Conserving Connectivity: Ecological Determinants of Gene Flow in Plants at the Landscape Scale

by

Michelle Francis DiLeo

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Ecology and Evolutionary Biology
University of Toronto

© Copyright by Michelle F. DiLeo 2016
Conserving Connectivity: Ecological Determinants of Gene Flow in Plants at the Landscape Scale

Michelle F. DiLeo
Doctor of Philosophy
Department of Ecology and Evolutionary Biology
University of Toronto
2016

Abstract

With intensifying global pressures of habitat loss and fragmentation, there is an increasing need to manage landscapes in a way that maintains gene flow among previously continuous populations. This requires a detailed understanding of the scale of gene flow in nature and of how aspects of the landscape support or inhibit movement. In plants, gene flow occurs through both pollen and seed, each of which may be carried by multiple vectors. This makes it challenging to characterize, and as a consequence, our understanding of the local and landscape drivers of gene flow are limited in plants compared to animals. My thesis addresses this deficiency by quantifying the contributions of landscape structure, individual plant characteristics, and directed seed dispersal to gene flow of plants in fragmented landscapes. First, I used simulations to test if models of pollen flow could be improved by incorporating individual plant characteristics that affect attractiveness to pollinators. The results showed that inter-individual variation in attractiveness explained significantly more variation than inter-mate distance. Second, I took advantage of a network of calcareous grasslands in Germany to quantify the determinants of pollen and seed-mediated gene flow in a
specialist herb, *Pulsatilla vulgaris*. Using 1,449 individuals from 57 populations, genotyped at seven newly developed markers, I tested the efficacy of an ecological shepherding network to maintain seed-mediated gene flow among *P. vulgaris* populations. I found that (i) shepherding distance explained genetic differentiation better than geographic distance among populations, (ii) populations that were well connected within the network had significantly higher genetic diversity, and (iii) genetic diversity was significantly positively correlated with fitness-related traits. Paternity analysis on a subset of seven populations revealed high rates of self-pollination. Within-population patterns of pollen flow correlated with floral resources measured at the scale of individuals and patches, whereas among-population pollen flow correlated with floral resources measured at the patch scale and landscape context measured at intermediate and large spatial scales. Together my results highlight the importance of using multi-scale and multi-vector approaches when modeling gene flow in plants, and suggests that typical models based on distance alone may be insufficient to capture the complexity of pollination and seed dispersal.
Acknowledgments

My PhD would not be possible without the amazing support network provided by my family, friends, colleagues and supervisory committee.

First I thank my supervisor Helene Wagner for her incredible mentorship and support throughout these last five years. I thank Helene for constantly pulling my head out of the details to keep me focused on the bigger picture – for papers, career, and life in general.

I thank my committee Marie-Josée Fortin and Marc Johnson for many helpful conversations, valuable feedback on my chapters, and for welcoming me into their labs during Helene’s sabbatical year. I thank Marc for working with me to help finalize my thesis on a tight timeline.

I thank the past and current members of the Wagner Lab, Yessica Rico, Shekhar Biswas, Luis Arias, Adam Hadley, and Mariana Chavez Pesqueira, and my fellow grad students at UTM for helpful academic discussions and much needed non-academic distractions. I especially thank my running buddy Laura Junker for keeping me fit but also knowing when to send emergency chocolate my way, and Ilona Naujokaitis-Lewis for providing a voice of reason during many late night conversations.

To my parents, Debra and Gerry DiLeo, for keeping me well fed and for their unwavering support despite their child’s odd fascination with plant sexy parts.

Finally, to my husband Bruno Guimarães, for keeping me going through the tough spots and always encouraging me to follow my dreams, even if it involves selling everything we own and moving half way across the world.
# Table of Contents

Acknowledgments ........................................................................................................ iv
Table of Contents .......................................................................................................... v
List of Tables .................................................................................................................. x
List of Figures ................................................................................................................ xii
List of Appendices ......................................................................................................... xiv
Chapter 1 ...................................................................................................................... 1
  1.1 Gene Flow as an Important Evolutionary Force .............................................. 1
  1.2 Managing Landscapes ...................................................................................... 1
  1.3 Quantifying Gene Flow in Plants .................................................................... 3
  1.4 Ecological Determinant of Gene Flow in Plants .......................................... 5
  1.5 Main Study System .......................................................................................... 7
  1.6 Main Study Species ......................................................................................... 8
  1.7 Thesis Outline and Objectives ........................................................................ 9
Chapter 2 ..................................................................................................................... 15
  2.1 Abstract ............................................................................................................. 15
  2.2 Introduction ....................................................................................................... 16
  2.3 Literature Review ............................................................................................. 20
  2.4 Characterizing Gaps in the Literature ............................................................. 22
  2.5 Habitat Amount Versus Fragmentation .......................................................... 22
  2.6 Landscape Effects on Gene Flow and Genetic Drift ...................................... 23
Chapter 2

2.7 Genetic Diversity versus Differentiation .......................................................... 26
2.8 Importance of Matrix ......................................................................................... 29
2.9 Testing the Importance of Habitat Amount versus Fragmentation with Genetic Data.......................................................... 31
2.10 Conclusions ..................................................................................................... 33
2.11 Acknowledgements ......................................................................................... 33

Chapter 3

3 The Gravity of Pollination: Integrating At-Site Features into Spatial Analysis of Contemporary Pollen Movement .......................................................... 39
3.1 Abstract ............................................................................................................ 39
3.2 Introduction ..................................................................................................... 40
3.3 Methods .......................................................................................................... 43
  3.3.1 Gravity Models ............................................................................................ 43
  3.3.2 Simulation Models ....................................................................................... 44
  3.3.3 Empirical Gravity Model with Cornus florida ............................................. 46
3.4 Results ............................................................................................................. 48
  3.4.1 Simulation Results ...................................................................................... 48
  3.4.2 Empirical Gravity Model with Cornus florida ............................................. 49
3.5 Discussion ....................................................................................................... 49
  3.5.1 Connectivity Networks Using Pollination Graphs ...................................... 50
  3.5.2 Effects of At-Site Features ......................................................................... 51
  3.5.3 Empirical Gravity Model: Cornus florida ............................................... 52
  3.5.4 Conclusions ............................................................................................... 53
3.6 Acknowledgements ......................................................................................... 54
3.7 Author Contributions ...................................................................................... 54

Chapter 4 ............................................................................................................... 59
# 4 Highly Polymorphic Microsatellite Markers in *Pulsatilla vulgaris* (Ranunculaceae) Using Next Generation Sequencing ................................................................. 59

4.1 Abstract ................................................................................................................................................................................................. 59

4.2 Introduction ........................................................................................................................................................................................................ 60

4.3 Methods and Results ......................................................................................................................................................................................... 60

4.4 Conclusions ............................................................................................................................................................................................................ 62

4.5 Acknowledgements ................................................................................................................................................................................................... 63

Chapter 5 .................................................................................................................................................................................................................... 67

# 5 An Ecological Network Maintains Genetic Diversity and Fitness of a Flagship Wildflower ........................................................................................................................................................................................................ 67

5.1 Abstract ................................................................................................................................................................................................................. 67

5.2 Introduction ............................................................................................................................................................................................................ 68

5.3 Methods ....................................................................................................................................................................................................................... 71

5.3.1 Data Collection .......................................................................................................................................................................................................... 71

5.3.2 Genetic Analysis of Functional Connectivity .................................................................................................................................................................. 73

5.3.3 Genetic Diversity ....................................................................................................................................................................................................... 74

5.3.4 Fitness .......................................................................................................................................................................................................................... 76

5.4 Results ......................................................................................................................................................................................................................... 77

5.4.1 Genetic Analysis of Functional Connectivity .................................................................................................................................................................. 77

5.4.2 Effects of Shepherding Connectivity on Genetic Diversity ................................................................................................................................................................ 78

5.4.3 Effect of Genetic Diversity on Fitness .................................................................................................................................................................. 79

5.5 Discussion .................................................................................................................................................................................................................. 79

5.5.1 Functional Connectivity .................................................................................................................................................................................................. 80

5.5.2 Effect of Shepherding Connectivity on Genetic Diversity .................................................................................................................................................................. 82

5.5.3 Effect of Genetic Diversity on Fitness .................................................................................................................................................................. 83

5.5.4 Conclusions ............................................................................................................................................................................................................ 84
5.6 Acknowledgements........................................................................................................85

Chapter 6................................................................................................................................92

6 Multi-Scale Drivers of Contemporary Pollen Flow in the Insect-Pollinated Herb, 
_Pulsatilla vulgaris_ ..............................................................................................................92

6.1 Abstract ............................................................................................................................92

6.2 Introduction .......................................................................................................................93

6.3 Methods ............................................................................................................................95

6.3.1 Study Species .................................................................................................................95

6.3.2 Study Area and Sampling ............................................................................................96

6.3.3 Genotyping ....................................................................................................................97

6.3.4 Reconstructing Contemporary Pollen Flow .................................................................97

6.3.5 Family-Level Analysis .................................................................................................99

6.3.6 Population-Level Analysis ..........................................................................................100

6.4 Results .............................................................................................................................103

6.4.1 Reconstructing Contemporary Pollen Flow .................................................................103

6.4.2 Family-Level Analysis .................................................................................................103

6.4.3 Population-Level Analysis ..........................................................................................105

6.5 Discussion .........................................................................................................................106

6.5.1 Pollen Flow as a Multi-Scale Process .........................................................................106

6.5.2 Landscape Effects on Contemporary Pollen Flow .....................................................107

6.5.3 Consequences for _P. vulgaris_ Populations .................................................................109

6.5.4 Conclusions .................................................................................................................110

Chapter 7................................................................................................................................120

7 Synthesis .............................................................................................................................120

7.1 Summary .........................................................................................................................120

7.2 Beyond Isolation-by-Distance .......................................................................................121
7.3 The Importance of Considering Multiple Vectors .................................................... 123

7.4 Conservation Implications for P. vulgaris ................................................................. 124

7.5 Limitations and Future Directions ............................................................................ 126

References ...................................................................................................................... 127

Appendices ..................................................................................................................... 148

Copyright Acknowledgements ....................................................................................... 162
List of Tables

Table 2-1: Summary of results of 11 node-based studies that controlled for the independent effects of local habitat amount and fragmentation on genetic variation. ..... 34

Table 2-2: Summary of the effects of habitat amount, patch size, and configuration on genetic variability for 56 node-based studies identified in the literature from 2011-2015. ................................................................. 35

Table 3-1: Independent predictor variables used to build gravity models to explain connectivity by pollen flow in C. florida.......................................................... 55

Table 3-2: Relative model fit of gravity models constructed with spatial distance and at-site and between-site parameters and fit to C. florida connectivity data. ................. 56

