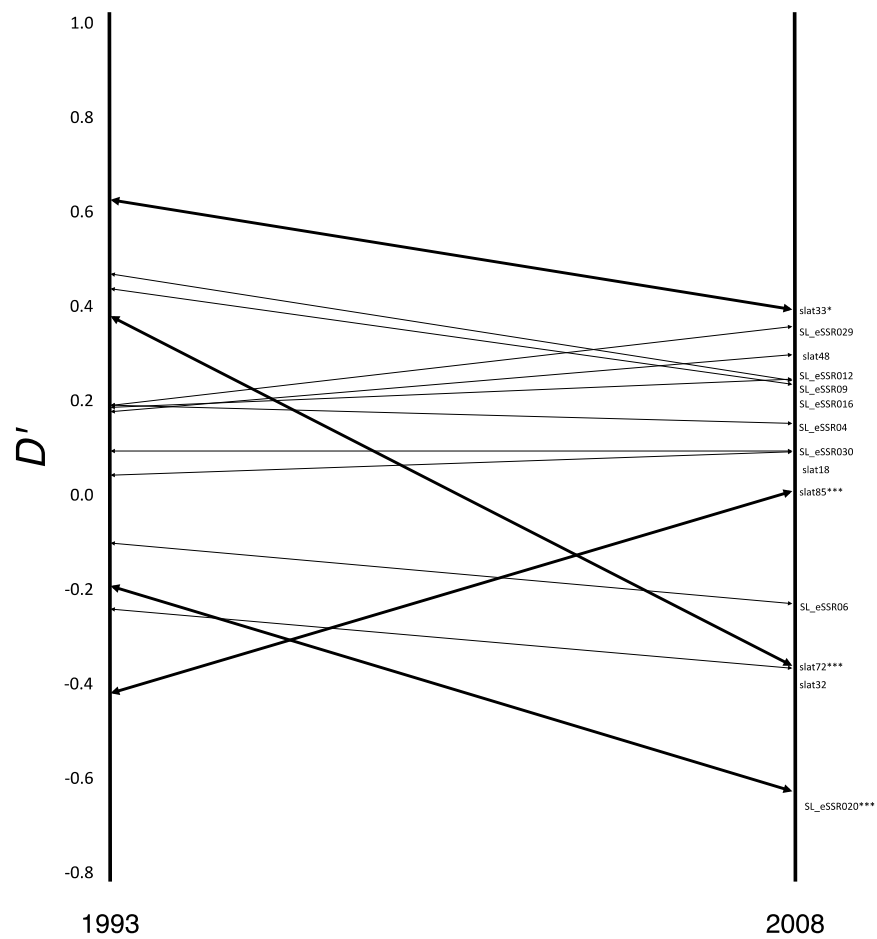


- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.
- MCCAULEY, D. E., and M. F. BAILEY, 2009 Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Annals of Botany* **104**: 611-620.
- MCCAULEY, D. E., and J. R. ELLIS, 2008 Recombination and linkage disequilibrium among mitochondrial genes in structured populations of the gynodioecious plant *Silene vulgaris*. *Evolution* **62**: 823-832.
- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.
- MCCAULEY, D. E., J. STEVENS, P. PERONI and J. RAVEILL, 1996 The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* **83**: 727-731.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- MOYLE, L. C., M. S. OLSON and P. TIFFIN, 2004 Patterns of reproductive isolation in three angiosperm genera. *Evolution* **58**: 1195-1208.
- NEI, M., and W. H. LI, 1973 Linkage disequilibrium in subdivided populations. *Genetics* **75**: 213-219.

- SAMBATTI, J. B. M., D. ORTIZ-BARRIENTOS, E. J. BAACK and L. H. RIESEBERG, 2008 Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters* **11**: 1082-1091.
- SLATE, J., and J. M. PEMBERTON, 2007 Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *Journal of Evolutionary Biology* **20**: 1415-1427.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**: 253-262.
- SLATKIN, M., 1994 Linkage disequilibrium in growing and stable populations. *Genetics* **137**: 331-336.
- SLATKIN, M., 2008 Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**: 477-485.
- SOKAL, R. R., and F. J. ROHLF, 2012 *Biometry: the principles and practice of statistics in biological research*. W.H. Freeman and Co., New York.
- STINCHCOMBE, J. R., and H. E. HOEKSTRA, 2007 Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158-170.
- TAYLOR, D. R., and S. R. KELLER, 2007 Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* **61**: 334-345.
- WADE, M., and C. GOODNIGHT, 2006 Cyto-nuclear epistasis: two-locus random genetic drift in hermaphroditic and dioecious species. *Evolution* **60**: 643-659.
- WHITLOCK, M. C., and D. E. MCCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.

ZAPATA, C., 2011 On the uses and applications of the most commonly used measures of linkage disequilibrium from the comparative analysis of their statistical properties. *Human Heredity* **71**: 186-195.

**Figure 1.** Standardized cyto-nuclear linkage disequilibrium ( $D'$ ) between fourteen nuclear loci and a SNP in the mitochondrial gene *atp1* found in populations of *Silene latifolia* sampled in 1993 and 2008 with a well-characterized metapopulation located in Giles County, VA. Values of  $D'$  can range from -1.0-1.0. Within the sampled plant populations,  $D'$  was found to range from -0.424-0.6211 in 1993 and -0.6316-0.3888 in 2008. There was a general trend for an overall decrease in  $D'$  between sampling periods, though only slat85 showed a significant (positive) change in overall cyto-nuclear  $D'$  based upon our t-test of samples (\* $p < 0.05$  after Benjamini-Hochberg correction). slat33, slat72, and SL\_eSSR20 all showed a significant decrease in  $D'$ .



**Table 1.** Global estimates of genetic diversity and population substructure within temporally separated samples.  $M$  = repeat size of the marker;  $N$  = number of observed alleles;  $p$  = allele frequency of the most frequent allele;  $H_O$  and  $H_E$  are observed and expected heterozygosity, respectively, of each sampled microsatellite locus;  $F_{IT}$  is the correlation of alleles within individuals relative to the metapopulation as a whole, and can be interpreted as an approximate measure of deviation from panmixia. Statistical deviations from panmixia are indicated by the symbols \*= $p < 0.05$ , \*\*= $p < 0.01$  and \*\*\* =  $p < 0.001$ ) and LD is an indication of the number of other nuclear loci for which each focal nuclear locus exhibits a non-random association, based upon an approximation of Fisher's exact test of random association and a  $p \leq 0.05$ .

<i>Locus</i>	1993							2008					
	<i>M</i>	<i>N</i>	<i>p</i>	$H_O$	$H_E$	$F_{IT}$	LD	<i>N</i>	<i>p</i>	$H_O$	$H_E$	$F_{IT}$	LD
SL_eSSR04	3	8	0.43	0.58	0.74	0.244***	7	8	0.55	0.57	0.65	0.143***	7
SL_eSSR06	3	15	0.32	0.50	0.82	0.405***	9	17	0.22	0.71	0.84	0.178***	10
SL_eSSR09	3	8	0.74	0.32	0.44	0.282***	8	8	0.75	0.39	0.43	0.101*	7
SL_eSSR12	2	12	0.22	0.72	0.87	0.185***	10	13	0.39	0.69	0.80	0.142***	11
SL_eSSR16	3	9	0.54	0.40	0.63	0.401***	10	6	0.54	0.48	0.62	0.245***	9
SL_eSSR20	3	5	0.94	0.12	0.11	-0.041	1	5	0.96	0.07	0.08	0.140*	1
SL_eSSR29	3	17	0.19	0.74	0.89	0.180***	11	22	0.22	0.84	0.90	0.078***	8
SL_eSSR30	3	17	0.43	0.69	0.77	0.111**	9	18	0.36	0.43	0.81	0.489***	11
slat_18	3	9	0.36	0.39	0.76	0.498***	10	11	0.43	0.37	0.76	0.519***	7
slat_32	4	7	0.34	0.44	0.78	0.455***	11	8	0.30	0.32	0.78	0.608***	11
slat_33	3	3	0.55	0.13	0.50	0.780***	12	2	0.65	0.11	0.46	0.792***	13
slat_48	3	5	0.45	0.27	0.66	0.597***	3	6	0.69	0.17	0.46	0.662***	7
slat_72	3	11	0.46	0.48	0.74	0.359***	9	13	0.19	0.38	0.80	0.582***	12
slat_85	3	17	0.44	0.54	0.76	0.303***	12	10	0.58	0.29	0.60	0.528***	8
<b>Overall</b>	<b>-</b>	<b>10.21</b>	<b>0.46</b>	<b>0.45</b>	<b>0.68</b>	<b>0.353***</b>	<b>8.71</b>	<b>10.5</b>	<b>0.49</b>	<b>0.42</b>	<b>0.64</b>	<b>0.377***</b>	<b>8.71</b>

**Table 2.** Normalized cyto-nuclear linkage disequilibrium ( $D'$ ) between fourteen nuclear loci and a SNP in the mitochondrial gene *atp1* found in populations of *Silene latifolia* sampled in 1993 and 2008, as well as a test for heterogeneity among pairs of correlation coefficients, or  $D'$ , between years. Statistical significance is indicated by the symbols \*= $p$ <0.05, \*\*=  $p$ <0.01 and \*\*\* =  $p$ <0.001. The  $t$ -statistic has infinite df, and so can be considered significant at 1.96.

<i>Locus</i>	$D'_{1993}$	$D'_{2008}$	$D'_{\text{pooled}}$	$t$ -value (1993 vs. 2008)
SL_eSSR04	0.1872	0.1476	0.1737*	0.347
SL_eSSR06	-0.1061	-0.2342*	-0.1751**	1.127
SL_eSSR09	0.1813	0.2412**	0.2161**	0.535
SL_eSSR12	0.4643***	0.2382**	0.3243***	2.216
SL_eSSR16	0.4328***	0.2303*	0.3258***	1.951
SL_eSSR20	-0.1964	-0.6316	-0.3721	8.042***
SL_eSSR29	0.1857	0.3529**	0.2841**	1.542
SL_eSSR30	0.0892	0.0897	0.0691	0.004
slat_18	0.0379	0.0879	0.071	0.428
slat_32	-0.2459**	-0.3709***	-0.3161***	1.180
slat_33	0.6211***	0.3888***	0.4868***	2.698*
slat_48	0.172	0.2939***	0.2415***	1.101
slat_72	0.375***	-0.366**	0.0193	6.634***
slat_85	-0.424***	0.0045	-0.1492*	3.897***

**Suppl. Table 1.** Nuclear-nuclear normalized linkage disequilibrium ( $D'$ ) for samples collected in 1993. Statistically significant associations amongst loci are indicated by the symbols \*=p<0.05.

<i>Locus</i>	SL_eSSR20	slat85	slat72	slat48	slat18	SL_eSSR30	slat33	SL_eSSR16	SL_eSSR09	SL_eSSR04	SL_eSSR09	SL_eSSR06	SL_eSSR12	slat32
SL_eSSR20	-													
slat85	0.20	-												
slat72	0.28	<b>0.48*</b>	-											
slat48	0.19	<b>0.22*</b>	0.20	-										
slat18	0.33	<b>0.29*</b>	<b>0.27*</b>	<b>0.24*</b>	-									
SL_eSSR30	0.51	<b>0.39*</b>	<b>0.40*</b>	0.13	0.25	-								
slat33	<b>0.35*</b>	<b>0.44*</b>	<b>0.32*</b>	0.07	<b>0.35*</b>	<b>0.38*</b>	-							
SL_eSSR16	0.33	<b>0.44*</b>	<b>0.39*</b>	0.11	<b>0.30*</b>	<b>0.40*</b>	<b>0.38*</b>	-						
SL_eSSR09	0.89	<b>0.54*</b>	0.27	0.12	<b>0.28*</b>	<b>0.42*</b>	<b>0.38*</b>	<b>0.40*</b>	-					
SL_eSSR04	0.16	<b>0.31*</b>	<b>0.27*</b>	0.18	0.21	<b>0.36*</b>	<b>0.33*</b>	0.16	0.20	-				
SL_eSSR09	0.38	<b>0.44*</b>	<b>0.41*</b>	0.16	<b>0.39*</b>	<b>0.40*</b>	<b>0.53*</b>	<b>0.44*</b>	<b>0.24</b>	<b>0.38*</b>	-			
SL_eSSR06	0.36	<b>0.43*</b>	<b>0.34*</b>	<b>0.29*</b>	<b>0.38*</b>	0.31	<b>0.27*</b>	<b>0.33*</b>	0.35	0.27	<b>0.35*</b>	-		
SL_eSSR12	0.33	<b>0.39*</b>	0.37	0.18	<b>0.33*</b>	<b>0.38*</b>	<b>0.56*</b>	<b>0.51*</b>	<b>0.37</b>	<b>0.39*</b>	<b>0.45*</b>	<b>0.36*</b>	-	
slat32	0.41	<b>0.31*</b>	<b>0.30*</b>	0.18	<b>0.22*</b>	<b>0.36*</b>	<b>0.29*</b>	<b>0.25*</b>	<b>0.21</b>	<b>0.24*</b>	<b>0.38*</b>	<b>0.24*</b>	<b>0.76*</b>	-



**Suppl. Table 2.** Nuclear-nuclear normalized linkage disequilibrium ( $D'$ ) for samples collected in 2008. Statistically significant associations amongst loci are indicated by the symbols \*=p<0.05.

<i>Locus</i>	SL_eSSR20	slat85	slat72	slat48	slat18	SL_eSSR30	slat33	SL_eSSR16	SL_eSSR09	SL_eSSR04	SL_eSSR09	SL_eSSR06	SL_eSSR12	slat32
SL_eSSR20	-													
slat85	0.43	-												
slat72	0.64	<b>0.32*</b>	-											
slat48	0.23	0.20	<b>0.27*</b>	-										
slat18	0.44	0.21	<b>0.32*</b>	0.20	-									
SL_eSSR30	0.37	0.23	<b>0.32*</b>	<b>0.35*</b>	<b>0.33*</b>	-								
slat33	<b>0.14*</b>	<b>0.46*</b>	<b>0.41*</b>	<b>0.23*</b>	<b>0.29*</b>	<b>0.30*</b>	-							
SL_eSSR16	0.25	<b>0.25*</b>	<b>0.35*</b>	0.17	0.19	<b>0.25*</b>	<b>0.34*</b>	-						
SL_eSSR09	0.59	<b>0.28*</b>	<b>0.37*</b>	0.12	<b>0.34*</b>	<b>0.24*</b>	<b>0.34*</b>	0.21	-					
SL_eSSR04	0.28	<b>0.19*</b>	<b>0.22*</b>	0.12	0.19	<b>0.33*</b>	<b>0.17*</b>	<b>0.18</b>	0.21	-				
SL_eSSR09	0.60	0.29	<b>0.36*</b>	<b>0.36*</b>	0.29	<b>0.19*</b>	<b>0.41*</b>	<b>0.38*</b>	0.18	<b>0.25*</b>	-			
SL_eSSR06	0.30	<b>0.24*</b>	<b>0.38*</b>	<b>0.28*</b>	<b>0.31*</b>	<b>0.31*</b>	<b>0.22*</b>	<b>0.24*</b>	0.38	<b>0.31*</b>	0.37	-		
SL_eSSR12	0.30	<b>0.25*</b>	<b>0.33*</b>	<b>0.34*</b>	<b>0.24*</b>	<b>0.27*</b>	<b>0.44</b>	<b>0.33*</b>	<b>0.25*</b>	0.25	<b>0.38*</b>	<b>0.37*</b>	-	
slat32	0.37	<b>0.19*</b>	<b>0.31*</b>	<b>0.25*</b>	<b>0.25*</b>	<b>0.29*</b>	<b>0.35*</b>	<b>0.25*</b>	<b>0.33*</b>	0.21	<b>0.37*</b>	<b>0.24*</b>	<b>0.29*</b>	-

## Chapter 7: Population Determinants of Migration and Evidence for Hard Selection during Colonization in a Plant Metapopulation <sup>7</sup>

---

<sup>7</sup> Formatted for submission as a coauthored manuscript: Fields, P.D., Austerlitz, F., Bollinger, A., and D. R. Taylor

**ABSTRACT**

The theory of interconnected populations (metapopulations) has shown that conclusions regarding ecological and evolutionary dynamics derived from single populations can be radically different when considered in a spatial context. However, metapopulation models are often stylized with parameters driving the system, such as migration, extinction, and colonization assumed to remain constant across populations and through time. Evolution in spatially structured populations will likely involve the evolution of the spatial ecology and genetic structure of the system itself, especially if there is inter-demic selection and/or the rates of migration, extinction, or colonization are genetically determined and subject to evolutionary change. Here we leverage a long-term demographic dataset, high-resolution genotyping, and a mixture of classical and recently derived statistical genetics to estimate the fundamental parameters of a natural metapopulation, but also identify and disentangle the ecological and genetic factors that influence those parameters and hence the evolution and spatial structure of the system. Population genetic structure is pervasive, but quantitatively different across sections of the metapopulation. Among population migration was generally low, with a large variance. The factors affecting migration between populations were consistent with theoretical expectations, including pair-wise population distance, as well as differences in population size and age. New colonists often occurred at distances much greater than the average dispersal distance amongst established populations, though not in all sections. Most significantly, we uncovered evidence for hard selection determining colonization success, as modulated through a disproportionately larger proportion of colonists originating from populations of higher genetic diversity, controlling for population size.