Table 4-1: Characteristics of 18 de novo developed microsatellite markers for Pulsatilla vulgaris ........................................................................................................... 64

Table 4-2: Results from HWE exact tests for diploid markers developed for Pulsatilla vulgaris in two populations, A03 and A25. ................................................................. 66

Table 5-1: Results from Mantel tests between genetic distance and geographic distance (IBD) among P. vulgaris populations, and genetic diversity and shepherding distance (IBR), and partial Mantel tests between genetic diversity and geographic distance controlling for shepherding distance (IBD|IBR), and genetic distance and shepherding distance controlling for geographic distance (IBR|IBD). .................................................. 86

Table 5-2 Results and model fit of linear mixed effect models testing the effect of geographic connectivity (\(S_{IBD}\)), shepherding connectivity (\(S_{IBR}\)) and population size on genetic diversity (\(A_r\)) of P. vulgaris populations. ................................................................. 87

Table 6-1: Sample sizes for the seven P. vulgaris populations where paternity analysis was conducted........................................................................................................... 112
Table 6-2: Results of model selection for linear mixed models estimating the relationship of mean pollination distance per mother plant with flower density around maternal plants and mean distance of maternal plants to all potential sires within the population (mean neighbour distance) (a), linear mixed models estimating the relationship of correlated paternity with flower density and mean neighbour distance (b), and generalized mixed models estimating the relationship of the proportion of selfed seeds per maternal plant with flower density and mean neighbour distance (c). ................................................................. 113

Table 6-3: Estimates for population-level logistic regressions, showing the relationship of the proportion of selfed seeds or proportion of immigrants per population with population size (number of flowering plants) and landscape composition (forest and seminatural habitat) measured at increasing radii from population centroids. ............................................. 114
List of Figures

Figure 1-1: An example of a calcareous grassland patch in the Franconian Alb, Germany. .................................................................................................................................................................................. 13

Figure 1-2: *Pulsatilla vulgaris* Mill. ............................................................................................................................................................................................................. 14

Figure 2-1: Approaches for quantifying the role of landscape structure on genetic variation. .................................................................................................................................................................................................................... 36

Figure 2-2: Mosaic plot showing the proportion of studies identified from the literature between 2011-2015 that were conducted at the link and node level across taxa (a), and barplots showing the number times configuration only, matrix, or habitat amount were used as predictors in genetic models.......................................................................................................................................................................................................................... 37

Figure 2-3: Predicted relationships between landscape structure, genetic drift and gene flow, and genetic diversity and differentiation. .................................................................................................................................................................................................................... 38

Figure 3-1: Pollen pool differentiation (Φ<sub>FT</sub>) of populations simulated at different pollen donor densities. ........................................................................................................................................................................................................ 57

Figure 3-2: The influence of tree detectability (simulated as *H*) in gravity models for high and low simulated pollen donor densities.................................................................................................................................................................................................................................. 58

Figure 5-1: Sampled populations and grazing treatment of *P. vulgaris* in the study region in the Franconian Alb, Germany.......................................................................................................................................................................................................... 88

Figure 5-2: Boxplots showing differences in genetic diversity (*A_r*) among grazing treatment of *P. vulgaris* populations........................................................................................................................................................................................................ 89

Figure 5-3: Scatterplots showing the relationships between genetic diversity (*A_r*) and a) connectivity index based on among-population shepherding distance (*S_i<sub>IBD</sub>*), b) connectivity index based on among-population geographic distance (*S_i<sub>IBR</sub>*), and c) population size class ........................................................................................................................................................................................................................................................................................................................................ 90
Figure 5-4: Scatterplots showing the relationships between a) seed set and genetic diversity ($A_r$), and b) mean seed mass and genetic diversity. ........................................... 91

Figure 6-1: Study area in the Franconian Alb, Germany, showing main land classification and all known *Pulsatilla vulgaris* populations in the region. ......................................... 115

Figure 6-2: Histograms showing observed (grey bars) and expected (white bars) within-population pollination distances. ................................................................. 116

Figure 6-3: Scatterplots showing the relationships of mean within-population pollination distance per mother and local floral density (a), mean pollination distance and mean neighbour distance (c), proportion of selfed seeds per mother and local floral density (b), and proportion of selfed seeds and mean neighbour distance (d)........................................... 117

Figure 6-4: Scatterplot showing the relationship of population size (number of flowering plants) with the proportion of selfed seeds per population and the proportion of immigrants per population. ................................................................. 118

Figure 6-5: Scatterplots showing the relationship of the proportion selfed seeds per population with the amount of forest measured within a 50 m radius (a) and seminatural habitat measured within a 1000 m radius (b), and the relationship of the proportion of immigrants per population with forest measured within 250 m (c) and seminatural habitat measured within 1000 m (d)................................................................. 119
List of Appendices

A 1: Boxplots showing correlation coefficients of gravity models estimated for *Cornus florida* as a function of bandwidth used to measure between-site ecological variables (Chapter 3). .................................................................................................................. 148

A 2: Boxplots showing the distribution of variance explained, $R^2$, for gravity models fit to simulated data based on saturated and pruned (by Pollination Graph) graph topologies as function of simulated pollen donor density (Chapter 3). .............................................. 149

A 3: Locations of ten populations where seeds were collect to measure fitness-related traits in Chapter 5................................................................................................................................................. 150

A 4: Supplementary Methods: The effect of connectivity and population size on genetic diversity ($A_r$) of *P. vulgaris* populations, with population size as a continuous variable (Chapter 5). ............................................................................................................. 151

A 5: Results of linear mixed effect models testing the effect of connectivity based on geographic distance among populations ($Si_{IBD}$), connectivity based on shepherding distance among populations ($Si_{IBR}$), and population size (continuous) on genetic diversity ($A_r$) of *P. vulgaris* populations (Chapter 5). .......................................................................................................................... 152

A 6: Scatterplot showing the relationship between population size and genetic diversity ($A_r$) across *P. vulgaris* populations (Chapter 5). .................................................................................................................. 153

A 7: Sequentially bonferroni-corrected *p*-values for Hardy-Weinberg exact tests conducted within populations for each microsatellite marker (Chapter 5)................. 154

A 8: Boxplots of alternate measure of genetic diversity, effective number of alleles (a), expected heterozygosity (b), and observed heterozygosity (c), across grazing treatments for *P. vulgaris* populations (Chapter 5). .................................................................................................. 156

A 9: Results of linear mixed effect models testing the effect of connectivity based on geographic distance among populations ($Si_{IBD}$), connectivity based on shepherding
distance among populations ($S_{IBR}$), and population size on three alternate measures of genetic diversity for *P. vulgaris* populations (Chapter 5)................................. 157

A 10 Results of linear mixed effect models showing the strength and significance of relationships between alternate genetic diversity metrics and fitness-related traits for ten populations of *P. vulgaris*. ........................................................................................................... 158

A 11: Spearman rank correlations (*rho*) among predictor variables for population-level analysis in Chapter 6.......................................................................................................................... 159

A 12: Scatterplot showing the relationship between the proportion of seed set and proportion of selfed seeds per maternal plant across seven *P. vulgaris* populations (Chapter 6). .......................................................................................................................... 160

A 13: Within population pollen flow trajectories, reconstructed using paternity analysis (Chapter 6). .......................................................................................................................... 161
Chapter 1

1 General Introduction

1.1 Gene Flow as an Important Evolutionary Force

Ecological and evolutionary dynamics are ultimately defined by the interplay of mutation, drift, gene flow and selection. Gene flow has traditionally been viewed as the great homogenizer - a force that opposes natural selection and inhibits local adaptation by swamping out the effects of new beneficial mutations in populations (Ellstrand 2014). In the early days of population genetics, gene flow was thus often treated as a nuisance, or viewed as unimportant relative to natural selection (Slatkin 1987). However, with the advent of high-resolution molecular markers, estimates of gene flow in wild populations have provided overwhelming evidence that it is indeed a pervasive and diversifying evolutionary force (Slatkin 1987, Ellstrand 2014). If habitat is fragmented and the population is thus subdivided, gene flow is essential to maintain regionally high effective population sizes, and to provide the levels of genetic diversity required for adaptation to future environmental change. In the same way, gene flow can promote the rapid spread of adaptive alleles throughout connected populations (Hanski et al. 2011), and allow genetic rescue of populations that are small, inbred, and facing extinction (Saccheri et al. 1998, Richards 2000, Keller and Waller 2002). In the face of climate change, gene flow will be essential for colonization of new habitat and for redistributing genetic variation as species’ ranges shift (Jump and Penuelas 2005, Kremer et al. 2012).

1.2 Managing Landscapes

With global habitat loss and fragmentation at an all-time high, there is an increasing need to manage landscapes in a way that maintains gene flow among previously contiguous populations
of plants and animals (Fischer and Lindenmayer 2007). This requires knowledge of the scale of contemporary gene flow to identify populations that are most at risk, and an understanding of how aspects of the landscape promote or inhibit dispersal and subsequent gene flow. Landscape genetics, which combines molecular markers with landscape and ecological data to test hypotheses about how landscape shapes gene flow, is particularly suited for this job and thus a useful tool for conservation (Manel et al. 2003, Keller et al. 2015). For example, a landscape genetic approach can be used to identify abrupt genetic discontinuities across the landscape where genetic connectivity can be restored (Riley et al. 2006, Ruiz-Gonzalez et al. 2015). Likewise, one can identify existing natural corridors that can be prioritized for protection (Wang et al. 2009).

However, managing populations in fragmented landscapes based on genetic connectivity alone may not be sufficient. Measures of gene flow only capture dispersal events that lead to mating, and thus may miss important aspects of how landscape structure influences demographic connectivity and population dynamics (Lowe and Allendorf 2010a). Theoretical and simulation work has shown that demographic (e.g. patch occupancy, species abundance and richness) and genetic outcomes (e.g. genetic differentiation, genetic diversity) can correlate with different aspects of landscape structure and at different spatial scales (Ezard and Travis 2006, Bruggeman et al. 2010, Cushman et al. 2012, Jackson and Fahrig 2014). It is well known from the empirical landscape ecological literature that patch occupancy, abundance, and richness of species are often more influenced by the amount of habitat in a landscape rather than the degree of fragmentation (Mortelliti et al. 2012, Fahrig 2013, Hornseth et al. 2014). In contrast, landscape genetic studies tend to emphasize the importance of fragmentation (i.e. spatial configuration of populations) and matrix permeability for determining gene flow and genetic population structure. Thus to effectively manage populations we must consider both demographic and genetic
conservation targets (Landguth et al. 2014), and a deeper understanding of how demography shapes genetic outcomes and vise-versa is urgently required.

### 1.3 Quantifying Gene Flow in Plants

In plants, gene flow occurs through the movement of two distinct propagules: seed and pollen, each of which may be carried by multiple abiotic and biotic vectors that contribute to overall genetic connectivity to different degrees. This makes it particularly challenging to characterize the landscape determinants of gene flow in plants, as the effects are indirect, and thus requires knowledge of how dispersal vectors move through the landscape. As a consequence, our understanding of how landscape shapes gene flow is limited in plants compared to animals (Holderegger et al. 2010, Storfer et al. 2010).

The small size of pollen and often seeds makes direct tracking of plant dispersal challenging, and thus estimating gene flow using molecular makers is preferred. Molecular markers can provide direct or indirect estimates of gene flow. Indirect methods employ neutral genetic markers to quantify genetic differentiation among populations (e.g. $F_{st}$) as a proxy of gene flow (i.e. populations that are less differentiated share more gene flow), or can be used to estimate migration rates using coalescent or Bayesian approaches (Beerli and Felsenstein 2001, Wilson and Rannala 2003). Nuclear microsatellites are a typical marker of choice because they are highly variable and evolve quickly, giving estimates of gene flow over more recent evolutionary timescales (Estoup and Angers 1998). However, nuclear markers such as microsatellites come with the caveat that they are biparentally inherited and thus carry the signal of both pollen and seed-mediated gene flow. This complicates hypothesis testing about specific landscape processes that may influence one propagule over the other. In angiosperms, chloroplasts are maternally inherited and thus seed-dispersed, and estimates of genetic
differentiation based on chloroplast markers can be used to strictly track seed-mediated gene flow (Ebert and Peakall 2009, Wheeler et al. 2014). By comparing genetic differentiation based on nuclear and chloroplast markers, one can test the relative contribution of pollen- versus seed-mediated gene flow to overall genetic structure (Ennos 1994). However chloroplast genomes tend to be highly conserved and unlikely to provide the resolution needed to detect gene flow at small spatial scales (but see McCauley 1997).

In contrast, direct approaches using parentage-based analysis can be used to separately track contemporary pollen and seed flow (Ashley 2010). Paternity analysis allows the reconstruction of realized pollen flow by assigning the paternal contribution of a seed genotype to a single most likely father, assuming the mother is known. When offspring (seeds or seedlings) can be reliably collected post-dispersal, a full parentage analysis can be conducted to reconstruct realized seed flow. However, paternity and full parentage analyses require complete sampling of potential fathers (and mothers in the case of parentage analysis). This is prohibitive for large populations or species found at high densities (e.g. herbs), and thus the vast majority of our knowledge of contemporary pollen and seed flow comes from trees sampled at small spatial scales (Sork and Smouse 2006, Holderegger et al. 2010). A related approach, two-generation analysis (TwoGener; Smouse et al. 2001, Dyer et al. 2004), can be applied to quantify pollen-mediated gene flow without the need to sample all potential fathers in a population. In this approach, multiple offspring are genotyped per mother and a “pollen cloud” is established by subtracting the maternal contribution from the offspring genotypes. Genetic differentiation is then measured among pollen clouds, giving an estimate of pollen flow among sampled mothers (Smouse et al. 2001).
Direct and indirect approaches allow estimation of gene flow over different temporal scales. Indirect measures of genetic differentiation (e.g. $F_{st}$) contain the signal of gene flow over many generations. This comes with the caveat that legacies of past population change may be reflected in measures of genetic differentiation, possibly confounding inference of contemporary landscape effects on gene flow (Anderson et al. 2010, Epps and Keyghobadi 2015). Similarly, contemporary landscape or environmental change may take many generations before it is reflected in the signature of genetic differentiation (Landguth et al. 2010). In contrast, direct estimation with parentage analysis allows tracking of gene flow within a single generation (i.e. the offspring generation). However, this approach does not take into account possible mortality of offspring or establishment success of seeds, and thus the contribution of contemporary gene flow to overall genetic structure of populations is often unclear. Ultimately, using a combination of indirect and direct approaches will lead to better inference of the spatial and temporal scale of gene flow in plants.

1.4 Ecological Determinant of Gene Flow in Plants

Genetic differentiation among individual plants or plant populations is typically modeled as a function of isolation-by-distance (IBD), where gene flow among pairwise individuals or populations decreases as a function of the distance between them (Wright 1943). However, this neutral model ignores potential interactions of dispersal vectors with aspects of the intervening landscape. In the most extreme case, physical landscape barriers may completely restrict gene flow between populations found on either side (i.e. isolation-by-barrier; Cushman et al. 2006). But the effect of landscape can be subtler, where certain types of features may be less amenable to movement of vectors and support lower rates of gene flow than others (McRae 2006). Effective distances based on this so-called “landscape resistance” have been found to explain substantially more variation in pairwise genetic differentiation than geographic distance for a
large number of animal species, but this has rarely been tested in plants (Holderegger et al. 2010; Storfer et al. 2010; Manel and Holderegger, 2013; but see McRae and Beier 2007, Kamm et al. 2010, Lander et al. 2011, Dyer et al. 2012).

The structure of the intervening landscape is expected to be a particularly important predictor of gene flow for animal-pollinated or dispersed species. For example, pollinators such as birds and bees have been shown to be susceptible to habitat fragmentation and change their movement based on the permeability of the intervening landscape (Levey et al. 2005, Hadley and Betts 2009, Davis et al. 2010, Cranmer et al. 2012, Aguirre-Gutierrez et al. 2015, Bartlett et al. 2016). However, many plant species interact with multiple pollen and seed vectors that may respond to the same landscape in different ways and contribute to gene flow over different spatial scales (Nathan et al. 2008, Kramer et al. 2011). Thus the first major challenge of quantifying the landscape determinants of gene flow in plants is to identify the different species that may contribute to pollination or seed dispersal, and from those, determine which are likely to contribute most strongly. This requires differentiating between flower visitors versus pollinators, and undirected versus directed dispersal of seeds (i.e. by specialized disperses that deposit seeds disproportionately in suitable habitat; Wenny 2001). Seed flow is generally thought to be negligible in its contribution to genetic connectivity compared to pollen flow in most plants (Petit et al. 2005). However, directed dispersal of seeds has been implicated as an important source of long-distance gene flow in plants (Wenny 2001, Manzano and Malo 2006, Nathan 2006).

In addition to the effects of geographic distance and resistance of the intervening landscape (i.e. between-site factors), at-site factors such as features of individual plants can influence pollen-mediated gene flow by altering the attractiveness of plants to pollinators
Likewise, aspects of the local landscape such as canopy openness might influence the detectability of a plant to its pollinators (Sork et al. 2005). The effects of individual traits such as phenology, floral morphology, display size and scent on pollinator visitation patterns and reproductive success have been well studied (Barrett and Harder 1996, Ishii et al. 2008, Mitchell et al. 2009), however they are rarely considered in models of pollen-mediated gene flow. By increasing the attractiveness of individual plants or populations, these traits may draw pollinators from further away, reducing the effective distance among mates and influencing the spatial scale of genetic structure.

1.5 Main Study System

My main study system is a 10 x 15 km actively managed network of calcareous grasslands in the Franconian Alb, Germany. The region is characterized by a series of valleys and plateaus ranging in elevation from 410-610 m (Wagner et al. 2013). The plateaus contain a mix of agricultural fields, forest, grasslands and settlements. Calcareous grassland habitat (Fig. 1-1) is found in shallow soils on the plateaus, or more typically on steep, eroded slopes at the margin of the plateaus and valleys. Although semi-natural, calcareous grasslands are of high conservation value in the study area and generally as they represent one of the most biodiverse ecosystems in central and northern Europe (Butaye et al. 2005). Abandonment of traditional grazing practices over the past century led to a significant loss of calcareous grassland habitat and fragmentation of previously connected patches by encroaching forest (Dolek and Geyer 2002, Walker and Pinches 2011). Since 1989 an ecological network management program was implemented in the Franconian Alb to reconnect previously abandoned patches to core areas of grasslands via rotational sheep grazing. The ecological network consists of 400-800 ewes herded in both directions along three non-overlapping routes. Prior to the implementation of the management plan, a survey was conducted to record all vascular plants in existing grassland patches
(Boehmer et al. 1990). In 2009 this survey was repeated, and leaf material was taken from individuals of the specialist herb *Pulsatilla vulgaris* (*n*=1,499) from all 57 populations in the study area. Of the 57 populations, 18 occur in ‘core areas’ that have been grazed since medieval times (Jacobeit and Hornberger 1962, Dolek and Geyer 2002), and the remaining 39 occur in ‘previously abandoned’ patches (abandoned since at least 1960) that, since the implementation of the management program, have been either consistently grazed (every year, 3-5 times per season), intermittently grazed (only later in season or few years), or remained ungrazed.

1.6 Main Study Species

*Pulsatilla vulgaris* (Ranunculaceae) is a perennial herb of conservation concern and a flagship species of calcareous grasslands across central Europe. Over the last century, *P. vulgaris* has witnessed rapid decline and local extinction across its range (Walker and Pinches 2011). It is self-compatible and produces hermaphroditic flowers that are protogynous (Wells and Barling 1971, Jonsson et al. 1991). Most plants produce between 1-3 purple flowers (Fig. 1-2), but large specimens may produce many more (Wells and Barling 1971). Flowering takes place early (March-April) and lasts 4-6 weeks, and flowers are pollinated by a variety of insects, but predominately small-bodied Hymenoptera Apoidae in the genera *Lasiglossum, Andrena, and Osmia* (Kratochwil 1988, Fay and Barlow 2014). Each flower produces 40-100 seeds, each with a long-feathery style (Fig. 1-2). Although seeds appear to be adapted for wind dispersal, those carried by wind rarely make it further than 20 cm from the plant (Wells and Barling 1971). *P. vulgaris* is a tetraploid (*2n = 4x = 32*), however multiple polyploid chromosomal arrangements have been recorded from the British Isles (*2n = 16, 24, 48*; Wells and Barling 1971).
1.7 Thesis Outline and Objectives

My thesis quantifies the effects of landscape, individual plant characteristics, and directed seed dispersal on gene flow in plants. The first chapter characterizes limitations in typical landscape genetic study designs that prevent researchers from testing the independent effects of habitat loss versus fragmentation on genetic variation. My second chapter uses simulations to quantify the importance of considering at-site and individual plant characteristics in genetic models. The remainder of my thesis assesses the ecological determinants of seed- and pollen-mediated gene flow in an actively managed network of calcareous grasslands in the Franconian Alb, Germany. Pollen-flow is generally assumed to be the most important driver of genetic structure in plants (Petit et al. 2005), but in this system, a large-flock shepherding network has the potential to move seeds through directed dispersal, and thus contribute to connectivity at the landscape scale. As the first species to flower in calcareous grasslands in the Franconian Alb, *P. vulgaris* is also an important early resource for wild bees, and thus pollen-flow is likely to contribute strongly to genetic structure. This provides a unique opportunity to quantify the spatial scale at which each process (i.e. seed versus pollen) contributes to connectivity, and test the strength of relationships with purported dispersal vectors. I outline the specific objectives of each chapter below.