## INTRODUCTION

The theory of interconnected populations (metapopulations) has shown that spatial structure can have profound effects on evolution, but there is a paucity of field data from spatially distributed populations over extended periods of time. Theoretical treatments of this fundamental issue in biology, that the process of evolution is affected by the reality that populations are distributed in space, are far beyond experimental studies because of the size and long-term nature of the datasets that are required.

A major focus of metapopulation research has been to elucidate the effect of extinction and colonization events on the creation and maintenance of neutral genetic diversity and population genetic structure (GILPIN 1991; MARUYAMA and KIMURA 1980; McCAULEY 1991; McCAULEY 1994a; SLATKIN 1977; WADE and McCAULEY 1988; WHITLOCK and McCAULEY 1990). In the models of Wright (1931), where populations are interconnected yet stable, genetic differentiation ( $F_{ST}$ ) is generated by drift or divergent selection, and diminished by gene flow. In metapopulations where demes experience extinction and recolonization, founder effects can be a powerful structuring mechanism (SLATKIN 1977; WHITLOCK and McCAULEY 1999), and population differentiation additionally depends on population size ( $N$ ) extinction ( $e$ ) and colonization ( $c$ ) rates, the number of founding propagules ( $k$ ), as well as the fraction of demes from which the colonists come ( $\phi$ ) (WADE and McCAULEY 1988; WHITLOCK and McCAULEY 1990). Specifically, an explanation of how the aforementioned metapopulation dynamics will affect population structure was described in WHITLOCK and McCAULEY (1990) and WHITLOCK (1992) as:

$$F_{ST} = \frac{[1 - e + e\phi(1 - 1/2k)]/(2N + e/2k)}{1 - [1 - e + e\phi(1 - 1/2k)](1 - m)^2(1 - 1/2N)} \quad \text{Eq (1)}$$

Direct manipulation of Eq. 1 has led to a number of useful inequalities that are useful for predicting the effect of particular metapopulation processes, such as colonization, in increasing or decreasing total population differentiation (GILES and GOUDET 1997). Each parameter has been shown to have a substantive effect on the genetic diversity distributed within and among demes, as well as on the potential outcome of evolution (PANNELL 2003; PANNELL and CHARLESWORTH 2000).

Population structure can, in turn, alter the outcome of selection in metapopulations. Consider the fate of a deleterious recessive allele. In structured populations, recessive alleles will be more exposed to selection because they experience a higher frequency, and homozygosity, in a subset of demes. This results in inbreeding depression over the short term, and perhaps a reduced genetic load (via more efficient purging of deleterious recessives) over the longer term (THRALL *et al.* 1998; WHITLOCK 2002). If population structure is severe, fixation of deleterious recessive alleles may occur, a process that can only be reversed by inter-demic processes such as genetic rescue via migration (WILLI and FISCHER 2005; WILLI *et al.* 2005). More generally, population structure influences the outcome of natural selection whenever individuals interact ecologically or behaviorally with a local subset of conspecifics relative to the total population (MCCAULEY 1994b). Individual fitness is then a function of local rather than global frequency, and the average fitness of a phenotype or genotype is better understood as a combination of the global frequency of the trait and the among group variance in the local frequency (MCCAULEY and TAYLOR 1997; WILSON 1980). In metapopulations, the redistribution of genetic variance from within demes to among demes will interact with the degree to which different

populations vary in the contribution to future generations (hard vs. soft selection) to make evolution in structured populations a multi-level selection process (WHITLOCK 2002).

Theoretical and empirical studies have indicated that multilevel selection in structured populations may be important in explaining diverse biological phenomena, including the evolution of social behavior (FRANK 1998; HAMILTON 1996; HAMILTON 1964; WILSON 1997), intragenomic conflict (HURST *et al.* 1996; INGVARSSON and TAYLOR 2002; TAYLOR *et al.* 2002), reduced pathogen virulence (KERR *et al.* 2006; O'KEEFE and ANTONOVICS 2002), sex ratio evolution (MCCAULEY and TAYLOR 1997; OLSON *et al.* 2005), and the origins of multi-cellularity (MICHOD 1997; MICHOD 1999; SZATHMARY and SMITH 1995). Recent conceptions of multilevel selection theory have broadened its applicability by recognizing that the effects of group membership on individual fitness, so-called contextual effects (HEISLER and DAMUTH 1987), is a group-level effect that can be distinguished from a group's potential to give rise to other groups (HEISLER and DAMUTH 1987; OKASHA 2004; so-called multilevel selection 1 (MS1) and MS2, respectively). Empirical studies have quantified multilevel selection within a generation (ASPI *et al.* 2003; BIJMA *et al.* 2007a; BIJMA *et al.* 2007b; DONOHUE 2003), but the fundamental question remains how these different forms of multilevel selection are acting to drive organic evolution (BIJMA and WADE 2008).

Previous work has captured the signature of metapopulation effects through an analysis of  $F_{ST}$  in age-structured demes (GILES and GOUDET 1997; INGVARSSON 1998; MCCAULEY *et al.* 1995). In the first of these studies, MCCAULEY *et al.* (1995), working in a metapopulation of the angiosperm *Silene latifolia*, and showed an elevated  $F_{ST}$  in newly colonized demes, consistent with the model where founder effect creates population structure, with gene flow eroding that initial genetic structure as demes age (i.e., high  $e$  and

$c$ , moderate  $m$ , but low  $k$  and  $\phi$ ).

While these aforementioned methods have been very powerful in describing some ecological and evolutionary processes, their reliance upon summary statistics will limit exact inference (e.g. identification of the exact parentage of new colonists, which allows for the identification of such evolutionary processes such as hard vs. soft selection). However, beginning with in the early 1990s with the introduction of methods such as genetic-stock identification (SMOUSE *et al.* 1990), computing-intensive methods paired with powerful, model based approaches were developed which allowed for additional insights beyond summary statistics such as  $F_{ST}$  (GAGGIOTTI *et al.* 2004; GAGGIOTTI *et al.* 2002). Of particular relevance is the so-called  $F$ -model. The  $F$ -model, which is a likelihood-based approach that defines  $F_{ST}$  as a parameter of the full distribution of allele frequencies (BALDING and NICHOLS 1995; BHATIA *et al.* 2013; GAGGIOTTI and FOLL 2010; KARHUNEN and OVASKAINEN 2012; NICHOLSON *et al.* 2002) accommodates differences in population size and migration rates across a species range (GAGGIOTTI and FOLL 2010), and thus has advantages over  $F_{ST}$  that estimates a single “global” value of differentiation.

In the present study, we utilize a combination of classical and recently derived statistical genetics methods to parameterize the models of WADE and MCCAULEY (1988); WHITLOCK and MCCAULEY (1990) and derive novel insights concerning the role of spatiotemporal metapopulation dynamics generating these particular summaries. We also show evidence of hard selection, or MS2, modulated through the differential colonization success of populations that are both larger and have a higher degree of genetic diversity.

## **MATERIALS AND METHODS**

## STUDY ORGANISM

*Silene latifolia* Poir. (= *S. alba*, Caryophyllaceae) is a short-lived perennial plant that is broadly used as a model system for studying sex determination and sex chromosome evolution, host-pathogen dynamics, species invasion, organelle evolution, sexual dimorphism, sex ratio evolution, and evolution in structured populations (BERNASCONI *et al.* 2009).

Our populations of *S. latifolia* are located within a metapopulation in Giles and Craig Counties, Virginia, USA (Figure 1). This region has been the subject of a 20+ year study of population dynamics and genetic structure in more than 800 spatially distributed populations (ANTONOVICS *et al.* 1994; MCCAULEY 1994a; MCCAULEY 1997; MCCAULEY *et al.* 1995; RICHARDS 2000; RICHARDS *et al.* 1999; RICHARDS *et al.* 2003). An annual census of an approximately 25x25-km area adjacent to Mountain Lake Biological Station has been conducted since 1988, recording the location, and gender, of *S. latifolia* individuals along ~150km of predominantly roadside habitat. The structure of the data and how it was collected is reported in Antonovics *et al.* (1994). Briefly, the roadside habitat is divided up into ~40m segments referred to as psilons. Which psilons are occupied and the number of plants in each occupied psilon are recorded annually. Important phase transitions, such as extinction and colonization events, are confirmed with a second census during the same season. Time since colonization (or population age) is based on the year plants were first observed in a given site. We identified extinction as the disappearance of plants from a psilon for a single year, though a previous study in the focal metapopulation has indicated the presence of seed bank with a possible viability of approximately four years (PERONI and



ARMSTRONG 2001). The census data provide the demographic data, spatial relationships among populations, and extinction/recolonization dynamics used in this study.

## GENOTYPE SAMPLING

We sampled plants from 77 and 97 spatially distinct populations during peak flowering in the summers of 2008 and 2010, respectively, spanning  $\sim 1/3$  of the metapopulation (Figure 1). In the summers of 2009 and 2011, individuals identified as colonists were sampled. We collected leaf tissue from every plant in the population, or up to 50 individuals in the largest populations, and stored the leaves in silica gel (Sigma). Genomic DNA was extracted and amplified following established microsatellite techniques for *S. latifolia* DNA was extracted from leaf tissue using the method described in (KELLER *et al.* 2012). We genotyped each individual plant at between 16 and 19 microsatellite loci. Our microsatellites are derived from multiple sources (ABDOULLAYE *et al.* 2010; JUILLET *et al.* 2003; MOCCIA *et al.* 2009; TEIXEIRA and BERNASCONI 2007). PCR amplification was conducted using published methods for each marker. PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following five minutes of denaturing at 95 °C, fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems). Alleles were binned using the software TANDEM (MATSCHINER and SALZBURGER 2009).

## POPULATION CHARACTERISTICS

We estimated characteristics of populations that were likely to play a role in metapopulation dynamics: population size, population isolation (a characteristic that will influence gene flow between populations, as well as the magnitude of drift a population is subject to), and population age (a factor that estimates the recency of founder effects and opportunities for subsequent migration).

We estimated recent population size based upon the number of plants (both flowering and vegetative) occupying population site at the time of tissue collection (summer 2008 and 2010). Pair-wise distances between populations were calculated using a network based upon the public roadway system, using ArcGIS (ESRI) Network Analyst tool. Given the mountain-valley geographic topology of the area, this network-based approach is more appropriate than standard Euclidean distances in order to predict likely routes of the predominant pollinators, noctuid moths. As such, larger isolation scores are indicative of a decrease in an individual population's probability of receiving migrants, whether through seeds or pollen.

To estimate population age, we assumed sites unoccupied for a single year as extinct. This reflects the simplifying assumption that recolonization from neighboring sites or from the seed bank involve similar bottlenecks of genetic diversity. Following extinction and re-colonization, age is calculated as the number of years a site has been occupied up until the time of collection. Given this operational definition of population age, individual sites could vary in age from one to 23 years (the extent of the demographic census started in 1988). Populations in the age class of 23 years might reasonably be considered as a heterogeneous grouping of extant populations given the limited compass of the current

long-term census. Individuals collected from sites that have been unoccupied from 1 to the full 23 years are assumed to be colonists.

## STATISTICAL ANALYSIS

We calculated population genetic summaries of genetic diversity, as well as estimates of genetic (sub)-structure via hierarchical  $F$ -statistics using the software GenoDive version 2.0b21 (MEIRMANS and VAN TIENDEREN 2004), with significant deviations from 0 assessed with 10,000 permutations and  $\alpha = 0.05$ .

We use a permutation-based Mantel test (MANTEL 1967) in order to determine the presence of isolation-by-distance (IBD). The resemblance matrix of pair-wise  $F_{ST}$  values were estimated using GenoDive and the resemblance matrix of pair-wise distance values was derived from the ArcGIS approach above. The tests utilized a total of 10,000 permutations, with significance assessed by a one-tailed test.

More exact determination of long-term dispersal dynamics were made by analyzing distance data and the pair-wise  $F_{ST}$  using SPAGeDi (HARDY and VEKEMANS 2002) using a nested AMOVA approach (MICHALAKIS and EXCOFFIER 1996; WEIR and COCKERHAM 1984). From the slope of the regression function assigned to the  $F_{ST}$  value plotted against distance, distance between interbreeding populations was found. Population pairs were assigned to distance classes every 500 meters of distance (i.e., less than 500 meters, between 500 and 1000 meters, between 1000 and 1500 meters, etc.).  $F_{ST}$  was calculated for all pairs of populations within pairwise distance intervals and the slope  $b$  of the regression of  $F_{ST}/(1 - F_{ST})$  against geographic distance was found. Since the environment was linear, average distance between interbreeding populations was defined with:

$$b = \frac{1}{4N\delta^2}$$

Slope was raised to -1 and divided by 4 to get  $N\delta^2$ . Average population size  $N$  was calculated as the average of each population in each section for 2010 and was used as a proxy for effective population size, and  $\sigma^2$  was found for each section in each year.

We estimated short-term migration amongst populations using the program BayesAss v. 3.03 (WILSON and RANNALA 2003). We subdivided the populations into three sections of the metapopulation, and analyzed the sections separately. These sections are isolated geographically, with migration between sections being very unlikely. A total of three runs per year were done, each using 50,000,000 Markov Chain Monte Carlo (MCMC) iterations and a burn-in of 500,000 iterations, and a thinning interval of 100, each with a different starting seed. In order to obtain appropriate mixing conditions, as determined by acceptance rate, in the MCMC chain, we modified the allele frequency, inbreeding coefficient, and migration rate parameters as per the BayesAss v. 3.03 manual suggestion. Chain convergence was assessed using the program Tracer v. 1.5 (RAMBAUT and DRUMMOND 2009).

To determine whether the population factors determined the magnitude of gene-flow between populations we applied the hierarchical Bayesian approach of FAUBET and GAGGIOTTI (2008) with the program BIMr. Each run consisted to 20,000 MCMC iterations of short pilot runs in order to tune proposal distributions to obtain acceptance rates between 25 and 45%; 10,000 iterations were discarded as burn-in; and a sample size of 10,000 iterations, with a thinning interval of 100, were retained to estimate posterior distributions

of model fit and parameter estimation (a total of 1,030,000 iterations per run) (FAUBET and GAGGIOTTI 2008). In order to control for MCMC-based variation in model convergence, we considered seven independent replicates of each run, and following FAUBET *et al.* (2007; also references therein), chose the run with the lowest deviance for estimation purposes. As reported in FAUBET and GAGGIOTTI (2008), deviance in migrant assignment, or  $D_{assign}$ , is calculated as  $-2\overline{\log Pr(M|S; m)}$ .

We used the program GENECLASS2 (PIRY *et al.* 2004) to assign colonists to their likely source populations. We used the RANNALA and MOUNTAIN (1997) Bayesian criterion, an assignment threshold of 0.05, and the L\_home/L\_max likelihood computation. Potential sources in individual analyses were limited to focal sections, with colonists in 2008-2009 and 2010-2011 assigned to 2008 and 2010 sources, respectively.

To test whether colonization was distance dependent, we calculated the pair-wise distance between the assigned sources and colonists across the 2008-2011 samples. The average distance between colonists for the assigned parents was compared to the average for the section as a whole. Significance was assessed using a two-sided Wilcoxon test in R v. 2.15.3 (R DEVELOPMENT CORE TEAM 2012). Finally, we used a General Linear Model with a Poisson error distribution to determine if population size or genetic diversity (our measure is independent of population size) have an effect on the probability of a given source giving rise to successful colonists in R.

## RESULTS

Global population differentiation as measured by  $F_{ST}$  was consistently high and significant across specific molecular markers as well as overall (Table 1). Additionally,

there was a large degree of substructure as measured by  $F_{IS}$ , which is consistent with previous work.

Mantel tests showed inconsistent IBD across the metapopulation. Significant IBD was identified in two of the three metapopulation sections ( $p < 0.05$ ) and almost so in the last ( $p = 0.078$ ).  $F_{ST}$  increased significantly with distance in two of the metapopulation sections, but not consistently so (Figure 2). The axial variance of dispersal distances estimated from the coefficient of regression between  $F_{ST}$  and distance ranged from 17 meters to 50 meters (Table 2).

Short-term migration rates between populations varied greatly amongst pairs of populations (0.001 – 0.15), though estimates were relatively consistent between years (Supplemental Tables 1-6). Overall migration among populations had a mean range between 0.007 and 0.015 (Table 3). Importantly, we were able to determine the metapopulation parameters responsible for the observed migration rates using BIMr. Models including factors related to pair-wise spatial distance, population size, and population age showed a higher posterior probability than the null model including an intercept (Table 4). As expected population isolation was part of the highest posterior model in all cases, showing a negative regression coefficient in 5 of the 6 datasets, indicating populations that are more isolated from one another share fewer migrants (Table 5). Similarly, population size occurs in the most probable model in all 4 of 6 datasets, and depending upon the section analyzed, showed both a positive and negative regression coefficient. In the positive case, the applied approach would indicate that larger differences in population size are leading to increased geneflow, potentially suggestive of a source-sink dynamic. A negative regression coefficient is indicative of geneflow happening

predominantly amongst populations of approximately the same size. Finally, population age was important in the highest posterior models. Positive regression coefficients are indicative of higher migration between populations of largely different age.

From 2008 to 2012, we observed roughly 100 new colonized sites in the roughly one third of the metapopulation that comprised our study area. This represented approximately 300 new individual colonists (Table 3). The mean number of colonists per colonization, or  $k$ , event varied from 1.67 to 4.20, and the number of sources,  $\phi$ , per colonist 1.67 to 2.67, suggesting a colonization model roughly intermediate between the propagules and migrant pool models of colonization. Figure 3 shows the distribution of weighted average of the pairwise dispersal distances between colonizers and their parental populations assigned by GeneClass2. We observed that the frequency of dispersal events decreased with distance for the three sections in 2008; however we did not observe such a pattern in 2010. In general, the average distances were quite high, ranging between 850m to 2280m in each section (Table 6). We assessed whether the populations that gave rise to colonists differed significantly from non-source populations for geographical distance. In only two cases, section 2 (2008-2009) and section 9 (2010-2011), was there evidence for a predominance of long-distance dispersal. We found that population size ( $p = 0.0205$ ;  $\chi^2 = 5.37$ ) and population genetic diversity ( $p = 0.0053$ ;  $\chi^2 = 7.77$ ) both had a significant, positive effect on colonization success (Figure 4).

## DISCUSSION

The presented analysis presents a number of results consistent with previous theoretical analyses and empirical studies, thereby lending a natural connection between

these previous population genetic analyses and our analytical extensions. Most significantly, and in addition to population size, we showed that population genetic diversity ( $H_s$ ), an estimator that takes into account sample size, significantly contributed to population success. This dynamic, while it has been observed in situations of heterozygosity fitness correlations (HFCs), has not been observed at the level of populations (DAVID 1998). As such, the presented analysis suggests a role of hard selection (WADE 1985) in driving colonization processes in *S. latifolia*.

Populations were highly genetically structured, and consistently so between sections and years (Table 1). This particular result was not surprising, as it was previously observed by McCauley et al. (1996). However, these previous studies utilized a different type of molecular marker (allozyme) and focused on a much smaller sample of the metapopulation as whole. In the present study, we can observe that while population structure is consistently observed across the metapopulation, subsection 6 actually has a slightly larger degree of population structure, which might lead to a heterogeneous distribution of genetic load across the metapopulation (WHITLOCK 2002).

We also observed a signature of IBD in the metapopulation, though standard Mantel tests suggested significant IBD in two of the three focal metapopulation sections. The utility of Mantel tests for understanding IBD has recently been questioned (GUILLOT and ROUSSET 2013). By plotting pairwise  $F_{ST}$  against distance class of individual populations, we are able to see inconsistencies in the pairwise  $F_{ST}$  distance relationship, potentially exposing problems that might arise in naïve application of standard Mantel tests (Figure 2). Inferences based on parameters like  $F_{ST}$  concern the long-term history of the populations, as the value reached by these parameters results from the equilibrium between genetic



drift and migration (see e.g. HUDSON 1998).  $F_{ST}$  generally increased with the pairwise distances between populations, as expected under an isolation-by-distance pattern (ROUSSET 1997), reaching quite high values ( $\sim 0.3$ ). From these values, we estimated dispersal distances that ranged from 17 to 43 meters. This is consistent with the results found by RICHARDS *et al.* (1999), where populations separated by 20 meters were almost panmictic, while populations separated by 80 meters were nearly isolated. This pattern likely results from limited seed and pollen dispersal within and among established populations. Regarding pollen dispersal, while moths can travel long distances, previous experimental work has shown that the pollen received while visiting a male flower is deposited on the next three female flowers (LABOUCHE and BERNASCONI 2010). This limits the extent of dispersal, as shown also by a paternity analysis (AUSTERLITZ *et al.* 2012). Also seed dispersal between established populations is likely to be limited in this species were seeds are gravity dispersed. An isolation by distance study considering both X-linked and autosomal markers showed also low levels of dispersal among established populations in Europe, both through seeds and pollen (BARLUENGA *et al.* 2011).

We were also able to estimate recent dispersal by applying the methods of BayesAss and BIMr (FAUBET and GAGGIOTTI 2008; WILSON and RANNALA 2003). Average pairwise dispersal was highly variable (Suppl. Tables 1-6). However, by applying BIMr we were able to estimate the spatiotemporal factors responsible for this variation. The most consistent factor affecting migration rates was pairwise distance (Table 4 and 5). While this result is generally consistent with theoretical expectations, there are some applications of BIMr that did not find this particular result (FAUBET and GAGGIOTTI 2008). More interestingly, we were able to identify particular samples in which the sign of the regression on factors affecting

migration shifted in sign, e.g. when migration took place disproportionately more between populations of different ages and sizes. This dynamic is likely to arise in situations of source-sink dynamics (PULLIAM 1988) and/or genetic rescue (RICHARDS 2000). By identifying the particular sources linking populations together, further experimentation would allow one to disentangle which of these mutually exclusive hypotheses is responsible for the observed patterns.

We observed generally contrasting patterns between the long-term level of gene flow, as inferred from the level of genetic differentiation among the established populations, and the level of dispersal events occurring during the colonization of empty areas, as inferred from the GeneClass2 analysis. Conversely, when inferring with GeneClass2 the probability of origin of the individuals growing in newly colonized patches, we clearly did not observe short dispersal distances during these colonization events. There was clearly no decrease of this probability of origin of a plant from a given patch with the distance from this patch for all three sections in 2010, as long-distance dispersers are common in all three sections during this year (Figure 3).

As established populations are usually clustered on the landscape, colonist seeds necessarily travel farther than their non-colonist counterparts – i.e., they have to leave their populations and surrounding populations to find their way to unoccupied territory – but such an explanation would still predict a downward sloping curve of dispersal distances, in contradiction with the results found in this study. Instead, colonization events involved seeds that traveled an unusually long distance, implying a different mechanism of seed dispersal than is used in short distances.

These results raise questions about the mechanics of seed dispersal during colonization events. Since *S. latifolia* seeds are normally distributed by gravity, and dispersal distances among established populations much lower than those found in colonization events, other mechanisms were likely at work during colonization events. ANTONOVICS (2004) suggested that recurrent extinction events must be compensated for with regular long distance colonization events. It is likely that human activity helps in seed dispersal (BARLUENGA *et al.* 2011), with road grading and mowing as possible mechanisms (RICHARDS *et al.* 1999). Considering how these populations were all located along the same stretches of road, vehicles may have played a part in dispersal. Such questions should clearly be further investigated through more direct methods, e.g. parentage analyses, but this would require a rather exhaustive sampling of potential parents among the potential parents of colonist individuals, a difficult task for this abundant species.

In addition to the distance analysis, we also showed that other factors, e. g. population size, contributed to population colonization success. This pattern results probably first mechanically from the fact that a larger population size means that there are more individuals producing seeds. Moreover, pollinators are attracted to populations with a larger number of flowers, and they visit more flowers while in larger populations (RICHARDS *et al.* 1999). Both of these factors increase the chance that a flower will be pollinated, which increases the chances that an individual will bear fruit and thus be able to produce seeds that can colonize empty patches.

Recent experiments have given insight into the role of multilevel selection in “everyday” natural plant populations. DONOHUE (2003) showed that sibling groups of the Great Lakes Sea Rocket (*Cakile edentula*) show a higher fitness (number of seed capsules)

on average than when the same families were arrayed with unrelated individuals. Both STEVENS *et al.* (1995) and KELLY (1996) show that certain phenotypic arrangements of jewelweed (*Impatiens capensis*) result in opposing group and individual level fitness components. ASPI *et al.* (2003) showed that individual and group selection were acting synergistically in a metapopulation of *S. tatarica* to increase plant height (tall plants receive more pollinators both at the group and individual level), while the two selective levels were acting in opposition in regards to the number of stems (herbivores tend to be attracted to more dense stands of plants). Though these experiments have suggested the role of MS1, ours is one of the first show evidence for MS2, acting via hard selection, in a natural metapopulation (though see WADE and GOODNIGHT 1991; WADE 1977 for experimental manipulation of MS2 ; and see WADE 2013 for a theoretical description of Wright's Phase III of the Shifting Balances Process). The presented results suggest there may be an important role for multi-level selection in “everyday” metapopulations.

**ACKNOWLEDGEMENTS** PF is funded by the National Science Foundation (NSF) DEB #0919335 to DT and JA, and NSF-OISE# 1139716 to DT and PF.

## LITERATURE CITED

- ABDOULLAYE, D., I. ACEVEDO, A. A. ADEBAYO, J. BEHRMANN-GODEL, R. C. BENJAMIN *et al.*, 2010  
Permanent genetic resources added to Molecular Ecology resources database 1  
August 2009-30 September 2009. *Molecular Ecology Resources* **10**: 232-236.
- ANTONOVICS, J., 2004 Long term study of a plant-pathogen metapopulation, pp. 471-488 in  
*Ecology, genetics and evolution of metapopulations*, edited by I. HANSKI and O.  
GAGGIOTTI. Academic Press.
- ANTONOVICS, J., P. THRALL, A. JAROSZ and D. STRATTON, 1994 Ecological genetics of  
metapopulations: the *Silene-Ustilago* plant-pathogen system, pp. 146-170 in  
*Ecological genetics*, edited by L. REAL. Princeton University Press, NJ.
- ASPI, J., A. JÄKÄLÄNIEMI, J. TUOMI and P. SIIKAMÄKI, 2003 Multilevel phenotypic selection on  
morphological characters in a metapopulation of *Silene tatarica*. *Evolution* **57**: 509-  
517.
- AUSTERLITZ, F., G. GLEISER, S. TEIXEIRA and G. BERNASCONI, 2012 The effects of inbreeding,  
genetic dissimilarity and phenotype on male reproductive success in a dioecious  
plant. *Proceedings of the Royal Society B-Biological Sciences* **279**: 91-100.
- BALDING, D. J., and R. A. NICHOLS, 1995 A method for quantifying differentiation between  
populations at multi-allelic loci and its implications for investigating identity and  
paternity. *Genetica* **96**: 3-12.
- BARLUENGA, M., F. AUSTERLITZ, J. A. ELZINGA, S. TEIXEIRA, J. GOUDET *et al.*, 2011 Fine-scale spatial  
genetic structure and gene dispersal in *Silene latifolia*. *Heredity* **106**: 13-24.
- BERNASCONI, G., J. ANTONOVICS, A. BIERE, D. CHARLESWORTH, L. F. DELPH *et al.*, 2009 *Silene* as a  
model system in ecology and evolution. *Heredity* **103**: 5-14.

- BHATIA, G., N. J. PATTERSON, S. SANKARARAMAN and A. L. PRICE, 2013 Estimating and interpreting  $F_{st}$ : the impact of rare variants. *Genome Research*.
- BIJMA, P., W. A. MUIR and J. A. M. VAN ARENDONK, 2007a Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* **175**: 277-288.
- BIJMA, P., W. M. MUIR, E. D. ELLEN, J. B. WOLF and J. A. M. VAN ARENDONK, 2007b Multilevel selection 2: Estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**: 289-299.
- BIJMA, P., and M. J. WADE, 2008 The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *Journal of Evolution Biology* **21**: 1175-1188.
- DAVID, P., 1998 Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**: 531-537.
- DONOHUE, K., 2003 The influence of neighbor relatedness on multilevel selection in the Great Lakes sea rocket. *American Naturalist* **162**: 77-92.
- FAUBET, P., and O. E. GAGGIOTTI, 2008 A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* **178**: 1491-1504.
- FAUBET, P., R. WAPLES and O. GAGGIOTTI, 2007 Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates, pp. 1149-1166 in *Mol Ecol*.
- FRANK, S. A., 1998 *Foundations of social evolution*. Princeton University Press, Princeton.
- GAGGIOTTI, O. E., S. P. BROOKS, W. AMOS and J. HARWOOD, 2004 Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Molecular Ecology* **13**: 811-825.