Chapter 2 presents a critical review of the landscape genetic literature of the last five years from a landscape ecological perspective. Landscape genetics, using molecular markers, is a useful tool for studying the effects of habitat fragmentation for species that are hard to track or have cryptic dispersal. However, the typical design of landscape genetic studies limits the ability to quantify the independent contributions of habitat loss versus fragmentation to population genetic structure. In this chapter I characterize this important caveat and draw on the landscape ecological literature to pose potential solutions. Specifically I ask: (1) Are landscape geneticists considering the independent effects of habitat amount versus fragmentation *per se* on
genetic variation? And (2) Do landscape geneticists equally consider the effects of landscape structure on genetic diversity versus genetic differentiation?

Chapter 3 uses simulations to test the importance of including features measured at the site of individual plants as predictors in models of contemporary pollen flow. Landscape genetic studies on connectivity typically test if aspects of the intervening landscape (“between-site” variables) explain gene flow beyond the effects of geographic distance. However, in animal-pollinated plants, aspects of individuals (i.e. “at-site” variables) such as floral display size may alter the attractiveness or detectability of plants to pollinators and may thus influence the scale of contemporary pollen flow. These at-site features are not typically considered in genetic models. I use simulations to test if including at-site variables improves the fit of pollen flow models beyond the effects of inter-plant distance. Using an empirical data set from the understory tree *Cornus florida*, this chapter further tests the relative contribution of geographic distance, at-site features, and intervening landscape features to patterns of contemporary pollen flow.

Chapter 4 describes newly developed nuclear microsatellite markers for *P. vulgaris*. Quantifying landscape effects on gene flow requires high-resolution, neutral molecular markers such as microsatellites. Most angiosperms are polyploid (Masterson 1994) and lack a reference genome, making the development of such markers particularly challenging for plants (Dufresne et al. 2014). Here I used next-generation sequencing to develop a suite of highly polymorphic nuclear microsatellite markers for the tetraploid *P. vulgaris*. Importantly, a subset of the markers I developed can be analyzed in a diploid fashion, which greatly simplifies downstream genetic analysis. Given that *P. vulgaris* is a declining species in Europe and is of high conservation
concern, these markers will be an important resource for the continued monitoring of populations.

**Chapter 5 tests the efficacy of an ecological shepherding network for maintaining seed-mediated gene flow, genetic diversity, and fitness of *P. vulgaris* populations.** Seed-mediated gene flow is commonly considered negligible in its contribution to genetic connectivity compared to pollen flow in most plant systems (Ennos 1994). However, an ecological shepherding network in the Franconian Alb presents a potentially important seed-dispersal vector for *P. vulgaris*. Here I test the contribution of the shepherding network to the maintenance of genetic connectivity, genetic diversity, and fitness of populations. Specifically, I ask: (1) Does the ecological network explain gene flow among *P. vulgaris* populations, as quantified by pairwise genetic differentiation? (2) Does the potential enhanced gene flow provided by the network translate to higher genetic diversity in connected populations? (3) Does higher genetic diversity translate to higher fitness of populations?

**Chapter 6 quantifies the ecological determinants of within- versus among-population contemporary pollen flow in *P. vulgaris*.** For generalist-pollinated plants like *P. vulgaris*, patterns of within- versus among-population contemporary pollen flow may depend on different sets of pollinators that respond to aspects of individual plants, patches, and local landscape in different ways and at different spatial scales. Specifically I ask: (1) Do within-population patterns of pollen flow follow a uniform distribution with respect to available potential fathers? (2) Are measures of within-population pollination (e.g. individual variation in selfing rates, correlated paternity, mean outcrossing distance) explained by local floral density and maternal plant isolation? (3) Are population-level selfing and pollen immigration rates
explained by landscape context of populations, and if so what is the most relevant spatial scale for each response?

By quantifying the ecological determinants of genetic connectivity in an herbaceous species, my thesis provides much needed empirical evidence of the interplay between landscape fragmentation, pollen and seed flow, and their contributions to genetic diversity, fitness and ultimately the viability of plant populations.
Figure 1-1: An example of a calcareous grassland patch in the Franconian Alb, Germany. Photo by M. DiLeo
Figure 1-2: *Pulsatilla vulgaris* Mill. flowers (left) and seed heads (right). Photos by M. DiLeo
Chapter 2

2 A Landscape Ecologist’s Agenda for Landscape Genetics

This invited review paper is currently in press as: DiLeo MF, and Wagner HH “A landscape ecologist’s agenda for landscape genetics”, Current Landscape Ecology Reports

Author contributions: HHW and MFD developed the concept for this paper after HHW was invited to submit a manuscript reviewing the last five years of landscape genetics. MFD conducted the literature review, analyzed the data, and wrote the manuscript. HHW helped to edit the manuscript and contributed ideas.

2.1 Abstract

This review examines the landscape genetic literature from 2011-2015 and summarizes the genetic evidence for the roles of habitat amount (i.e. total area of habitat), configuration (i.e. spatial arrangement of habitat), and matrix (i.e. nature of the intervening landscape between habitat patches) in shaping genetic differentiation and diversity of populations. We found that the vast majority of landscape genetic studies focused on the effects of habitat configuration and intervening matrix permeability on genetic differentiation of populations, and very few consider the consequences of habitat loss (i.e. change in habitat amount) versus fragmentation per se (i.e. change in habitat configuration). In addition, disproportionately few studies consider genetic diversity as a response variable in landscape genetic models. We argue that by ignoring the effects of habitat amount, landscape geneticists are missing an important component of how landscape structure shapes patterns of genetic variation. On the other hand, landscape ecologists may need to consider the confounding role of the matrix to resolve the ongoing debate about the relative importance of habitat loss versus fragmentation per se in determining biological diversity.
2.2 Introduction

Landscape genetics is an interdisciplinary field combining tools and concepts from both landscape ecology and population genetics to relate landscape structure to patterns of genetic variation (Manel et al. 2003, Holderegger and Wagner 2008). The field has evolved tremendously since Manel’s landmark paper in 2003 (Manel et al. 2003), moving from descriptive assignment tests used to define population boundaries to a more explicit analytical framework including landscape variables as predictors in genetic models (Storfer et al. 2007). A recent review categorized the types of questions and methods used in landscape genetic studies compared to papers published in its predecessor fields, and found that most self-identifying landscape genetic studies fall more into the realm of population genetics (e.g. using terms like "genetic", "gene", and "barrier"; Dyer 2015) than landscape ecology (e.g. using terms like “vegetation”, “resource”, “properties”; Dyer 2015). This begs the question: what can genetics contribute to the field of landscape ecology? Could this trend reflect a lack of initiative from the landscape ecology community to drive the agenda for landscape genetics?

As a tool, molecular genetics can make hard to observe processes visible and thus should be useful for landscape ecologists working on species who are cryptic or whose movement is hard to track. For example, genetic markers have been used to estimate both contemporary and historical effective population size (Chiucchi and Gibbs 2010), assess sex-biased dispersal (Goudet et al. 2002, Wang et al. 2012, Vangestel et al. 2013), identify population bottlenecks (Tucker et al. 2012), and characterize meta-population dynamics (Andreasen et al. 2012). Genetics can be used to quantify actual functional connectivity either directly (e.g. parentage analysis) or indirectly (e.g. estimates of genetic differentiation among populations), and thus provides the means to test hypotheses about how aspects of the intervening landscape matrix support or inhibit dispersal and gene flow (Wang et al. 2009, Lowe and Allendorf 2010b,
Most studies in landscape genetics focus on this aspect – i.e. testing if the landscape matrix (i.e. permeability of the intervening landscape between habitat patches) matters beyond the effects of the spatial configuration of populations in explaining among-population genetic differentiation. Such studies typically use link-level analysis (Fig. 2-1; Wagner and Fortin 2013), where each data point, or row in a data table, refers to a pair of sampling units and quantifies e.g. their pairwise genetic, geographic or ecological distance. Far fewer studies have tested the effects of habitat amount (i.e. total area of habitat) and habitat configuration (i.e. spatial arrangement of habitat) on genetic diversity within populations (Storfer et al. 2010). Such studies typically use node-level analysis (Fig. 2-1; Wagner and Fortin 2013), where each data point refers to a focal patch. Few papers combine node- and link-based analysis e.g. in gravity models. Note that another set of landscape genetic studies focuses on adaptive genetic variation (landscape genomics; Manel et al. 2010, Rellstab et al. 2015), which is beyond the scope of this review.

There is an ongoing and lively debate in ecology about the relative importance of habitat loss (i.e. change in habitat amount) versus fragmentation (i.e. change in habitat configuration independently of habitat amount) in shaping populations and communities. The two processes can be measured independently, where the same amount of habitat can be in the shape of one large patch (not fragmented) or several smaller patches (fragmented). Thus fragmentation reduces patch size and increases the proportion of edge to interior habitat. Studies in landscape ecology have provided strong theoretical and empirical support for the role of habitat amount in determining patch occupancy (e.g. Betts et al. 2007, Mortelliti et al. 2011, Scherer et al. 2012, Hornseth et al. 2014), species diversity (e.g. Fischer et al. 2005, Flick et al. 2012), and abundance (e.g. McGarigal and McComb 1995, Holbrook et al. 2000), with habitat fragmentation taking a back seat to the overwhelming effects of habitat loss (Fahrig 2003). These
finding have culminated into a provocative new hypothesis – Fahrig’s “habitat amount hypothesis”, which posits that the amount of habitat in a local landscape determines species richness and that metrics of configuration such as patch size (e.g. size of focal patch) and isolation (e.g. distance to neighbouring patches) can largely be ignored (Fahrig 2013). In contrast, existing evidence from metapopulation ecology indicates a strong role of habitat configuration in explaining species richness (e.g. Hanski et al. 2013, Rybicki and Hanski 2013) and population persistence (e.g. Hanski et al. 1995). In response to Fahrig, Hanski argues that the habitat amount hypothesis may only be valid at small spatial scales and when overall habitat amount is large (Hanski 2015). Landscape configuration should be important in real landscapes where the amount of remaining habitat is often quite low (Hanski 2015). Evidence from population and landscape genetics seems to support the metapopulation view that configuration and isolation (or its opposite, connectivity) are key determinants of population outcomes. For example, gene flow among populations decreases as a function of population isolation (i.e. isolation-by-distance; Wright 1943) and this relationship can be modified by the permeability of the intervening landscape matrix (i.e. isolation-by-resistance; McRae 2006). These patterns have been demonstrated in a variety of plant (e.g. Kamm et al. 2010, Dyer et al. 2012, Rico et al. 2014b) and animal (e.g. Row et al. 2010, Trumbo et al. 2013) systems. However, the amount of habitat at the local (within a buffer around each focal patch) or landscape scale (within the total study area) should also play an important role in shaping within-population genetic variation, as larger amounts of habitat can accommodate higher effective populations sizes (\(N_e\)), and consequently, populations will experience lower levels of genetic drift and retain higher levels of diversity (Wright 1931). Ultimately, it is the balance of genetic drift and gene flow that determines neutral genetic diversity within- and genetic differentiation among- populations.
(Wright 1931), and only by considering the two together can we fully appreciate the role of landscape structure on genetic variation.