- GAGGIOTTI, O. E., and M. FOLL, 2010 Quantifying population structure using the F-model. *Molecular Ecology Resources* **10**: 821-830.
- GAGGIOTTI, O. E., F. JONES, W. M. LEE, W. AMOS, J. HARWOOD *et al.*, 2002 Patterns of colonization in a metapopulation of grey seals. *Nature* **416**: 424-427.
- GILES, B., and J. GOUDET, 1997 Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist* **149**: 507-526.
- GILPIN, M., 1991 The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**: 165-175.
- GUILLOT, G., and F. ROUSSET, 2013 Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**: 336-344.
- HAMILTON, W., 1996 Hamilton's Rule: The Genetical Evolution of Social Behaviour, I and II, pp. 1-21.
- HAMILTON, W. D., 1964 The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* **7**: 1-16.
- HARDY, O., and X. VEKEMANS, 2002 SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618-620.
- HEISLER, I. L., and J. DAMUTH, 1987 A method for analyzing selection in hierarchically structured populations. *The American Naturalist* **130**: 582-602.
- HUDSON, R. R., 1998 Island models and the coalescent process. *Molecular Ecology* **7**: 413-418.



- HURST, L. D., A. ATLAN and B. O. BENGTTSSON, 1996 Genetic conflicts. *Quarterly Review of Biology* **71**: 317-364.
- INGVARSSON, P., 1998 Kin-structured colonization in *Phalacrus substriatus*. *Heredity* **80**: 456-463.
- INGVARSSON, P., and D. TAYLOR, 2002 Genealogical evidence for epidemics of selfish genes. *PNAS* **99**: 11265-11269.
- JUILLET, N., H. FREYMOND, L. DEGEN and J. GOUDET, 2003 Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **3**: 358-359.
- KARHUNEN, M., and O. OVASKAINEN, 2012 Estimating population-level coancestry coefficients by an admixture F model. *Genetics* **192**: 609-617.
- KELLER, S. R., K. J. GILBERT, P. D. FIELDS and D. R. TAYLOR, 2012 Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-4734.
- KELLY, J. K., 1996 Kin selection in the annual plant *Impatiens capensis*. *American Naturalist* **147**: 899-918.
- KERR, B., C. NEUHAUSER, B. J. M. BOHANNAN and A. M. DEAN, 2006 Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'. *Nature* **442**: 75-78.
- LABOUCHE, A. M., and G. BERNASCONI, 2010 Male moths provide pollination benefits in the *Silene latifolia*-*Hadena bicruris* nursery pollination system. *Functional Ecology* **24**: 534-544.

- MANTEL, N., 1967 The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research* **27**: 209-220.
- MARUYAMA, T., and M. KIMURA, 1980 GENETIC-VARIABILITY AND EFFECTIVE POPULATION-SIZE WHEN LOCAL EXTINCTION AND RECOLONIZATION OF SUB-POPULATIONS ARE FREQUENT. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **77**: 6710-6714.
- MATSCHINER, M., and W. SALZBURGER, 2009 TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-1983.
- MCCAULEY, D. E., 1991 GENETIC CONSEQUENCES OF LOCAL-POPULATION EXTINCTION AND RECOLONIZATION. *Trends Ecol Evol* **6**: 5-8.
- MCCAULEY, D. E., 1994a Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8127-8131.
- MCCAULEY, D. E., 1994b Intrademic group selection imposed by a parasitoid-host interaction. *American Naturalist*: 1-13.
- MCCAULEY, D. E., 1997 The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* **88**: 257-263.
- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.
- MCCAULEY, D. E., J. STEVENS, P. PERONI and J. RAVEILL, 1996 The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* **83**: 727-731.

- MCCAULEY, D. E., and D. R. TAYLOR, 1997 Local population structure and sex ratio: evolution in gynodioecious plants. *American Naturalist* **150**: 406-419.
- MEIRMANS, P. G., and P. H. VAN TIENDEREN, 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792-794.
- MICHALAKIS, Y., and L. EXCOFFIER, 1996 A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061-1064.
- MICHOD, R. E., 1997 Evolution of the individual. *American Naturalist* **150**: S5-S21.
- MICHOD, R. E., 1999 *Darwinian Dynamics, Evolutionary Transitions in Fitness and Individuality*. Princeton University Press, Princeton, N.J.
- MOCCIA, M. D., C. OGER-DESFEUX, G. A. MARAIS and A. WIDMER, 2009 A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.
- NICHOLSON, G., A. SMITH, F. JÓNSSON, Ó. GÚSTAFSSON, K. STEFÁNSSON *et al.*, 2002 Assessing Population Differentiation and Isolation from Single-Nucleotide Polymorphism Data, pp. 695-715 in *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*.
- O'KEEFE, K., and J. ANTONOVICS, 2002 Playing by different rules: the evolution of virulence in sterilizing pathogens. *American Naturalist* **159**: 597-605.
- OKASHA, S., 2004 Multilevel selection and the partitioning of covariance: a comparison of three approaches. *Evolution* **58**: 486-494.

- OLSON, M., D. McCAULEY and D. TAYLOR, 2005 Genetics and adaptation in structured populations: sex ratio evolution in *Silene vulgaris*. *Genetica* **123**: 49-62.
- PANNELL, J. R., 2003 Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* **57**: 949-961.
- PANNELL, J. R., and B. CHARLESWORTH, 2000 Effects of Metapopulation Processes on Measures of Genetic Diversity. *Philosophical Transactions: Biological Sciences* **355**: 1851-1864.
- PERONI, P. A., and R. T. ARMSTRONG, 2001 Density, dispersion and population genetics of a *Silene latifolia* seed bank from southwestern Virginia. *Journal of the Torrey Botanical Society* **128**: 400-406.
- PIRY, S., A. ALAPETITE, J. M. CORNUET, D. PAETKAU, L. BAUDOUIN *et al.*, 2004 GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**: 536-539.
- PULLIAM, H. R., 1988 Sources, sinks, and population regulation. *American Naturalist* **132**: 652-661.
- R DEVELOPMENT CORE TEAM, 2012 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.
- RAMBAUT, A., and A. J. DRUMMOND, 2009 Tracer v1.5 [Internet]. **Available from:** URL <http://beast.bio.ed.ac.uk/Tracer>.
- RANNALA, B., and J. L. MOUNTAIN, 1997 Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 9197-9201.

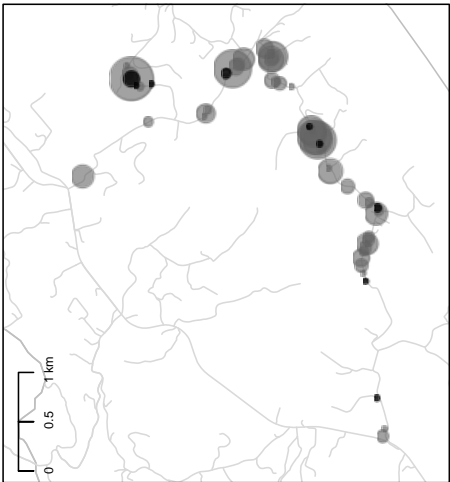
- RICHARDS, C., 2000 Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* **155**: 383-394.
- RICHARDS, C., S. CHURCH and D. MCCAULEY, 1999 The Influence of Population Size and Isolation on Gene Flow by Pollen in *Silene alba*. *Evolution* **53**: 63-73.
- RICHARDS, C. M., S. N. EMERY and D. E. MCCAULEY, 2003 Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity* **90**: 181-186.
- ROUSSET, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219-1228.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**: 253-262.
- SMOUSE, P. E., R. S. WAPLES and J. A. TWOREK, 1990 A Genetic Mixture Analysis for use with Incomplete Source Population Data. *Canadian Journal of Fisheries and Aquatic Sciences* **47**: 620-634.
- STEVENS, L., C. J. GOODNIGHT and S. KALISZ, 1995 Multilevel selection in natural-populations of *Impatiens capensis*. *American Naturalist* **145**: 513-526.
- SZATHMARY, E., and J. M. SMITH, 1995 The major evolutionary transitions. *Nature* **374**: 227-232.
- TAYLOR, D. R., C. ZEYL and E. COOKE, 2002 Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 3690-3694.
- TEIXEIRA, S., and G. BERNASCONI, 2007 High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology* **16**: 4370-4379.

- THRALL, P., C. RICHARDS, D. E. MCCAULEY and J. ANTONOVICS, 1998 Metapopulation collapse: the consequences of limited gene flow in spatially structured populations, pp. 83-104 in *Modeling spatiotemporal dynamics in ecology*, edited by J. BASCOMPTE and R. V. SOLE. Springer-Verlag Berlin.
- WADE, M., and C. GOODNIGHT, 1991 Wright's shifting balance theory: an experimental study, pp. 1015-1018 in *Science*.
- WADE, M. J., 1977 Experimental-study of group selection. *Evolution* **31**: 134-153.
- WADE, M. J., 1985 Soft selection, hard selection, kin selection, and group selection. *American Naturalist* **125**: 61-73.
- WADE, M. J., 2013 PHASE III OF WRIGHT'S SHIFTING BALANCE PROCESS AND THE VARIANCE AMONG DEMES IN MIGRATION RATE. *Evolution* **67**: 1591-1597.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating f-statistics for the analysis of population-structure. *Evolution* **38**: 1358-1370.
- WHITLOCK, M. C., 1992 NONEQUILIBRIUM POPULATION-STRUCTURE IN FORKED FUNGUS BEETLES - EXTINCTION, COLONIZATION, AND THE GENETIC VARIANCE AMONG POPULATIONS. *American Naturalist* **139**: 952-970.
- WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191-1202.
- WHITLOCK, M. C., and D. E. MCCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717-1724.

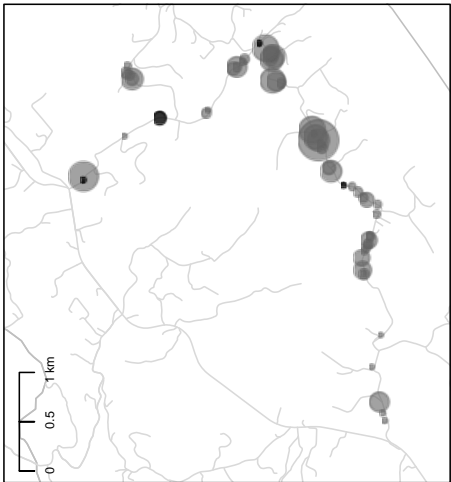
- WHITLOCK, M. C., and D. E. McCAULEY, 1999 Indirect measures of gene flow and migration:  $F_{ST}$  not equal  $1/(4Nm+1)$ . *Heredity* **82**: 117-125.
- WILLI, Y., and M. FISCHER, 2005 Genetic rescue in interconnected populations of small and large size of the self-incompatible *Ranunculus reptans*. *Heredity* **95**: 437-443.
- WILLI, Y., J. VAN BUSKIRK and M. FISCHER, 2005 A threefold genetic allee effect: population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. *Genetics* **169**: 2255-2265.
- WILSON, D., 1980 *The natural selection of populations and communities*. Benjamin-Cumming, Menlo Park, CA.
- WILSON, D., 1997 Introduction: Multilevel selection theory comes of age, pp. 1-21 in *Am Nat*.
- WILSON, G. A., and B. RANNALA, 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177-1191.
- WRIGHT, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97-159.

**Figure 1.** Spatial distribution of source and colonist *S. latifolia* populations in the three focal metapopulation section, from 2008-2011. Populations colored grey were considered as sources, while those colored black are colonists. Circle size is representative of the population size at the time of tissue collection.