Habitat loss and fragmentation tend to be highly correlated in nature, and as a consequence, their independent effects are often difficult to tease apart. Beyond this analytical challenge, there are inconsistencies in terminology and metrics used to measure fragmentation across disciplines (Lindenmayer and Fischer 2007). For example, metapopulation studies tend to operate at the patch scale, measuring patch size and isolation individually for each focal patch. However, in landscape ecology, patch size is measured at the landscape scale (e.g. mean patch size in landscape) and is not considered a measure of habitat amount, but as a metric of configuration (i.e., fragmentation). At the patch scale, patch size is an ambiguous metric of fragmentation as it does not account for local landscape context (Fahrig 2003). Fragmentation occurs at the landscape level and thus can only be measured \textit{per se} after accounting for effects of habitat amount. This mismatch in approach and terminology across disciplines has led to a large number of publications that measure the effects of certain aspects of fragmentation on a variety of population and genetic processes (e.g. Keyghobadi 2007, Aguilar et al. 2008), but very few actually quantify fragmentation \textit{per se} (i.e. control for habitat amount Fahrig 2003, Hadley and Betts 2012).

Here we examine the last five years of the landscape genetic and relevant population genetic literature and summarize the genetic evidence for the roles of habitat amount, configuration, and matrix resistance in shaping connectivity and diversity of populations. We identify key gaps in our current knowledge and ask: (1) are landscape geneticists considering the independent effects of habitat amount versus configuration on genetic variation? And (2), are the effects of landscape structure on genetic diversity versus genetic differentiation equally
considered by landscape geneticists? We end with a discussion of how researchers can address the identified gaps to help resolve a key debate in ecology.

2.3 Literature Review

We conducted a final literature search on 26 October 2015 of articles published between 2011-2015 in the ISI Web of Science Core Collection and BIOSIS Citation Index using the following search terms in the TOPIC field: landscape AND genet* AND (frag* OR “habitat loss” OR connectivity OR “gene flow” OR isolation). We refined results to include only the following research areas: ecology, genetics and heredity, evolutionary biology, biodiversity, conservation, plant science, zoology, environmental sciences, multidisciplinary sciences, biology, and ornithology. This search returned 1346 papers. As a first pass, we looked at titles and abstracts to remove papers that did not incorporate genetic data, were reviews or opinions, or were mainly methodological in nature and did not use a new empirical dataset. We also removed papers that did not explicitly test the influence of landscape structure (habitat amount, configuration, or matrix, as described below and in Fig. 2-1) on genetic variation (i.e. did not include them as predictors). This resulted in the exclusion of papers that investigated the genetic consequences of certain aspects of fragmentation (e.g. studies that compare genetic diversity between a reference and fragmented landscape) but these types of studies have been reviewed elsewhere (e.g. Keyghobadi 2007, Aguilar et al. 2008). Studies comparing genetic variation amongst islands were only included if habitat amount was defined and included as a predictor in genetic models (i.e. not just island size as predictor), and we excluded studies conducted in marine systems or seascapes. We placed the remaining 541 articles in one of three categories based on the landscape process investigated and whether the genetic response was measured at the link- (i.e. pairwise genetic differences among populations or individuals; Fig. 2-1) or node- (i.e. genetic variation measured per population; Fig. 2-1) level: 1) influence of habitat configuration on
genetic differentiation measured at the link level (e.g. isolation-by-distance, IBD), 2) influence of matrix permeability on genetic differentiation measured at the link level (e.g. isolation-by-resistance, IBR), or 3) influence of habitat amount and/or configuration on within-population genetic variation (e.g. genetic diversity or relatedness) or genetic differentiation (e.g. population-specific $F_{st}$), measured at the node-level. Many papers considered two or more of the above processes, and we categorized them hierarchically: papers that tested IBR in addition to IBD were placed in the second category, and papers that considered IBD or IBR and the influence of landscape structure on genetic variation measured at the node level were placed in the third category. For each paper that fell into the third category, we recorded the genetic response variable and information about the landscape predictors measured. Each landscape predictor was categorized as one of the following: 1) habitat amount, measured at the landscape or local-landscape scale (i.e. percent habitat in buffered area around focal patches or populations), 2) patch size measured at the patch scale (i.e. size of the focal patch), or 3) habitat configuration. Metrics of habitat configuration included patch-level measures of isolation (e.g. nearest neighbour distances, connectivity index), patch number and density metrics measured at the landscape scale (e.g. mean patch size, number of patches), and landscape-scale aggregation indices (e.g. clumping/dispersion of patches). We considered a study to have controlled for the independent effects of habitat loss versus fragmentation per se (Fahrig 2003) if they included habitat amount and at least one (patch- or landscape-level) metric of habitat configuration as predictors in genetic models, or used an experimental approach. We did not consider patch size measured at the patch scale to be a metric of habitat configuration as it lacks local landscape context, and is often used instead as a proxy of population size in the metapopulation literature. We further summarized results of the retained node-level studies by recording the direction and statistical significance of the effects of habitat amount, patch size, and configuration on genetic
variation. We did not carry out formal meta-analysis because few studies reported sufficient information.

2.4 Characterizing Gaps in the Literature

Out of the 541 retained studies, 485 were conducted at the link level (IBD: n=298; IBR: n=187) and 56 at the node level (Fig. 2-2a). We found similar biases in study taxa as previous reviews (Fig. 2-2a; Storfer et al. 2010). Given the over-representation of link-level studies, it was thus not surprising to find that the vast majority of retained studies used genetic differentiation as the response variable in genetic models, and that the most common landscape predictors were habitat configuration and matrix permeability (Fig. 2-2b). In contrast, only 23 studies – all of which were conducted at the node level - included habitat amount as a predictor, with either genetic differentiation (n=5) or genetic diversity (n=18), as the response variable. Only 11 of these controlled for the confounding effects of habitat amount versus fragmentation (Table 2-1). We have thus uncovered two key gaps in our knowledge of the effects of landscape structure on genetic variation: 1) most studies in landscape genetics are conducted at the link level and thus only consider effects of configuration and/or matrix on genetic variation, ignoring possible effects of habitat amount; and 2) genetic diversity is severely underrepresented as a response variable in landscape genetic models.

2.5 Habitat Amount Versus Fragmentation

Current evidence from landscape genetics would suggest a strong and almost singular effect of landscape configuration on functional connectivity and genetic structure across a variety of taxa. This goes against considerable evidence from landscape ecology, which suggests that habitat amount is the most important determinant of demographic outcomes (e.g. patch occupancy, abundance, species richness Fahrig 2003) and that configuration only becomes important below
a critical threshold of habitat loss (Fahrig 1998). However, our search of the recent literature suggests that this discordance among disciplines likely reflects a bias in the type of questions being asked and approaches used in landscape genetics compared to landscape ecology. For example, the vast majority of the studies that we identified were conducted at the link level – i.e. they quantified the relationship between pairwise genetic differentiation among populations and their spatial configuration (IBD; Fig. 2-2a) or aspects of the intervening landscape matrix (IBR; Fig. 2-2a). A much smaller proportion incorporated approaches at the node level, where aspects of the local landscape around focal patches were related to genetic variation. It is in this second scenario where the best opportunity exists to quantify the effects of habitat amount on genetic variation. This can also be accomplished at the link level, but it requires sampling of multiple landscapes, which is often not feasible. A third approach, using network-based gravity models, has the potential to integrate both node- and link-based data, but to our knowledge has not been used to include the effects of local habitat amount in genetic models. This approach will be discussed in a later section of this review.

2.6 Landscape Effects on Gene Flow and Genetic Drift

To understand the importance of including habitat amount in genetic models, we must consider the landscape determinants of gene flow and drift (Fig. 2-3). The balance between these two opposing forces ultimately defines both within- and among-population genetic variation (Wright 1948). When the relative strength of drift is higher than gene flow, populations are expected to exhibit low genetic diversity and high among-population differentiation. When the strength of gene flow outweighs that of drift, populations are expected to exhibit the opposite pattern, where genetic diversity within population is high and differentiation among them is low (Wright 1948). Whereas gene flow is the product of migration and thus should be influenced heavily by habitat configuration and matrix permeability, the strength of drift is determined by the effective size of
populations ($N_e$), and thus is directly influenced by habitat amount (Fig. 2-3). A strong link between genetic variation and $N_e$ has been empirically demonstrated in a large number of species (Ellstrand and Elam 1993, Frankham 1996, Leimu et al. 2006), yet the potential landscape determinants of $N_e$ and thus genetic drift are largely ignored in landscape genetics in favour of landscape hypotheses relating to gene flow. This is despite the fact that measures of both genetic diversity and differentiation will contain signals of both processes. Including $N_e$ as a covariate can potentially increase the strength of correlations between landscape structure and both genetic diversity (Mendez et al. 2014, Carvalho et al. 2015), and genetic differentiation (Weckworth et al. 2013, Prunier et al. 2015). Including $N_e$ is expected to be most important for species with low dispersal ability, but should not be ignored for vagile organisms. For example, Prunier et al. (2015) showed through simulations that including a proxy of $N_e$ explained up to 50% of the variation in pairwise population genetic differentiation when migration rates were low, and still explained up to 20% of the variation for high simulated migration rates.