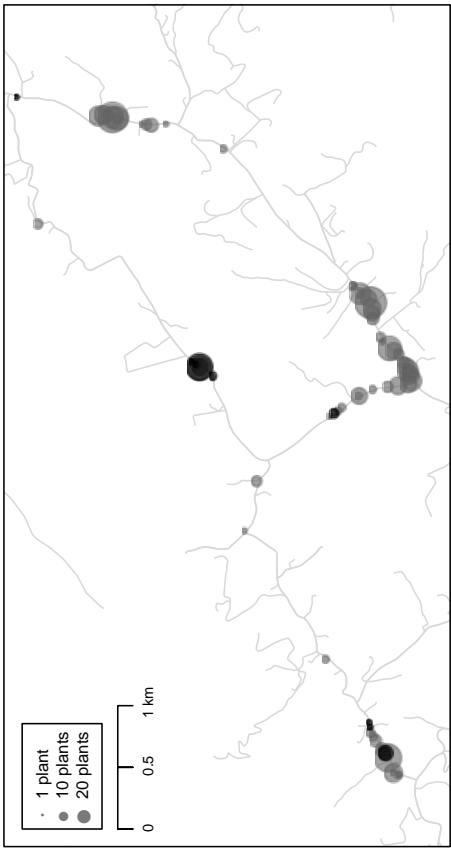




Section 9

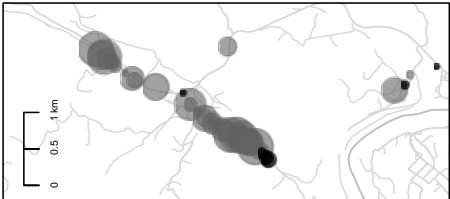


Section 6

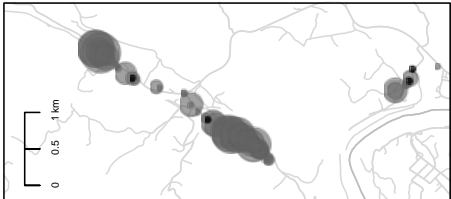


2008 – 2009

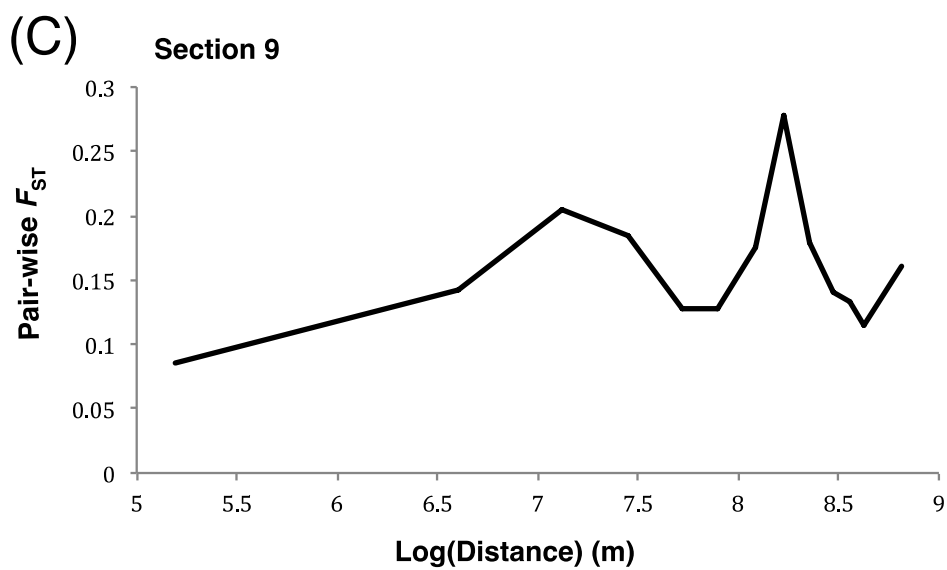
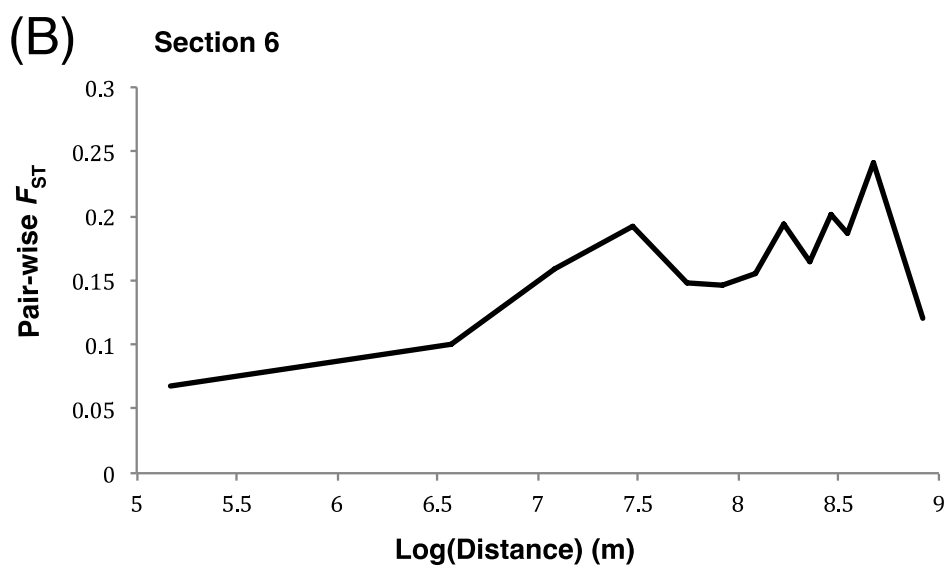
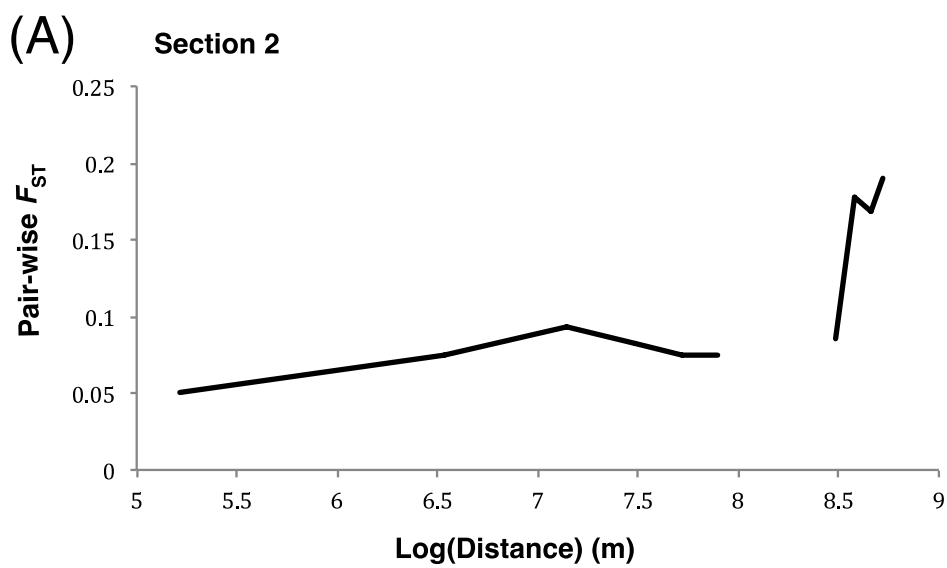
2010 – 2011



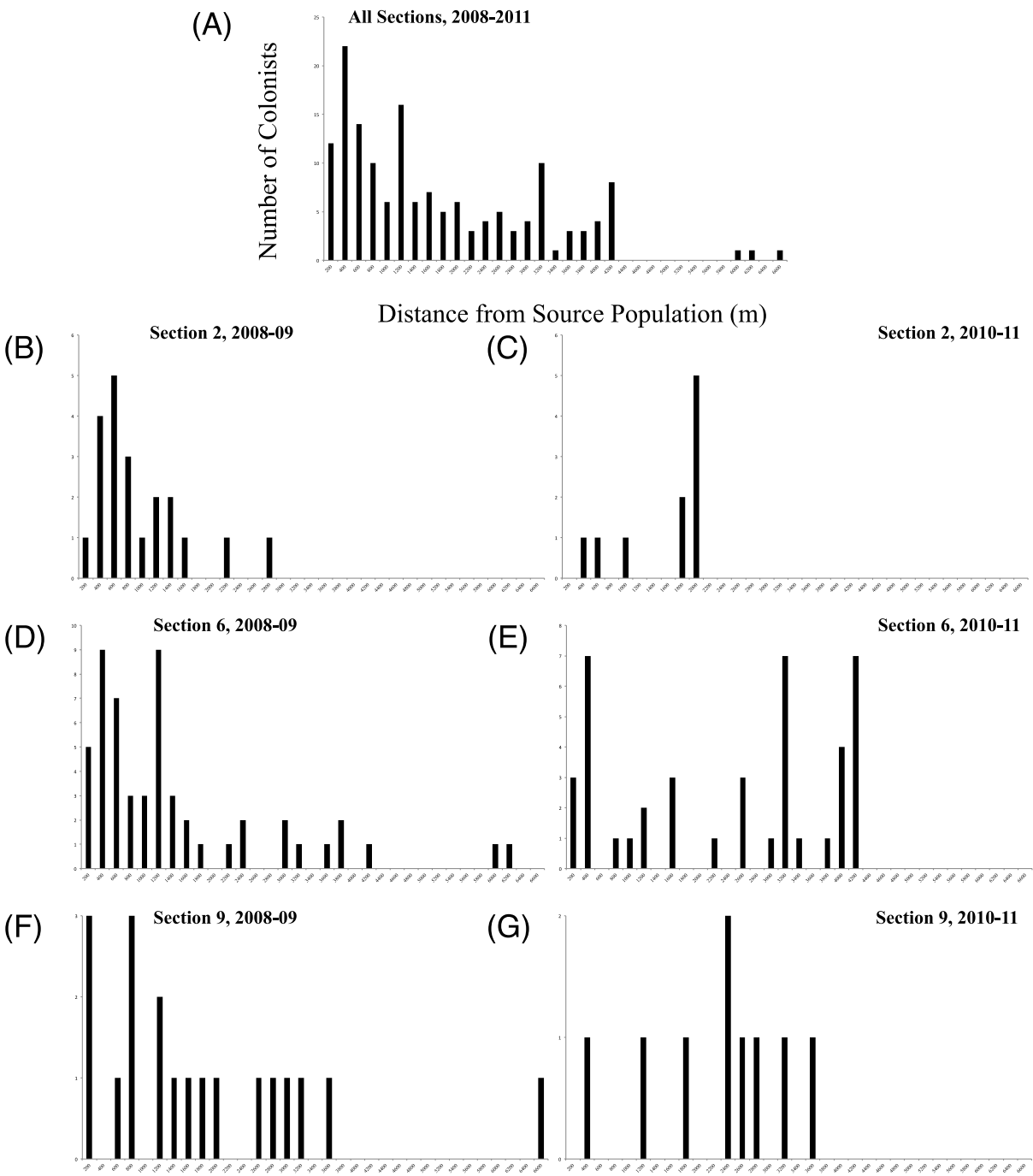
Section 2



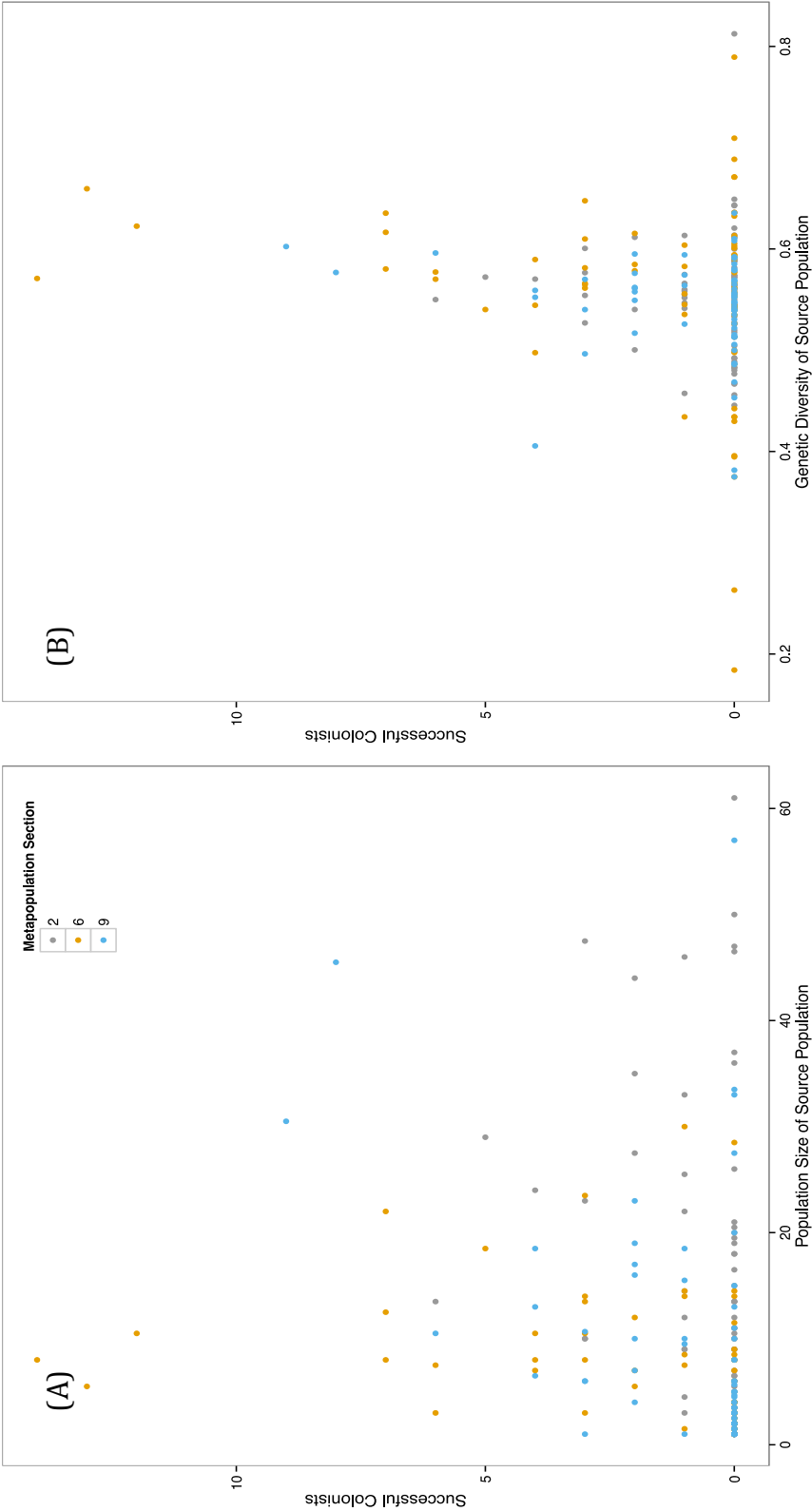
**Figure 2.**  $F_{ST}/(1-F_{ST})$  for established populations, plotted against distance class. Each point in the dataset depicts the average pairwise distance between populations as a function of the distance class to which they belong. Subpopulations without data points represent pairwise distances not found in that section. (A), (B), and (C) represent results for sections 2, 6, and 9, respectively.



**Figure 3.** Distance between most probable source and colonist as estimated with GeneClass2. (A) is the combined distribution of distances amongst all sources and colonists, and (B-G) are the individuals samples (2008 and 2010) for sections 2, 6, and 9, respectively.



**Figure 4.** Population source properties generating successful colonists. (A) is the distribution of successful colonists as a result of population size, and (B) the result of population genetic diversity.



**Table 1.** Global population genetic summaries. Variables are the number of alleles ( $N$ ), the effective number of alleles ( $E_N$ ), observed heterozygosity ( $H_O$ ), Expected heterozygosity ( $H_E$ ), and the inbreeding coefficient ( $F_{IS}$ ), global among-population allelic variation ( $F_{IS}$ ), and Jost’s measure of population differentiation ( $D_{EST}$ ). (A) and (B) represent summaries from the 2008 and 2010 samples, respectively.

(A)

<i>Locus</i>	Section 2							Section 6							Section 9						
	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>	<i>N</i>	<i>E_N</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>ST</sub></i>	<i>D<sub>EST</sub></i>
SL_eSSR01	-	-	-	-	-	-	-	7	1.818	0.224	0.488	0.542	<b>0.114</b>	0.129	6	1.722	0.161	0.452	0.644	<b>0.099</b>	0.094
SL_eSSR04	10	2.282	0.577	0.584	0.012	<b>0.05</b>	0.076	7	2.399	0.524	0.62	0.154	<b>0.05</b>	0.089	6	2.475	0.519	0.629	0.174	<b>0.105</b>	0.206
SL_eSSR06	18	3.762	0.699	0.765	0.086	<b>0.041</b>	0.145	14	3.333	0.503	0.751	0.33	<b>0.069</b>	0.233	13	3.108	0.39	0.725	0.462	<b>0.1</b>	0.304
SL_eSSR09	12	1.407	0.287	0.301	0.049	<b>0.037</b>	0.017	7	1.65	0.393	0.416	0.055	<b>0.073</b>	0.059	9	1.584	0.382	0.386	0.01	<b>0.091</b>	0.065
SL_eSSR12	20	3.296	0.724	0.724	-0.001	<b>0.044</b>	0.123	14	3.557	0.674	0.762	0.116	<b>0.117</b>	0.444	15	4.093	0.702	0.795	0.117	<b>0.087</b>	0.381
SL_eSSR16	8	1.569	0.334	0.378	0.117	<b>0.099</b>	0.068	4	2.245	0.583	0.584	0.002	<b>0.073</b>	0.116	8	2.248	0.537	0.583	0.079	<b>0.141</b>	0.238
SL_eSSR17	-	-	-	-	-	-	-	13	2.747	0.608	0.674	0.098	<b>0.106</b>	0.256	14	2.66	0.556	0.658	0.154	<b>0.069</b>	0.147
SL_eSSR20	11	1.161	0.121	0.145	0.168	<b>0.065</b>	0.012	3	1.057	0.056	0.057	0.016	0.028*	0.002	3	1.048	0.038	0.049	0.218	0.039*	0.002
SL_eSSR22	-	-	-	-	-	-	-	10	2.817	0.561	0.686	0.183	<b>0.149</b>	0.403	11	3.167	0.501	0.726	0.31	<b>0.084</b>	0.253
SL_eSSR27	-	-	-	-	-	-	-	9	2.462	0.58	0.628	0.077	<b>0.167</b>	0.355	11	2.568	0.595	0.641	0.071	<b>0.128</b>	0.271
SL_eSSR28	-	-	-	-	-	-	-	6	1.79	0.469	0.465	-0.008	<b>0.22</b>	0.257	6	1.353	0.24	0.275	0.126	<b>0.245</b>	0.127
SL_eSSR29	25	5.488	0.835	0.85	0.018	<b>0.054</b>	0.332	19	3.871	0.715	0.785	0.09	<b>0.103</b>	0.441	18	3.683	0.7	0.765	0.085	<b>0.083</b>	0.305
SL_eSSR30	21	3.407	0.65	0.737	0.119	<b>0.056</b>	0.17	21	2.683	0.345	0.679	0.492	<b>0.111</b>	0.278	18	2.705	0.281	0.678	0.585	<b>0.083</b>	0.197
slat_18	10	2.837	0.418	0.683	0.388	<b>0.058</b>	0.137	9	2.045	0.234	0.556	0.578	<b>0.208</b>	0.346	11	2.522	0.276	0.649	0.575	<b>0.119</b>	0.259
slat_32	9	2.709	0.29	0.67	0.568	<b>0.137</b>	0.331	7	2.271	0.364	0.602	0.396	<b>0.159</b>	0.301	9	2.384	0.34	0.62	0.451	<b>0.181</b>	0.375
slat_33	3	1.128	0.061	0.12	0.49	<b>0.237</b>	0.044	2	1.515	0.184	0.368	0.502	<b>0.265</b>	0.221	3	1.211	0.105	0.186	0.434	<b>0.597</b>	0.35
slat_48	6	1.253	0.087	0.214	0.593	<b>0.099</b>	0.031	5	1.766	0.194	0.472	0.589	<b>0.108</b>	0.113	5	1.666	0.233	0.427	0.455	<b>0.139</b>	0.125
slat_72	12	2.858	0.349	0.688	0.493	<b>0.048</b>	0.116	12	2.545	0.326	0.657	0.504	<b>0.097</b>	0.217	16	2.86	0.282	0.7	0.596	<b>0.171</b>	0.499
slat_85	15	2.085	0.311	0.55	0.434	<b>0.101</b>	0.142	16	1.829	0.284	0.489	0.418	<b>0.156</b>	0.185	16	1.618	0.232	0.408	0.432	<b>0.194</b>	0.172
SI_8	11	2.153	0.324	0.566	0.428	<b>0.132</b>	0.203	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A11	43	4.382	0.707	0.806	0.122	<b>0.062</b>	0.283	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Overall	14.625	2.611	0.423	0.549	0.229	<b>0.074</b>	0.099	9.737	2.337	0.412	0.565	0.272	<b>0.128</b>	0.2	10.421	2.351	0.372	0.545	0.317	<b>0.138</b>	0.198



(B)

<i>Locus</i>	Section 2							Section 6							Section 9						
	<i>N</i>	<i>E_N</i>	<i>H</i> <sub>O</sub>	<i>H</i> <sub>E</sub>	<i>F</i> <sub>IS</sub>	<i>F</i> <sub>ST</sub>	<i>D</i> <sub>EST</sub>	<i>N</i>	<i>E_N</i>	<i>H</i> <sub>O</sub>	<i>H</i> <sub>E</sub>	<i>F</i> <sub>IS</sub>	<i>F</i> <sub>ST</sub>	<i>D</i> <sub>EST</sub>	<i>N</i>	<i>E_N</i>	<i>H</i> <sub>O</sub>	<i>H</i> <sub>E</sub>	<i>F</i> <sub>IS</sub>	<i>F</i> <sub>ST</sub>	<i>D</i> <sub>EST</sub>
SL_eSSR01	6	1.974	0.311	0.521	0.405	<b>0.054</b>	0.064	6	2.004	0.315	0.54	0.416	<b>0.051</b>	0.065	6	1.758	0.213	0.466	0.544	0.064*	0.063
SL_eSSR04	7	2.443	0.608	0.615	0.011	<b>0.036</b>	0.062	8	2.745	0.635	0.672	0.055	<b>0.033</b>	0.072	8	2.839	0.623	0.684	0.09	<b>0.085</b>	0.209
SL_eSSR06	11	3.105	0.776	0.702	-0.105	<b>0.102</b>	0.273	11	1.774	0.552	0.454	-0.215	<b>0.06</b>	0.055	11	3.571	0.624	0.764	0.184	<b>0.09</b>	0.334
SL_eSSR09	7	1.402	0.286	0.298	0.043	<b>0.087</b>	0.042	7	1.859	0.428	0.49	0.127	<b>0.156</b>	0.184	10	1.731	0.417	0.445	0.063	<b>0.078</b>	0.071
SL_eSSR12	18	3.307	0.673	0.728	0.075	<b>0.053</b>	0.152	14	3.972	0.729	0.792	0.08	<b>0.079</b>	0.335	16	3.868	0.678	0.785	0.137	<b>0.091</b>	0.381
SL_eSSR16	7	1.617	0.402	0.397	-0.013	<b>0.063</b>	0.045	9	2.253	0.574	0.587	0.021	<b>0.109</b>	0.179	8	2.203	0.547	0.576	0.051	<b>0.13</b>	0.211
SL_eSSR17	18	4.001	0.577	0.788	0.268	<b>0.048</b>	0.194	15	2.295	0.425	0.604	0.297	<b>0.104</b>	0.184	14	2.544	0.524	0.645	0.187	<b>0.104</b>	0.218
SL_eSSR20	5	1.15	0.137	0.135	-0.015	<b>0.022</b>	0.004	5	1.21	0.14	0.185	0.245	<b>0.163</b>	0.046	4	1.08	0.051	0.079	0.358	<b>0.03</b>	0.003
SL_eSSR22	11	3.267	0.563	0.728	0.227	<b>0.044</b>	0.127	10	3.459	0.662	0.754	0.122	<b>0.096</b>	0.338	10	3.243	0.529	0.738	0.284	<b>0.075</b>	0.238
SL_eSSR27	8	1.973	0.484	0.514	0.058	<b>0.063</b>	0.072	10	2.704	0.623	0.666	0.065	<b>0.115</b>	0.269	9	2.714	0.635	0.666	0.046	<b>0.111</b>	0.258
SL_eSSR28	11	2.801	0.675	0.668	-0.01	<b>0.091</b>	0.207	9	1.712	0.456	0.437	-0.043	<b>0.191</b>	0.19	6	1.353	0.31	0.272	-0.138	<b>0.234</b>	0.119
SL_eSSR29	20	5.705	0.814	0.859	0.053	<b>0.052</b>	0.343	18	4.064	0.787	0.795	0.01	<b>0.094</b>	0.416	18	3.912	0.765	0.784	0.024	<b>0.06</b>	0.242
SL_eSSR30	22	3.724	0.715	0.763	0.063	<b>0.044</b>	0.152	22	3.633	0.491	0.779	0.37	<b>0.042</b>	0.16	16	2.75	0.57	0.675	0.156	<b>0.101</b>	0.243
slat_18	11	2.463	0.474	0.624	0.24	<b>0.045</b>	0.081	12	2.576	0.538	0.651	0.174	<b>0.079</b>	0.165	12	2.702	0.462	0.673	0.313	<b>0.067</b>	0.153
slat_32	7	3.317	0.396	0.74	0.466	<b>0.061</b>	0.188	7	2.462	0.421	0.637	0.34	<b>0.069</b>	0.135	8	2.282	0.352	0.604	0.418	<b>0.129</b>	0.235
slat_33	2	1.106	0.077	0.1	0.23	<b>0.301</b>	0.049	2	1.537	0.249	0.375	0.336	<b>0.248</b>	0.204	2	1.253	0.17	0.215	0.21	<b>0.557</b>	0.358
slat_48	5	1.222	0.087	0.193	0.548	<b>0.077</b>	0.02	4	1.718	0.223	0.453	0.508	<b>0.136</b>	0.134	4	1.758	0.327	0.46	0.288	<b>0.051</b>	0.048
slat_72	13	2.429	0.354	0.623	0.431	<b>0.056</b>	0.1	13	2.613	0.391	0.665	0.412	<b>0.103</b>	0.235	11	2.256	0.315	0.6	0.475	<b>0.153</b>	0.283
slat_85	15	2.185	0.434	0.569	0.237	<b>0.162</b>	0.261	18	2.103	0.398	0.562	0.292	<b>0.095</b>	0.139	16	2.864	0.811	0.677	-0.197	<b>0.051</b>	0.118
SI_8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Overall	10.737	2.589	0.465	0.556	0.163	<b>0.068</b>	0.094	10.526	2.458	0.476	0.584	0.186	<b>0.1</b>	0.162	9.947	2.457	0.47	0.569	0.175	<b>0.113</b>	0.175

**Table 2.** Distance between genetic subpopulations, calculated from the regression coefficient of  $F_{ST}$ .  $b$  is the slope of the regression of  $F_{ST}$  and distance,  $N$  an approximate estimate of population size, and  $\sigma$  is an estimate of dispersal distance.

Section, year	Distance parameters					Error
	b (regression)	$N\sigma^2$	$N$	$\sigma^2$	Distance (m)	p
2, 2010	3.89E-05	6432.41	22.48	286.08	16.91	<0.001
6, 2010	1.27E-05	19728.22	10.69	1845.40	42.96	<0.001
9, 2010	7.82E-06	31969.43	12.38	2582.68	50.82	0.116

**Table 3.** Metapopulation dynamic summary. Here,  $m$  is the average of the pairwise migration rates between all extant populations,  $c$  is the number of newly colonized sites,  $e$  is the number of extinctions,  $k$  mean number of individuals per new colonization, and  $\phi$  is the mean number of sources per new colonized site.

	$m$	$c$	$e$	$k$	$\phi$
Section2, 2008-9	0.008	6	24	3.50	2.67
Section2, 2010-11	0.007	8	14	2.75	1.63
Section6, 2008-9	0.015	15	24	4.20	2.60
Section6, 2010-11	0.011	22	13	2.55	1.86
Section9, 2008-9	0.010	18	21	2.74	2.00
Section9, 2010-11	0.012	9	25	1.67	1.67

**Table 4.** Posterior model probabilities for fine scale samples across three sections of the metapopulation in 2008 and 2010. Posterior model probabilities derive from consideration of three spatiotemporal variables, pairwise population isolation (G1), population size (G2), and population age (G3). The null model includes only an intercept. Within the BIMr framework only pairwise interactions are estimable.

Model	Factors Included	Section 2, 2008	Section 2, 2010	Section 6, 2008	Section 6, 2010	Section 9, 2008	Section 9, 2010
		Pr	Pr	Pr	Pr	Pr	Pr
1	null	0.00	0.00	0.30	0.00	0.00	0.00
2	G1	0.00	0.00	<b>0.40</b>	0.00	0.00	0.00
3	G2	0.00	0.00	0.03	0.00	0.00	0.00
4	G1 + G2	0.00	0.00	0.05	0.00	0.00	0.00
5	G3	0.00	0.00	0.09	0.00	0.00	0.00
6	G1 + G3	0.00	0.00	0.05	0.00	0.00	0.00
7	G2 + G3	0.00	0.00	0.01	0.00	0.00	0.00
8	G1 + G2 + G3	0.00	0.00	0.01	0.00	0.00	0.00
9	G1 + G2 + G1*G2	<b>0.83</b>	0.00	0.01	0.00	0.00	0.00
10	G1 + G2 + G3 + G1*G2	0.15	0.00	0.00	0.00	0.00	0.00
11	G1 + G3 + G1*G3	0.00	0.00	0.03	0.00	0.00	0.00
12	G1 + G2 + G3 + G1*G3	0.00	0.00	0.02	<b>0.63</b>	0.00	0.00
13	G1 + G2 + G3 + G1*G2 + G1*G3	0.01	0.00	0.00	0.05	0.00	0.00
14	G2 + G3 + G2*G3	0.00	0.18	0.00	0.00	0.00	0.00
15	G1 + G2 + G3 + G2*G3	0.00	0.28	0.01	0.00	0.00	0.00
16	G1 + G2 + G3 + G1*G2 + G2*G3	0.01	<b>0.51</b>	0.00	0.00	<b>0.97</b>	<b>0.95</b>
17	G1 + G2 + G3 + G1*G3 + G2*G3	0.00	0.01	0.00	0.30	0.00	0.00
18	G1 + G2 + G3 + G1*G2 + G1*G3 + G2*G3	0.00	0.02	0.00	0.02	0.03	0.05

**Table 5.** Posterior estimates for regression parameters of the model with the highest posterior probability. Numbers in bold are the mean parameter estimate, italics alone the mode, and the bracket values the 95% confidence interval. Here, the  $\alpha$  followed by a subscript correspond to the spatiotemporal variables described in Table 4, and  $\sigma^2$  is an indication of the quantity of variation unexplained by the regression.

Regression Coefficient	Factor	Section 2, 2008	Section 2, 2010	Section 6, 2008	Section 6, 2010	Section 9, 2008	Section 9, 2010
$\sigma^2$	-	<b>1.43</b> 1.57 [-0.174;2.53]	<b>43.5</b> 47 [38.4;62.2]	<b>1.54</b> 1.55 [0.242;3.53]	<b>0.316</b> 0.255 [0.145;0.667]	<b>1.77</b> 1.65 [1.3;2.53]	<b>0.632</b> 0.659 [0.292;0.957]
$\alpha_0$	Intercept	<b>0.314</b> 0.34 [-0.185;0.704]	<b>3.17</b> 3.04 [2;3.93]	<b>2.8</b> 3.3 [-0.376;3.76]	<b>2.99</b> 3.01 [1.53;3.78]	<b>6.05</b> 5.87 [4.69;6.62]	<b>4.02</b> 4.02 [2.76;4.49]
$\alpha_1$	G1	<b>0.0795</b> 0.101 [-0.481;0.659]	<b>-0.221</b> -0.222 [-0.59;-0.0127]	<b>-0.561</b> -0.558 [-1.13;0.274]	<b>-0.44</b> -0.436 [-0.77;-0.0615]	<b>-0.121</b> -0.12 [-0.376;0.163]	<b>-0.62</b> -0.558 [-1.02;-0.292]
$\alpha_2$	G2	- - [-0.481;0.659]	<b>-0.016</b> -0.13 [-0.349;0.53]	- - [-1.13;0.274]	<b>-0.392</b> -0.41 [-0.655;-0.0989]	<b>1.28</b> 1.26 [0.951;1.57]	<b>0.775</b> 0.52 [0.226;1.37]
$\alpha_3$	G3	- <b>-0.622</b> -0.713 [-0.968;-0.244]	<b>-0.723</b> -0.774 [-1.04;-0.24]	- - [-1.13;0.274]	<b>0.552</b> 0.622 [0.0793;1.05]	<b>-0.285</b> -0.242 [-0.744;-0.0263]	<b>0.0547</b> 0.125 [-0.326;0.419]
$\alpha_4$	G1*G2	- - [-0.5;0.134]	<b>-0.282</b> -0.319 [-0.5;0.134]	- - [-1.13;0.274]	- - [-1.13;0.274]	<b>0.372</b> 0.367 [0.108;0.61]	<b>0.481</b> 0.464 [0.2;0.711]
$\alpha_5$	G1*G3	- - [-0.5;0.134]	- - [-0.5;0.134]	- - [-1.13;0.274]	<b>0.517</b> 0.604 [0.225;0.755]	- - [-0.887;-0.467]	- - [-0.835;-0.128]
$\alpha_6$	G2*G3	- - [0.268;0.77]	<b>0.543</b> 0.55 [0.268;0.77]	- - [-1.13;0.274]	- - [-1.13;0.274]	<b>-0.665</b> -0.662 [-0.887;-0.467]	<b>-0.472</b> -0.464 [-0.835;-0.128]

**Table 6.** Mean values for each parameter among assigned parental populations and among all populations, and level of significance of the difference between these two values as obtained with a Wilcoxon test. The mean values among assigned parental populations are obtained by weighting each potential parental population by its probability of paternal origin on each seed, as obtained with GeneClass2.