Although a variety of methods and programs exist to estimate $N_e$ from genetic markers, effective populations sizes are difficult to quantify for real landscapes because these estimates can be biased when populations are spatially structured (Neel et al. 2013, Ryman et al. 2014). A landscape proxy for $N_e$ may thus be preferred. But what is the appropriate metric? Patch size measured at the patch-scale is often used as a proxy of population size in metapopulation models, but multiple meta-analyses have revealed highly inconsistent relationships between patch size and abundance across species (Bowers and Matter 1997, Thornton et al. 2011). Measured at the patch scale, patch size lacks local landscape context and may only be an appropriate measure of habitat amount and thus population size when patches are quite isolated (Fahrig 2003). In our search of the recent literature, only 11 out of the 37 models that included patch size found it to be a significant predictor of genetic variation, and in two instances the results showed the opposite
of the expected relationship (Table 2-2). Alternatively, habitat amount, measured at the local landscape scale (i.e. in buffered area around focal patches) may be a more appropriate proxy for population size (Fahrig 2013). However, of the 34 instances where local-landscape habitat amount was included as a predictor in genetic models, just half (n=17) found the predicted relationship, while fifteen found no significant effect and two found an opposite effect (Table 2-2). The small number of studies retained in our literature search precluded an in-depth meta-analysis, but there are a number of potential reasons for these inconsistencies. First, the relevant scale for measuring a landscape proxy of \( N_e \) is likely species-specific and will depend on how species perceive patch boundaries (Bender et al. 1998, Betts et al. 2014). Second, \( N_e \) is not only determined by habitat amount, but is also indirectly impacted by habitat configuration and matrix permeability by allowing migration among nearby patches (Fig. 2-3). We expect these effects to be most important when local habitat amount is small, where small patches will only be able to maintain a viable population if they are connected via gene flow (i.e. genetic rescue). Third, in addition to its impact on genetic drift, habitat amount can have contrasting effects on gene flow, which may obscure predicted relationships. For example, Robinson et al. (2012) and Dharmarajan et al. (2014) found that high local habitat availability was associated with higher levels of average relatedness among individuals within populations of white-tailed deer (\textit{Odocoileus virginianus}) and raccoons (\textit{Procyon lotor}), respectively. This suggests that these species exhibit higher natal philopatry when local resources and habitat are plentiful - a process that can lead to a pattern of low genetic diversity in populations surrounded by large amounts of habitat. Similarly, density-dependent migration can lead to the movement of individuals from large populations that have surpassed carrying capacity to smaller populations (Matthysen 2005). In this case accounting for the differences in \( N_e \) among pairwise populations may produce stronger models (Weckworth et al. 2013, Prunier et al. 2015).
Habitat quality - shaped by local environmental conditions, resource availability, and species interactions - can also limit the size of populations and impact rates of migration. It is increasingly recognized that these effects can shape genetic outcomes (e.g. Pitra et al. 2011, Wang 2012, Kahlilainen et al. 2014, Koen et al. 2016) and thus should be incorporated into landscape genetic models (see Pflueger and Balkenhol 2014 for a review on the subject). A binary definition of habitat and non-habitat is unlikely to capture the full range of these effects, and ideally, variation in habitat quality should be integrated directly into measures of habitat amount. For example, habitat suitability models are often used as a starting point for parameterizing resistance surfaces in landscape genetics, and these same models can be translated to measure the amount of high quality habitat in buffered areas around focal patches or populations.

2.7 Genetic Diversity versus Differentiation

Our literature search revealed another important gap; very few studies consider the effects of landscape structure on genetic diversity within populations, as opposed to genetic differentiation among them. Genetic diversity is a strong predictor of individual and population fitness, extinction risk, and the ability to respond to future environmental change, and thus is an important indicator for landscape-level conservation planning (Reed and Frankham 2003, Frankham 2005). Understanding the causes and consequences of patterns of genetic diversity has been a longstanding goal of biogeography, population and conservation genetics. Theoretical and empirical evidence from these fields have given us a strong understanding of the effects of population size and population isolation on genetic diversity. More recently, the field of landscape genomics has made great strides relating local environmental conditions to adaptive genetic diversity (reviewed in Manel et al. 2010, Rellstab et al. 2015). However, our knowledge
of the landscape determinants of neutral genetic diversity in complex landscapes is limited, and has largely been ignored in favour of studies on genetic differentiation in landscape genetics.

Recent simulations have demonstrated that within-population and among-population genetic variation can respond quite strongly to different aspects of landscape structure. For example, using individual-based simulations of gene flow for five hypothetical organisms that varied in dispersal ability, Jackson and Fahrig (2014) found that habitat amount explained 88% of variation in genetic diversity compared to 7% explained by fragmentation. Likewise, Ezard and Travis (2006) found that the fixation time of neutral alleles within simulated populations was predicted best by the amount of habitat. In contrast, simulations by both Bruggeman et al. (2010) and Cushman et al. (2012) demonstrated a stronger role of habitat configuration compared to habitat amount in explaining patterns of genetic differentiation among populations. These results appear to be reflected in a subset of the empirical studies that we retained in our literature search: Barr et al. (2015), Taylor and Hoffman (2014), and Capurucho et al. (2013; Table 2-1) found correlations between habitat configuration and genetic differentiation among populations, but habitat amount and not fragmentation correlated best with within-population genetic diversity. These results are again supported by population genetic theory. While habitat amount is expected to play a stronger role in determining population size and thus the local pool of alleles (i.e. genetic diversity; a measure of alpha diversity), the configuration of habitat determines rates of migration and thus leaves a stronger signal in measures of differentiation among populations (a measure of beta diversity; Fig. 2-3).

Genetic diversity and differentiation reach migration-drift equilibrium at different rates, and are thus expected to respond to landscape structure at different spatiotemporal scales (Varvio et al. 1986). When gene flow or population size is reduced by a change in landscape structure,
populations will become genetically differentiated from each other faster than genetic diversity is lost within populations. This difference in lag time is important to consider when quantifying the impact of landscape structure on genetic variation (Anderson et al. 2010). For example, measures of genetic differentiation might correlate with current landscape structure while genetic diversity might be better explained by a historical landscape (e.g. Keyghobadi et al. 2005, Flavenot et al. 2015).

Both genetic diversity and differentiation are the products of many generations of dispersal and mating. As the lag time in the ability to detect a signal of landscape change increases, the more important spatially distant individuals and their landscape context will be for explaining genetic variation. For example, using simulations, Jackson and Fahrig (2014) found that habitat amount explained genetic diversity best when it was measured at a broad spatial scale. In contrast, demographic outcomes (patch occupancy and abundance) correlated best with habitat amount measured at smaller spatial scales. Similarly, Millette and Keyghobadi (2015) found that fine-scale genetic structure of a pitcher-plant insect (Metriocnemus knabi) was best explained by broad-scale habitat configuration. These findings are significant from a conservation perspective as they may lead to different decisions about how to conserve the landscape depending on the outcome measured, further emphasizing the need for both demographic and genetic approaches.

If it takes longer for a signal of landscape change to emerge in measures of genetic diversity compared to differentiation, we might also expect the relevant spatial scale to be larger for within-population diversity. However, it appears that this is not always the case. For example, Taylor and Hoffman (2014) tested the influence of habitat amount and fragmentation on the white footed mouse (Peromyscus leucopus) at three spatial scales, and found that habitat amount
measured at the smallest radius (500 m) around focal populations was the single best predictor of genetic diversity. Balkenhol et al. (2013) found that landscape-level habitat amount influenced the contribution of local-landscape and patch-scale measures of connectivity to both genetic diversity and differentiation in the forest dwelling marsupial *Marmosops incanus*. Similarly, Millette and Keyghobadi (2015) found an interactive effect of habitat amount and patch connectivity; genetic differentiation of *M. knabi* populations increased with patch isolation when habitat amount was low, but decreased with patch isolation when habitat amount was high. Together these results suggest that both habitat amount and configuration can impact within- and among-population measures of genetic variation at multiple spatial and temporal scales and underlines the importance of multi-scale approaches in landscape genetics.

2.8 Importance of Matrix

The last decade of landscape genetics has produced overwhelming evidence that functional connectivity is affected by the permeability of the intervening landscape matrix beyond the effects of the spatial configuration of populations (e.g. Stevens et al. 2006, Perez-Espona et al. 2008, Lange et al. 2012), and that different species can respond to the same landscape in vastly different ways (Shanahan et al. 2011, Poelchau and Hamrick 2012, Richardson 2012, Amos et al. 2014, Engler et al. 2014, Paquette et al. 2014). Thus, using a binary classification of habitat versus non-habitat when quantifying landscape structure is unrealistic and can lead to misleading conclusions about the impact of habitat loss and fragmentation on populations (Prugh et al. 2008, Betts et al. 2014). However, our knowledge of how habitat amount, configuration, and the matrix interact to determine gene flow is quite limited, as it requires sampling across replicate landscapes, or the used of node-based or gravity model approaches (Fig. 2-1) which are quite rare in landscape genetics. The few studies that have considered multiple landscapes have found that different aspects of the landscape matrix predicted gene flow for the same species in
different regions (e.g. Moore et al. 2011, Trumbo et al. 2013). We know of only one study with sufficient replication to determine if these differences in landscape effects on gene flow could be explained by landscape-level differences in habitat amount and configuration: Short Bull et al. (2011) contrasted patterns of genetic differentiation in black bears (*Ursus americanus*) across 12 landscapes, and found that landscape features were only identified as significant predictors of population differentiation if they were highly variable or fragmented at the landscape scale. For example, forest (considered suitable habitat for black bears) was more likely to be identified as a predictor of genetic differentiation in landscapes where forest was more fragmented (Bull et al. 2011). This study did not control for differences in overall habitat amount among the landscapes, so it is unclear if fragmentation or if the amount of different features in each landscape was driving the effect. This suggests that broad-scale habitat amount and configuration can alter species response to the matrix or, alternatively, alters our ability to detect landscape genetic correlations. Simulations by Cushman et al. (2012) suggests that the reverse can also be true; the relative resistance of the intervening landscape matrix can modify the relative contribution of landscape-level habitat amount versus fragmentation to patterns of genetic differentiation. Although they identified habitat configuration as the best overall predictor of genetic differentiation in simulated landscapes, the amount of variance explained by habitat amount (alone, and interactively with configuration) increased with relative matrix resistance. At the highest level of matrix resistance simulated, habitat amount outperformed three of the four tested fragmentation metrics. This result suggests that landscape ecologists need to take into account the intervening matrix in order to resolve the ongoing debate on the relative importance of habitat loss vs. fragmentation. However, more evidence from empirical studies and further simulations are urgently needed.
2.9 Testing the Importance of Habitat Amount versus Fragmentation with Genetic Data

How can we test the relative importance of the effects of habitat amount versus fragmentation on genetic variation? We see three possible approaches: 1) Link-based analysis in multiple landscapes (Fig. 2-1): by quantifying gene flow across multiple landscapes that vary in overall habitat amount and degree of fragmentation, we can start to understand the interactive effects of habitat amount, configuration and matrix. Link-based analysis is by far the most common approach in landscape genetics, but usually only a single landscape is considered. We recognize that replication at the landscape scale is not always feasible, but as shown by Short Bull et al. (2011), it is desperately needed to build a comprehensive understanding of how landscape structure influences gene flow. 2) Node-based analysis (Fig. 2-1): habitat amount and fragmentation are measured either at the landscape-scale or locally within a given radius around focal patches, and are related to patch-specific measures of genetic variation. The genetic response variable can either be a measure of genetic diversity, or population-specific $F_{st}$ that represents the relative genetic differentiation of a local population from all others in a metapopulation (Gaggiotti and Foll 2010). This focal patch approach is commonly used in landscape ecology and can be adapted to directly test Fahrig’s habitat amount hypothesis (Fahrig 2013), but can also incorporate matrix effects by measuring connectivity of the focal patch to all others in the local landscape as a function of matrix permeability (e.g. using an incidence function; Hanski 1994); 3) Network-based gravity models (Fig. 2-1): this approach models genetic differentiation among populations as a function of both local-landscape variables that are measured at the node level, and matrix permeability measured at the link level. This method has been used to model the effects of local environment and habitat quality on gene flow (e.g.
Murphy et al. 2010, Dileo et al. 2014), but to the best of our knowledge has not yet been employed to quantify the relative contribution of local habitat amount, fragmentation and matrix.