Section, year		Mean pairwise distance
2, 2008	Assigned parents	2003.70
	All populations	1720.37
	<i>p</i> value	<b>0.02957</b>
2, 2010	Assigned parents	2251.94
	All populations	2111.27
	<i>p</i> value	0.4027
6, 2008	Assigned parents	2357.81
	All populations	2624.38
	<i>p</i> value	0.9132
6, 2010	Assigned parents	3221.86
	All populations	2933.72
	<i>p</i> value	0.4785
9, 2008	Assigned parents	2137.00
	All populations	2164.26
	<i>p</i> value	0.5898
9, 2010	Assigned parents	2218.97
	All populations	2000.09
	<i>p</i> value	<b>0.037</b>











**Suppl. Figure 5.** Pairwise migration estimates for Section 9, 2008. The diagonal represents resident proportion while the off diagonals represent short-term migration rates between populations.

**Suppl. Figure 6.** Pairwise migration estimates for Section 9, 2010. The diagonal represents resident proportion while the off diagonals represent short-term migration rates between populations.

Appendix 1: Isolation and characterization of microsatellite loci in the white campion, *Silene latifolia* (Caryophyllaceae)<sup>8</sup>

---

<sup>8</sup> Formatted for submission as a coauthored manuscript: Fields, P.D., Keller, S.K., Ingvarsson, P.K., Pedersen, A.B., and D. R. Taylor

**ABSTRACT**

We report the isolation and characterization of 8 polymorphic microsatellite loci in *Silene latifolia* (Caryophyllaceae). We surveyed allelic variability in a sample of individuals collected across *S. latifolia*'s European range, and a single population located within a metapopulation in Virginia, US. These loci were polymorphic at both spatial scales, yielding 2-35 alleles per locus. Expected heterozygosity ranged from 0.0678-0.899. These microsatellite loci should be valuable tools for population genetic analysis at multiple spatial scales.

Keywords: white campion, *Silene latifolia*, microsatellite

*Silene latifolia* (= *S. alba*, Caryophyllaceae) is a short-lived, dioecious perennial plant, with X-Y sex determination. The species has a broad geographic distribution in Europe where it is native, and has become widespread in North America where it was introduced in the mid-1800's (MCNEILL 1977). At local spatial scales, *S. latifolia* occurs in networks of patches subject to recurrent extinction and recolonization (MCCAULEY *et al.* 1995). *Silene latifolia* is an emerging model system for research on metapopulations, population structure, evolution of breeding systems, sex chromosome evolution, sexual dimorphism, and the evolution of invasiveness (BERNASCONI *et al.* 2009). In this note, we describe the development and characterization of 8 variable microsatellite loci that expand the resources currently available for studies of genetic diversity, paternity analysis, and linkage mapping.

We extracted 100µg of total genomic DNA from one male *S. latifolia* plant using a DNAEasy Plant Mini kit (Qiagen). This DNA was sent to Genetic Identification Services, Inc. (GIS, Chatsworth, CA, USA, <http://www.genetic-id-services.com/>) and used to construct 4 microsatellite libraries (JONES *et al.* 2002). Genomic DNA was partially restricted with blunt-end cutting enzymes (Rsa I, Hae III, Bsr B1, Pvu II, Stu I, Sca I, Eco RV). Fragments in the size range of 300 to 750 bp were adapted and subjected to magnetic bead capture (CPG, Inc., Lincoln Park, New Jersey), using biotinylated capture molecules. These libraries were enriched for CA, GA, ATG and TGAA sequence repeats. Captured molecules were amplified and restricted with HindIII to remove the adapters. The resulting fragments were ligated into the HindIII site of pUC19 and stored in glycerol stock tubes at -70° C. Colonies were grown from 60µL of enriched glycerol stock on blue-gal/IPTG/ampicillin LB agar plates and incubated at 37° C overnight. White colonies from the spread stock plates were transferred

using sterilized toothpicks onto gridded blueo-gal/IPTG/ampicillin LB agar plates and grown overnight at 37° C. To release the plasmids, selected colonies were heated to 100° C for 10 minutes with 10µl of PCR reaction mix (M13 universal primers) and Rnase A. Following lysing, Taq polymerase (0.075 µL) was added and the plasmids were amplified using a PE Gene Amp PCR system 9700 (30 s denaturing at 94° C, 30 s annealing at 57° C, 30 s extension at 72° C, for 21 cycles). Size of the inserts was estimated using a Phi-X/Hae III standard on a 1% agarose gel. Amplified plasmids were purified using a QIAquick PCR purification kit (Qiagen). Cycle sequencing reactions were prepared using Big Dye sequencing kits (Applied Biosystems), and purified through sephadex columns. Plasmid inserts were then sequenced using an automated ABI Prism 377 DNA sequencer.

We designed primers from sequenced clones using the software PRIMER 3 (ROZEN and SKALETSKY 1998). Amplification reactions were prepared in a 10µL volume containing approximately 5-10 ng of template DNA, 1 X PCR buffer (Stratagene, containing 2.0mM MgCl<sub>2</sub>), 0.25mM of each dNTP (ABI Prism), 0.5µM of each designed primer and 0.1 U of Paq5000 polymerase (Stratagene). Amplifications were performed in a PTC-200 thermal cycler (MJ Research) with the following cycling conditions: 2 min at 95 °C,  $n$  cycles composed of 20 s denaturing at 95 °C, 20 s annealing at  $T_a$ , (see Table 1 for  $T_a$  and  $n$ ), 15 s extension at 72 °C, followed by a final extension at 72 °C for 1 hour. PCR products were amplified with the forward primer pair end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE, or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following 5 min of denaturing at 95 °C,



fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems).

We assessed range-wide polymorphism by surveying a sample of 30 plants collected across *S. latifolia*'s native European range (1 plant per site). To assess diversity at a finer scale, we surveyed a sample of 31 plants from within a single population located in a well-characterized *S. latifolia* metapopulation in Giles County, VA, US. The number of alleles per locus ( $N_a$ ), observed ( $H_o$ ) and expected ( $H_E$ ) heterozygosities, and exact tests of Hardy-Weinberg equilibrium were calculated with ARLEQUIN v3.1 (SCHNEIDER *et al.* 2000). Pairwise tests of genotypic linkage disequilibrium among loci were conducted using FSTAT v2.9.4, with 1000 permutations for significance testing (GOUDET 1995). Evidence for null alleles and genotyping errors were evaluated using MICROCHECKER v. 2.2.3 (VAN OOSTERHOUT *et al.* 2006).

All loci were polymorphic at both scales. The number of alleles per locus ranged from 2-35 (mean 14) across the European samples and 2-14 (mean 7) within the Giles County population. Expected heterozygosity was high in both samples, averaging 0.771 in Europe and 0.586 in Giles County. Analysis of genotype frequencies within the Giles County population revealed a consistent excess of homozygotes across loci (Table 1), resulting in significant departures from Hardy-Weinberg equilibrium at each locus ( $P \leq 0.01$ ), and all but one of these loci (*slat48*) were still significant after Bonferroni correction. Because this pattern was evident across loci, we interpret this to be the product of fine-scale substructuring within this population, a characteristic observed in other *Silene* species distributed in metapopulations (TERO and SCHLOTTERER 2005). While an excess of homozygotes across loci may also be caused by the presence of null alleles, MICROCHECKER revealed no evidence of large allele dropout or scoring error due to stuttering. Pairwise

genotypic disequilibria between loci ranged from  $R^2 = 0.002 - 0.057$ , and none were significant following Bonferroni correction.

## LITERATURE CITED

- BERNASCONI, G., J. ANTONOVICS, A. BIERE, D. CHARLESWORTH, L. F. DELPH *et al.*, 2009 *Silene* as a model system in ecology and evolution. *Heredity* **103**: 5-14.
- GOUDET, J., 1995 FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**: 485-486.
- JONES, K., K. LEVINE and J. BANKS, 2002 Characterization of 11 polymorphic tetranucleotide microsatellites for forensic applications in California elk (*Cervus elaphus canadensis*). *Molecular Ecology Resources* **2**: 425-427.
- MCCAULEY, D. E., J. RAVEILL and J. ANTONOVICS, 1995 Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630-636.
- MCNEILL, J., 1977 Biology of Canadian weeds .25. *Silene-alba* (Miller) Krause,EHL. *Can J Plant Sci* **57**: 1103-1114.
- ROZEN, S., and H. SKALETISKY, 1998 Primer 3. Code available at [http://www.genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www.genome.wi.mit.edu/genome_software/other/primer3.html), pp.
- SCHNEIDER, S., D. ROESSLI and L. EXCOFFIER, 2000 Arlequin: A software for population genetics data analysis. Ver. 2.0., pp. in *Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva*.
- TERO, N., and C. SCHLOTTERER, 2005 Isolation and characterization of microsatellite loci from *Silene tatarica*. *Molecular Ecology Notes* **5**: 517-518.

VAN OOSTERHOUT, C., D. WEETMAN and W. HUTCHINSON, 2006 Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Mol Ecol Resour* **6**: 255-256.

**Table 1.** Characterization of 8 microsatellite loci in *Silene latifolia*. Allelic size range, number of alleles ( $N_a$ ), observed ( $H_o$ ) and expected ( $H_E$ ) heterozygosities based upon a screen of 31 plants sampled within a single population in Giles County VA, USA and from 30 plants collected across geographically dispersed populations of *S. latifolia*'s European range. A Brookfield 1 estimate ( $\hat{r}_b$ ) was used to quantify the presence of null alleles in our sampling.

Locus	Primer sequence 5'-3'	Repeat Motif	$T_a$	n	Size range	Virginia, USA			Europe		$\hat{r}_b$	GenBank accession
						$N_a$	$H_o$	$H_E$	$N_a$	$H_E$		
<i>slat</i> 16	F: CCGTTTCCTTCCCAATTC R: TGTGGGTCTTACAAGAGGTTCA	(ATG) <sub>18</sub>	62	30	164-239	14	0.63	0.893	14	0.919	0.132	FJ573199
<i>slat</i> 18	F: TCGCCGAGTTAAGTGCTGT  R: TCCCGAACTAACCTGTTCTT	(GAT) <sub>8</sub>	60	30	128-154	7	0.4	0.794	12	0.913	0.214	FJ573200
<i>slat</i> 28	F: CCCTTTCTTTACTTCATTAACCACA  R: TGGCGGAGAAAGAGAAAATG	(CT) <sub>20</sub>	70*	8 + 30	165-182	14	0.667	0.899	35	0.976	0.115	FJ573202
<i>slat</i> 32	F: CCTGAATTGTTCTTTATTTCTTTCC  R: TGCTTCGAGAAGTCTCCTATCAT	(TATG) <sub>6</sub>	70*	8 + 30	120-145	8	0.29	0.793	12	0.865	0.275	FJ573203
<i>slat</i>			70*		194-197	2						FJ573204

33

	F: CACAACACCCCGTTAAAA	(ACC) <sub>3</sub>		8 + 30				0.03 2	0.1 51	2	0.42 7	0.101	
	R: GGATGGTGTCTAGGTTTTTGG												
slat 48	F: AGATGACAAGACTCGCGTAGG	(GGA) <sub>4</sub>	70*	8 + 30	174-177	2	0	0.0 68		2	0.25 9	0.062 4	FJ573206
	R: TCCAGTGCCTCCTCCATATC												
slat 72	F: TGATGGGAAATCGTGCTGTA	(ATG) <sub>9</sub>	70*	8 + 30	136-183	5		0.45 2	0.6 87	16	0.89 6	0.134 1	FJ573207
	R: GGAACCAAGGCAAACTACCA												
slat 85	F: TCGAGCACCTTGTTGAAATG	(CAT) <sub>8</sub>	70*	8 + 30	192-265	4	0.29	0.4 06		17	0.91 3	0.077 7	FJ573209
	R: GAGCCACGGTTTAGCACCTA												
Mean						7		0.34 5	0.5 86	14	0.77 1	0.139	

---

Subscript after repeat motif gives the number of repeats observed in the original sequence.  $T_a$ , annealing temperature (°C);  $n$  number of cycles. \*Starting from 70°C, the annealing temperature is decreased by -2.0°C per cycle for the first 8 cycles, then is stabilized at 54°C for the remaining 30 cycles.

Appendix 3: *Transcriptome* resources for two highly divergent *Silene latifolia* populations <sup>9</sup>

---

<sup>9</sup> Formatted for submission as a coauthored resource note: Fields, P.D., Weingartner, L. A., Taylor, D. R., and L. F. Delph

## INTRODUCTION

In order to understand the genetic architecture of local adaptation in two highly divergent populations of *Silene latifolia* we collected RNA-seq data on four individuals, two from each population (one male and one female). We provide information concerning the number of high-quality SNPs that differentiate the two divergent species.

## DATA ACCESS

*Sequence files* – Sequence files (.fq) can be found on NCBI Sequence Read Archive under project number: *input* (see table 1 for individual accession numbers)

*Reference file* – Reference transcriptome (Pooled\_reads\_transcriptome.fa, 341,132 contigs, 357.7 million base pairs, median contig size = 660bp, mean contig size = 1048bp, maximum contig size = 15252, N50 = 1741, N90 = 433) and is accessible on DRYAD (*input later*)

*Sequence alignment files* – Sequence alignments (one .bam file per individual) can be found on NCBI Sequence Read Archive under project number: *input*

*SNP file* – SNP tables (one .txt file per transcriptome) are accessible on DRYAD (*input*)

*Coverage file* – Coverage per gene and per individual (one .txt file) is accessible on DRYAD (*input*)

*Adaptor contaminant file* – Adaptors were removed by GENEWIZ, Inc.

*Script files* – Generating the reference transcriptome, as well as individual transcriptomes, was done through the use of individual bash commands.

**META INFORMATION**

*Sequencing center* – GENEWIZ, Inc. (South Plainfield, NJ, USA,

[http://www.genewiz.com/public/complete\\_solution.aspx](http://www.genewiz.com/public/complete_solution.aspx)).

*Platform and model* – Illumina HiSeq2000

*Design description* – We sampled two individuals per population (one male, one female), from two populations known to harbor most of the phenotypic diversity seen within the species European range. The goals were to identify population specific polymorphism in coding sequence.

*Run date* – 2012-12-04

**LIBRARY**

*Strategy* – normalized cDNA

*Taxa* – *Silene latifolia*

*Tissue* – Young leaf, bud, and root tissue from plants approximately x years old were extracted individually, and then pooled in equal proportions to generate whole RNA samples to send to GENEWIZ, Inc., for library preparation and sequencing.

*Location* – see Table 1

*Sample handling to prevent possible contamination* – We germinated all achenes at the University of British Columbia (Vancouver, Canada) and grew them for approximately two months in growth chambers (12 hours of daylight at 22 degrees). Then, we harvested young



leaf/stem tissue, flash froze it in liquid nitrogen and kept it at -80 degrees. Once sequencing was performed, sequences were cleaned to remove low quality reads and potential adaptors sequences using TRIMMOMATIC (Lohse *et al.* 2012). Alignment to the reference dataset also reduced contaminating reads (see pipeline description below).