Ideally, future research in landscape genetics should employ all three approaches, as each captures the potential effects of landscape structure on genetic variation at different spatio-temporal scales. Certain approaches will be more appropriate or feasible in particular systems or for particular species. For example, the use of the individual as the study unit is common in link-based studies for continuously distributed species (i.e. genetic differentiation is measured among individuals rather than among populations), however it is less clear if individual-based methods can be translated to node- or gravity-model frameworks. It has been suggested that landscape features could be measured around sampling locations of focal individuals at radii relevant to the species’ dispersal ability or home-range size (see Pflueger and Balkenhol 2014), however more theoretical work is needed to demonstrate the validity and strength of this approach.

Node-based analysis is most intuitively applicable when habitat is discrete. For more continuously distributed species, binary habitat classification may not be appropriate, and this may explain why we found an under-representation of node-based studies. In comparison, link-based analysis of either IBD or IBR does not necessarily require the quantification of landscape pattern with a strict definition of patch boundaries. For example, matrix permeability can be modeled as a continuous variable such as habitat suitability or outputs from species distribution models. Translating these continuous representations of landscapes to categorical ones that can be used to quantify the amount and configuration of habitat is possible, but not always straightforward (Betts et al. 2014, Lausch et al. 2015).
2.10 Conclusions

Here we have identified three key gaps in our knowledge of the effects of landscape structure on genetic variation. (1) Very few empirical studies have included habitat amount as a covariate in genetic models, and as a result, empirical evidence of the relative influence of habitat loss versus fragmentation on genetic diversity and differentiation is limited. (2) In addition, disproportionately few studies in landscape genetics consider genetic diversity as a response variable. Habitat amount can have important impacts on both genetic drift and gene flow at multiple spatial scales. Moreover, the effects of drift and gene flow are unequally represented in the signals of genetic diversity and differentiation and thus may correlate with different aspects of landscape structure. (3) The effects of habitat amount, configuration and matrix (and ideally also habitat quality) should be considered simultaneously to help resolve the debate about the relative importance of habitat loss versus fragmentation in determining biological diversity. This will require a more balanced analytical framework that incorporates node-based and landscape-level approaches commonly used in landscape ecology. We thus encourage landscape ecologists take a more active role in setting the agenda for landscape genetics.

2.11 Acknowledgements

We thank A. Hadley for helpful discussion during the preparation of this manuscript, and Matt Betts and two anonymous reviewers for their comments. This research was funded by the Natural Sciences and Engineering Research Council of Canada Discovery grant to HHW an Ontario Graduate Scholarship to MFD.
Table 2-1: Summary of results of 11 node-based studies that controlled for the independent effects of local habitat amount and fragmentation on genetic variation. The direction of the relationship of the predictor variables “habitat amount” and “habitat fragmentation” with the genetic response variable (genetic diversity, relatedness, or differentiation) is given. For all studies, the genetic response was measured per population and the predictors were measured in buffered regions around populations.

<table>
<thead>
<tr>
<th>Study</th>
<th>Taxon</th>
<th>Response Variable</th>
<th>Effect of Habitat Amount</th>
<th>Effect of Fragmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toma et al. 2015</td>
<td>plant</td>
<td>genetic diversity</td>
<td>positive</td>
<td>negative</td>
</tr>
<tr>
<td>Barr et al. 2015</td>
<td>bird</td>
<td>genetic diversity</td>
<td>positive</td>
<td>no effect</td>
</tr>
<tr>
<td>Flavenot et al. 2015</td>
<td>amphibian</td>
<td>genetic diversity</td>
<td>positive</td>
<td>no effect</td>
</tr>
<tr>
<td>Taylor and Hoffman 2014</td>
<td>mammal</td>
<td>genetic diversity</td>
<td>positive</td>
<td>no effect</td>
</tr>
<tr>
<td>Capurucho et al. 2013</td>
<td>bird</td>
<td>genetic diversity</td>
<td>positive</td>
<td>no effect</td>
</tr>
<tr>
<td>Levy et al. 2013</td>
<td>reptile</td>
<td>genetic diversity</td>
<td>no effect</td>
<td>positive</td>
</tr>
<tr>
<td>Robinson et al. 2012</td>
<td>mammal</td>
<td>average relatedness</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>Millette &amp; Keyghobadi 2015</td>
<td>insect</td>
<td>genetic differentiation</td>
<td>small spatial scale: no effect</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium scale: no effect</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>large scale: negative</td>
<td>negative&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Coster et al. 2015</td>
<td>amphibian</td>
<td>genetic differentiation</td>
<td>no effect</td>
<td>no effect</td>
</tr>
<tr>
<td>Peterman et al. 2015</td>
<td>amphibian</td>
<td>genetic differentiation</td>
<td>no effect</td>
<td>no effect</td>
</tr>
<tr>
<td>Hahn et al. 2013</td>
<td>plant</td>
<td>genetic differentiation</td>
<td>no effect</td>
<td>no effect</td>
</tr>
</tbody>
</table>

<sup>a</sup> a significant interaction between habitat amount and isolation was found: genetic differentiation increased with isolation when habitat amount was low, but decreased with isolation when habitat amount was high
Table 2-2: Summary of the effects of habitat amount, patch size, and configuration on genetic variability for 56 node-based studies identified in the literature from 2011-2015. An effect on genetic diversity was classified as ‘predicted’ if genetic diversity exhibited a significant positive relationship with habitat amount or patch size, or a negative relationship with fragmentation (configuration). An effect on relatedness or genetic differentiation was classified as ‘predicted’ if relatedness or differentiation exhibited a significant negative relationship with habitat amount or patch size, or a positive relationship with fragmentation. Note that many studies included multiple species or tested multiple predictors and are represented more than once.

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Habitat Amount</th>
<th></th>
<th>Patch Size</th>
<th></th>
<th>Configuration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted</td>
<td>No Effect</td>
<td>Opposite</td>
<td>Predicted</td>
<td>No Effect</td>
<td>Opposite</td>
</tr>
<tr>
<td>genetic diversity</td>
<td>15</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>relatedness</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>genetic differentiation</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>15</td>
<td>2</td>
<td>11</td>
<td>24</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 2-1: Approaches for quantifying the role of landscape structure on genetic variation. Letters represent focal patches or populations, and grey shapes represent habitat in a matrix of non-habitat (white background). In link-based approaches and gravity models, the genetic response (genetic differentiation) is measured in a pairwise fashion between focal patches/populations. In node-based approaches, the genetic response is measured per focal patch/population. Link-based approaches model genetic differentiation as a function of the distance (e.g. isolation-by-distance, IBD) or matrix permeability (e.g. isolation-by-resistance, IBR) between populations or individuals. For node-based approaches, genetic diversity or population-specific genetic differentiation is modeled as a function of landscape structure measured in buffered areas around focal patches/populations (dashed circles). Gravity models combine the two approaches, where genetic differentiation is modeled as a function of landscape predictors measured at both nodes (e.g. in buffered areas around each patch/population) and links (e.g. pairwise distance or matrix resistance). Note: incorporating the effects of habitat amount on genetic differentiation in link-based approaches is possible, but requires sampling of multiple landscapes.
Figure 2-2: Mosaic plot showing the proportion of studies identified from the literature between 2011-2015 that were conducted at the link and node level across taxa (a), and barplots showing the number times configuration only, matrix, or habitat amount were used as predictors in genetic models with either genetic differentiation or diversity as the response variable (b). Link-based studies were categorized as either testing hypotheses related to isolation-by-distance (IBD) or isolation-by-resistance (IBR). One link-based IBR study included three separate taxa, and one node-based study included two taxa and are therefore represented more than once in the mosaic plot. Non-insect invertebrates were included in the insect category and studies on fungus were included in the plant category.
Figure 2-3: Predicted relationships between landscape structure, genetic drift and gene flow, and genetic diversity and differentiation. Habitat amount and quality are expected to strongly determine local effective population size ($N_e$) and thus the local pool of alleles (i.e. genetic diversity), whereas habitat configuration and matrix will affect the probability of migration and gene flow among populations and thus leave a stronger signal in measures of genetic differentiation. Habitat configuration and matrix can indirectly influence local $N_e$ by promoting migration among nearby populations. Habitat amount and quality can directly impact migration by promoting philopatry when local resources are high or indirectly through density-dependent processes (e.g. emigration when a population approaches carrying capacity).
Chapter 3

3 The Gravity of Pollination: Integrating At-Site Features into Spatial Analysis of Contemporary Pollen Movement


Author contributions: This paper is the result of group project initiated during a Landscape Genetics Distributed Graduate Seminar. RJD conceived the study, provided the empirical data, and wrote the simulation code. MFD and RJD analyzed the data. All authors contributed to the design of the study and writing of the manuscript. MFD took the lead in coordinating the writing of the manuscript, integrating parts from co-authors, and revising the manuscript for publication.

3.1 Abstract

Pollen-mediated gene flow is a major driver of spatial genetic structure in plant populations. Both individual plant characteristics and site-specific features of the landscape can modify the perceived attractiveness of plants to their pollinators and thus play an important role in shaping spatial genetic variation. Most studies of landscape-level genetic connectivity in plants have focused on the effects of inter-individual distance using spatial and increasingly ecological separation; yet have not incorporated individual plant characteristics or other at-site ecological variables. Using spatially explicit simulations, we first tested the extent to which the inclusion of at-site variables influencing local pollination success improved the statistical characterization of genetic connectivity based on examination of pollen pool genetic structure. The addition of at-site characteristics provided better models than those that only considered inter-individual spatial distance (e.g., IBD). Models parameterized using conditional genetic covariance (e.g., Population Graphs) also outperformed those assuming panmixia. In a natural population of
Cornus florida L. (Cornaceae), we showed that the addition of at-site characteristics (clumping of primary canopy opening above each maternal tree and maternal tree floral output) provided significantly better models describing gene flow than models including only between-site spatial (IBD) and ecological (Isolation By Resistance) variables. Overall, our results show that including inter-individual and local ecological variation greatly aids in characterizing landscape-level measures of contemporary gene flow.