*Additional sample information* – see Table 1 *Layout* – Paired end reads (2 X 100 bp or 2 X 101 bp)

*Library construction protocol* –For each individual, we extracted RNA using a modified RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and the RNA integrity (RIN) was checked with the RNA6000 Nano Assay using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). cDNA library preparations, sequencing reactions, and initial bioinformatics analysis were conducted at GENEWIZ, Inc. (South Plainfield, NJ, USA). Illumina TruSeq RNA Library Preparation, clustering, and sequencing reagents were used throughout the process following the manufacturer's recommendations (Illumina, San Diego, CA, USA). Briefly, mRNAs were purified using poly-T oligo-attached magnetic beads and then fragmented. The first and the second strand cDNAs were synthesized and end repaired. Adapters were ligated after adenylation of the 3'ends followed by cDNA template enrichment and barcode addition for multiplexing by limited cycle PCR. cDNA libraries were validated using a High Sensitivity Chip on the Agilent 2100 Bioanalyzer. The cDNA library was quantified using Qubit 2.0 Fluorometer and by qPCR. The samples were clustered on two lanes of a flow cell, using the cBOT from Illumina. After clustering, the samples were loaded on the Illumina HiSeq 2000 instrument according to manufacturer's instructions.

*Nominal size (paired) of fragments sequenced – 500 bp*

## **SEQUENCE PROCESSING**

*Pipeline* – Verification of the full removal of Illumina adapters was made with the use of Trimmomatic (Lohse *et al.* 2012). Read quality was subsequently assessed through the use of FastQC (Patel& Jain 2012). Quality inspection did not indicate a distinct drop off in read quality necessitating trimming of read lengths. Paired reads were interleaved through the use of custom Python scripts part of the khmer package (Brown *et al.* 2012). FASTX toolkit (Pearson *et al.* 1997) was used to remove reads with Q scores < 30. FastQC was used once again to verify the removal of low quality reads. Removal of duplicate reads was made with SAMtools (Li *et al.* 2009), leaving a mean of 82.6% reads retained from the original raw reads.

*De novo* assembly of the four individual transcriptomes, as well as a consensus transcriptome (concatenation of the left and right reads for each sample was concatenated) was made with the use of Trinity (Grabherr *et al.* 2011), with a default kmer value of 25 (other *de novo* transcriptome assemblers that allow for variation in kmer value, for example, Velvet-Oases or SOAPdenovo-Tran v.2, might reasonably be expected to yield longer contigs in some cases, but we believe the utilized protocol is sufficient for obtaining adequate numbers of potential SNPs given our experimental goals). Assembly of the concatenated reads was not possible with local machines, and so application was made to the XSEDE grid computing services for allocation to use Trinity on the Pittsburgh supercomputing center's Blacklight supercomputer. The number of contigs resulting from individual samples varied between 151835 and 174912, with a mean of 163053. Average

contig length ranged from 970bp to 1077bp, and a mean 1036bp. The N50 contig length ranged from 1701bp and 1846bp, and a mean 1793bp. The N90 contig length ranged from 373bp and 431bp, and a mean 407bp. Our consensus assembly resulted in 352654 contigs, with 917bp, 1524bp, and 367bp for the mean contig, N50, and N90 values.

Individual samples were aligned to the consensus Trinity.fasta assembly file using Bowtie2 (Langmead & Salzberg 2012) using paired-end options. The percentage of reads successfully mapped to the consensus assembly was consistently at least 90%. SAMtools mpileup was used to identify SNP polymorphisms, and output to a vcf file using bcftools. Vcftools were used to output all potential SNPs with a Phred score > 20. Of these high-confidence SNPs, we used a custom pipeline to retain only those SNPs that are polymorphic between the two populations (e.g., both parents from Spain are homozygous for “T” at a site, while both from Croatia are homozygous for “A” at that same site). Taking into consideration the potential presence of isoform transcripts part of the consensus Trinity assembly, a total of **7956 SNPs** differentiating the two focal populations.

Sequencing files were cleaned to remove low quality reads and potential adaptor sequences using TRIMMOMATIC (Lohse *et al.* 2012). The trimming parameters for adaptor removal (ILLUMINACLIP) were as follow: seed mismatch of 2, palindrome clip threshold of 40, simple clip threshold of 15. For trimming based on quality, the parameters were: minimum leading and trailing base quality of 2, minimum length of 36, minimum average base quality of 15 for sliding window of size 10.

Cleaned reads were then aligned against the reference transcriptome (51,468 contigs,

51.3M bp) using the Burrows-Wheeler Aligner (BWA, ALN with -q 20 and SAMPE commands, Li & Durbin 2009). SAMTOOLS (MPILEUP with -C50 and BCFTOOLS, Li *et al.* 2009) was used to call Single Nucleotide Polymorphisms (SNPs) using information from all samples for each species separately. SNPs therefore include both fixed differences from the *H. annuus* reference and intraspecific polymorphisms. Genotypes with Phred-scaled likelihoods below 20 were considered as missing, which corresponds to a genotyping accuracy of at least 99%. Custom R (R Core Team 2012) scripts were used to automate analysis.

*Runs* – 18 runs were submitted to NCBI SRA. Each run contains two (\_1.fq and \_2.fq) files. Runs were submitted as two different experiments given that samples were sequenced on two different platforms (see Table 1).

## RESULTS

*Total number of reads, percentage of reads surviving filtering, mean length, number of reads aligned, percentage of reads aligned, mean (median) number of reads aligned per contig* – Table 2

*Number of contigs with coverage > 0, number of base pairs with coverage > 0, total number of SNPs, number of fixed differences, Mean number of SNPs per 100 aligned base pairs* – Table 3

*Quality scoring system* – phred+33 *Quality scoring ASCII character range* – "!" to "J"

**LITERATURE CITED**

Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH (2012) A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. *arXiv:1203.4802v2*.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9, 357-359.

Lohse M, Bolger AM, Nagel A, *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research* 40, W622-W627.

Patel RK, Jain M (2012) NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE* 7, e30619.

Pearson WR, Wood T, Zhang Z, Miller W (1997) Comparison of DNA Sequences with Protein Sequences. *Genomics* 46, 24-36.

Appendix 3: LinkC: A software for the measurement of variance component estimation of linkage disequilibria in subdivided populations<sup>10</sup>

---

<sup>10</sup> Formatted for submission as a coauthored manuscript: Fields, P.D. and M. Morgan

**ABSTRACT**

LinkC is a reimplementation of the software softwares LinkDis and LinkDos, and allows for estimation of a number of population genetic summaries useful for understanding the presence of linkage disequilibrium, and its biological basis, including a hierarchical portioning of LD for subdivided populations, on large-scale, recently derived molecular population genetic datasets. LinkC is available as a binary for all modern operating systems, and is co-distributed with the GENEPOP package, as well in the form of source code. LinkC can also be accessed with any internet capable computer through Genepop on the Web: <http://genepop.curtin.edu.au/>

Keywords: linkage disequilibrium, GENEPOP, population subdivision, epistatic-selection.

Linkage disequilibrium (LD), or the non-random association of alleles at two or more loci, is a central concept in population genetics (SLATKIN 2008). LD will be generated by both selective and neutral processes (OHTA 1982) above and beyond structural linkage, including natural selection on non-additive multi-locus interactions (epistasis), genetic drift, population subdivision and bottlenecks, and non-random mating/inbreeding (SLATKIN 2008).

While there exist a number of software packages for measuring different forms of LD, perhaps the most frequently used package for analysis of a small number of molecular markers for ecological genetic study remains Garnier-Gere and Dillmann's (1992) LinkDos (originally a revision of the program LinkDis provided by BLACK and KRAFSUR (1985)). Here we describe a full revision of the LinkDos, LinkC, which provides the full functionality of the aforementioned software packages on datasets more consummate with ongoing population genetic research. Specifically, in recent years, population genetic studies have benefitted from using an increasing number of individuals and loci (LASCoux and PETIT 2010). Using a comprehensive literature search, Guichoux *et al.* (2011) identified approximately 8000 published population genetic analyses utilizing simple sequence repeats (SSRs) and single nucleotide polymorphism (SNPs) in 2009 alone. Continued advances in second-generation sequencing technologies and precipitous decreases in costs, combined with advances in analytical techniques focused upon these specific datasets, promise to accelerate these trends (CAO *et al.* 2011; WANG and HEY 2010). LinkC is not limited by any particular, pre-defined allocation of populations and/or loci, rather it is only limited by computer hardware limits.



The utility of the program, both in its original conception and the presented revision, is the computation of population specific allele frequency tables, the estimation of the COCKERHAM and WEIR (1977) measure of linkage disequilibrium (with a significance test based upon the  $\chi^2$  statistic), and, most significantly, a variance partitioning of linkage disequilibrium as described by OHTA (1982). This later summary, to the author's knowledge, is currently estimable only through the LinkDos software package, though it has been shown to be particularly important in reference to the further understanding of epistatic interactions their role in local adaptation (HUANG *et al.* 2012; MA *et al.* 2010).

Inputs for LinkC inputs can be generated with any form of text editor or spreadsheet program, or through the use of a utility available in the GENEPOP package (ROUSSET 2008). LinkC can also be accessed on any computer with an Internet connection through the a publically available webserver, <http://genepop.curtin.edu.au/>, replacing the current "LinkDos on the Web". Results are available as tab-delimited files suitable for import into spreadsheet or database programs.

LinkC is written entirely in the C programming language, and so is capable of functioning on any modern OS (flavors of Linux, Mac OSX, and Windows 7/8). Multi-platform binaries are available as part of the GENEPOP package, <http://kimura.univ-montp2.fr/~rousset/Genepop.htm>. Source code is available via request from the corresponding author, or through github at <https://github.com/crd1859/linkc>.

## LITERATURE CITED

- BLACK, W. C., and E. S. KRAFSUR, 1985 A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. TAG Theoretical and Applied Genetics **70**: 491-496.
- CAO, J., K. SCHNEEBERGER, S. OSSOWSKI, T. GUNTHER, S. BENDER *et al.*, 2011 Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nature Genetics **43**: 956-U960.
- COCKERHAM, C. C., and B. S. WEIR, 1977 Digenic descent measures for finite populations. Genetical Research **30**: 121-147.
- GARNIER-GERE, P., and C. DILLMANN, 1992 A computer program for testing pairwise linkage disequilibria in subdivided populations, pp. 239 in *J Hered.*
- GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LEGER *et al.*, 2011 Current trends in microsatellite genotyping. Molecular Ecology Resources **11**: 591-611.
- HUANG, W., S. RICHARDS, M. A. CARBONE, D. ZHU, R. R. H. ANHOLT *et al.*, 2012 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. Proceedings of the National Academy of Sciences.
- LASCOUX, M., and R. J. PETIT, 2010 The 'New Wave' in plant demographic inference: more loci and more individuals. Molecular Ecology **19**: 1075-1078.
- MA, X.-F., D. HALL, K. R. ST ONGE, S. JANSSON and P. K. INGVARSSON, 2010 Genetic Differentiation, Clinal Variation and Phenotypic Associations With Growth Cessation Across the *Populus tremula* Photoperiodic Pathway. Genetics **186**: 1033-1044.

- OHATA, T., 1982 Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**: 1940-1944.
- ROUSSET, F., 2008 GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**: 103-106.
- SLATKIN, M., 2008 Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**: 477-485.
- WANG, Y., and J. HEY, 2010 Estimating divergence parameters with small samples from a large number of loci. *Genetics* **184**: 363-U390.

## Appendix 4: Modification in BayeScan to incorporate GESTE-like environment association<sup>11</sup>

---

<sup>11</sup> Modifications conducted by Peter D. Fields, Matthieu Foll, and Oscar Gaggiotti

## NEW MODEL

In BayeScan we decompose the genetic differentiation in two components:

$$\ln(\theta_{ij}) = \alpha_i + \beta_j$$

and:

$$F_{ST}^{ij} = \frac{1}{1 + \theta_{ij}}$$

In this model  $\theta_{ij}$  is not an actual parameter of the model, as it is replaced by  $\alpha_i$  and  $\beta_j$ .

We add a GESTE-like prior on the  $\beta_j$  components:

$$\beta_j \sim N(\tau_j; \sigma)$$

with:

$$\tau_j = \omega_0 + \sum_{k=1}^F \omega_k G_{jk}$$

$G_{jk}$  represents the  $k^{\text{th}}$  factor for population  $j$ , and here also  $\tau_j$  is not an actual parameter of the model. We had to change some notations compared to the original paper describing GESTE:  $\tau$  replaces what we used to write  $\mu$  (as it is now used for AFLP band intensity in BayeScan), and  $\omega$  replaces what we used to write  $\alpha$  (the locus-specific components in BayeScan).

The DAG of the new model is given in Figure 1, and the posterior distribution can now be written as:

$$\pi(p, \alpha, \beta, \sigma, \omega | N) \sim L(p, \alpha, \beta) p(\beta | \omega, \sigma) p(\alpha) p(p) p(\sigma) p(\omega)$$

$L(p, \alpha, \beta)$  is the same likelihood and  $p(\alpha)$  the same prior as in BayeScan. Note that the different likelihood functions available in BayeScan are also used in this new model (codominant markers, dominant markers, SNPs and AFLP band intensity).  $p(\beta | \omega, \sigma)$ ,  $p(\sigma)$  and  $p(\omega)$  are the same prior as in GESTE.  $p(p)$  was already the same in BayeScan and GESTE. We perform the same reversible jump on  $\omega$  coefficients as in GESTE, and the same reversible jump on  $\alpha$  as in BayeScan (see implementation below).

## IMPLEMENTATION

In practice I started from the code of BayeScan 2.1 command line version and added the new parameters and functions by copying portions of the code of GESTE 2.0. The only modification we had to do was because in GESTE we had the normal prior on  $\ln(\theta_j)$ , and now it is directly on  $\beta_j$ . Fortunately the change is minimal as  $\beta_j$  already had a Normal prior in BayeScan, and it affects only the update functions of  $\beta$ ,  $\omega$  and  $\sigma$ . In these function we had to: (i) replace the parameter of the normal prior for  $\beta_j$  which are now  $\tau_j$  and  $\sigma$ , and (ii)  $\ln(\theta_j)$  by  $\beta_j$ .

The new program is now able to perform pure BayeScan type of analyses by setting the number of factors to 0 using “-nf 0” command line option. It is also able to perform

pure GESTE type of analyses, by ignoring locus-specific parameters  $\alpha$  using “-fstat” command line option. We added an option (“-norj\_locus”) not originally present in BayeScan to estimate the locus-specific parameters  $\alpha$  without performing a reversible jump. Note that the reversible jump on environmental factors is still performed in this case but can also be disabled using the “-norj” command line option just as in the original GESTE program.

Both GESTE-like output file (compatible with the plot program provided) and BayeScan like output files (compatible with the R plot function provided) are produced. Note that some information are redundant in these files but we them for back compatibility reasons. As in BayeScan original program, locus-specific components are not output by default, but can be using the “-all\_trace” option.

Figure 1: Directing acyclic graph (DAG) of the model. For the sake of simplicity, we show the DAG for codominant markers. For other likelihood functions, see BayeScan references. The square node denotes known quantity (i.e., data) and circles represent parameters to be estimated. Lines between nodes represent direct stochastic relationships within the model. The variables within each node correspond to the different model parameters discussed in the text. Variable represented within a dashed circle are not actually parameters of the model as they be calculated directly from actual parameters, but we represent it for a better understanding of the diagram.