3.2 Introduction

Pollen is the most pervasive vector of gene exchange for most plant species (Ennos 1994, Moran and Clark 2011). As such, it has the potential to exert the largest influence on within and among population genetic structure (Sork and Smouse 2006). Parameterization of gene flow rates have most commonly been measured using only spatial context under models of isolation by distance (IBD; Wright 1943), though more recently models have begun to include characteristics of heterogeneity in the intervening landscape (Isolation By Resistance, IBR; McRae 2006) and measures of environmental dissimilarity among populations (Isolation By Environment, IBE; Wang et al. 2013). These more integrated approaches to quantifying contemporary gene flow provide additional insights into factors influencing spatial dynamics in contemporary gene flow (Sork et al. 2005, Dyer et al. 2010).

Physical and ecological separation are not the only factors influencing genetic connectivity; there is an equally robust literature describing differential pollination success driven by inter-individual heterogeneity in individual-level reproductive traits. Variation in total floral output as well as specific floral traits such as size, color, scent, and nectar reward often increase pollinator attraction, and their effects on plant reproductive success have been well studied (e.g., Schemske and Bradshaw 1999, Ishii et al. 2008, Mitchell et al. 2009, Streinzer et al. 2009, Macukanovic-Jocic et al. 2011, Riffell 2011, Ye et al. 2011, Malerba and Nattero
What is currently missing from models evaluating landscape-level patterns of gene flow are methodologies that integrate the effects of among-individual landscape separation with features shown to be influential to pollination success at the local scale.

There are several ways in which localized demographics may influence patterns of genetic connectivity in such a way as to modify overall population genetic structure. In sparse populations, pollinators tend to visit more flowers on each plant resulting in reduced inter-individual pollination (Ghazoul 2005). Conversely, patches with high floral density tend to attract more pollinators and have more inter-individual pollination (Totland and Matthews 1998). The net effect of density on population genetic structure may be two-fold. First, pollinator foraging distance is inversely related to plant density, resulting in quantitative changes in fine-scale population genetic structure (Nattero et al. 2011). At larger spatial scales, genetic structure becomes more granular due to spatially restricted mating and concomitant reductions in genetic effective neighborhood size. Second, genetic diversity may also be indirectly influenced by changes in inbreeding. In plants with mixed mating systems, geitonogamy increases selfing and reduces heterozygosity (De Jong et al. 1992, Barrett et al. 1994). Even for species with obligate outcross mating systems, inbreeding may increase if mating is among spatially proximate individuals and adult genotypes are spatially autocorrelated due to spatial limitations in seed dispersal (e.g., Dyer et al. 2012).

The location of individuals on the landscape and their local ecological conditions may also impact genetic connectivity. Dyer & Sork (2001) showed that the density of *Quercus* and *Carya* canopy trees within a stand influenced the genetic diversity of *Pinus* seedlings. Across a gradient of timber management techniques based on removal of primary canopy trees, Sork *et al.* (2005) also showed significant reductions in pollen pool genetic structure for the understory tree
Cornus florida L. More recently, Dyer et al. (2012) showed that the genetic structure of pollen pools among Cornus florida individuals located within primary canopy openings had elevated biparental inbreeding and correlated paternity. They hypothesized that understory trees within the openings were preferentially pollinated by nearest neighbor mating. While these analyses showed that local conspecific and heterospecific variables measured at the location of the individual influenced future population genetic structure, their impact on describing landscape-level patterns of contemporary gene flow in relation to between-individual spatial and ecological isolation has yet to be determined.

In this study we examine the extent to which local density, ecological variables, and individual reproductive characteristics aid in describing the spatial pattern of genetic connectivity. Our interest here is to determine if landscape-level analysis of pollen-mediated gene flow, which has traditionally relied on IBD, IBR, and IBE processes, is improved with the addition of localized variability. We adopted a regression approach that included predictors based on at-site and between-site variables. Regression models that have parameters representing both attractive (at-site) and repulsive (between-site) predictors have been used in many contexts, including analyses of landscape genetic connectivity (Murphy et al. 2010), and are analogous to classical gravity models (e.g., Fotheringham and O'Kelly 1989). We use reverse-time stochastic coalescent simulations to first determine the magnitude of variation in both local density and at-site heterogeneity that are necessary to be statistically detected by means of pollen pool genetic analysis. We then apply this regression approach to an experimental population of Cornus florida L. to evaluate both model efficacy and the relative contributions of at-site and between-site features for describing contemporary gene flow. We close by discussing the general utility of using gravity models for landscape genetic analyses of pollen connectivity.
3.3 Methods

3.3.1 Gravity Models

The regression approach we adopted has parameters for two kinds of predictor variables, those that will increase genetic covariance among sampled pollen pools and those that are expected to decrease it. Predictors of the first kind include reproductive characteristics such as floral output and plant size as well as at-site ecological features such as canopy opening and local density, all of which are measured at the location of the individual plant. Variables that act to decrease pollen pool covariance include Euclidean distance between individuals (IBD) and ecological features that may differentially modify pollinator movement (IBR) across the landscape. We assumed that this is a mixed-effects model as:

\[ Y_{ij} = k d_{ij}^\alpha m_j^\beta e_{ij}^\gamma \]  

explicitly following the form presented in Anderson (1979) and Murphy et al. (2010) for continuity. Our response variable, \( Y_{ij} \), is defined by Bray-Curtis (shared allele) distance (Bowcock et al. 1994). This distance is estimated from sampled pollen pools and consists of the set of paternal pollen haplotypes sampled by each plant. In cases where the maternal individual had the same heterozygous genotype as the offspring, making the paternal contribution ambiguous, we followed Smouse et al. (2001) to estimate pollen donor haplotypes probabilistically.

The three categories of predictors in [1] included spatial distance between maternal trees \( (d_{ij}) \) to account for IBD, features measured at the location of the maternal individual \( (m_j) \) that may aid in attracting pollinators, and ecological features of the landscape measured between maternal individuals that may act to modify genetic connectivity \( (e_{ij}) \) under the assumptions of IBR. The regression models were built according to hypothesized connectivity networks where pollen pools of georeferenced maternal individuals represented network nodes, and at-site and
between-site variables were measured at nodes and along edges, respectively. We used simulation data (see below) to select the connectivity network providing the greatest statistical resolution. All model parameters were estimated using the nlme package (Pinheiro et al. 2012) in R.

3.3.2 Simulation Models
We developed an individual-based stochastic, reverse-time coalescent simulation model to create pollination pedigrees for a set of maternal individuals (C code available from RJD at http://dyerlab.bio.vcu.edu). Pollinator movement was based on random walks on a fixed grid system. Twenty offspring were simulated for each of twenty maternal individuals. For each offspring we simulated a pollination path across the landscape to identify the putative father from which we extracted multilocus pollen haplotypes at ten codominant loci. For simplicity we assumed that all individuals were self-incompatible, there was no limitation in pollen production (e.g., a father could sire many offspring), and all plants had unique spatial coordinates. In addition to paternal haplotypes, we also retained maternal and paternal location on the landscape for estimates of isolation by distance (IBD) and the value assigned for relative maternal tree detectability ($H$, a surrogate for differential at-site features, see below). All simulations were repeated 50 times with each set of parameter combinations to provide confidence intervals.

Pollen flow was simulated under alternate pollen donor densities, ranging from 0.1% to 10% site occupancy, to explore the influence local donor density may have on pollen pool structure. For comparison to existing empirical studies, we calculated multilocus genetic structure using TwoGener analysis ($\Phi FT$; Smouse et al. 2001, Dyer and Nason 2004), a robust measurement of among-pollen pool genetic covariance commonly used in analyses of contemporary gene flow. All genetic data was imported into R 3.1 (R Development Core Team, 2010) and analyzed using the packages gstudio (Dyer 2014) and pegas (Paradis 2010). After
corrections for non-normality, we assessed differences in estimates of the $\Phi_{FT}$ parameter across pollen donor densities with an ANOVA and examined significant differences using a TukeyHSD test.

A gravity model was constructed for each level of pollen donor density. The specification of the underlying connectivity network is critical because it determines the set of pairwise measurements used to estimate the gravity model. Two connectivity networks were examined. We first fit a full connectivity network (hereafter referred to as the saturated network) where all pair-wise connections were used. This saturated connectivity network is the null case that is commonly used for Mantel Tests and evaluations of IBD. We also estimated models whose connectivity network was based on Pollination Graphs, an extension of Population Graphs (Dyer and Nason 2004) that uses pollen pool haplotypes to estimate a conditional genetic covariance network (Dyer et al. 2012). Pollination Graphs have fewer edges than saturated networks, but those edges describe the spatial shape of multilocus genetic covariance on the landscape and have been shown to significantly increase statistical power in reconstructing connectivity (e.g., Dyer et al. 2010). We compared gravity models from alternate connectivity networks as follows. First, we compared the model fit for saturated vs. Pollination Graph (hereafter pruned) networks. We assumed that the pruned network (based on conditional genetic covariance) would provide a better fit than the saturated network (all pairs of individuals connected), as the latter would be over-parameterized. For direct comparison of models we used $R^2$, as a model fit statistic such as AIC would be inappropriate because saturated and pruned models have different number of observations, not different sets of predictors. We used the more robust connectivity network for all subsequent analyses. Second, we estimated the effect size that heterogeneity in at-site characters requires before it is statistically detectable by looking at genetic variation in the sampled pollen pools. We simulated heterogeneity in at-site features by
modifying the distance at which pollinators moving across the landscape will be able to detect the presence of a flowering individual. We parameterized this detectability as $H$ and assumed it would operate as a surrogate for variation in floral output or size of individual floral display. The null set of simulations imposed homogeneity in detectability whereby everyone had equal detectability. Variability in $H$ was introduced by adding variation to individual detectability at 2X and 5X greater than the null simulations. The effects of local density and detectability were evaluated using a factorial ANOVA comparing high vs. low density (using the 10% and 0.1% occupancy data) as one factor and the three levels of detectability as the other factor. All data were transformed when necessary.

3.3.3 Empirical Gravity Model with *Cornus florida*

The flowering dogwood, *Cornus florida* L. (Cornaceae), is commonly found in eastern deciduous forests of North America. Blooming in the early spring, this is a predominantly outcrossing species, with inflorescences consisting of 15-35 perfect flowers surrounded by four large white bracts (Dyer et al. 2012). Dogwoods in the piedmont region of Virginia are pollinated by a variety of bees from the Halictidae and Andrenidae families, as well as several species of small beetles and flies (Mayor et al. 1999, Carr 2010).

The Virginia Commonwealth University maintains the Inger and Walter Rice Center for Environmental Life Sciences in Charles City County, Virginia (latitude 37.33, longitude -77.20). This field research station along the north bank of the James River comprises 494 acres of mixed upland deciduous and pine forest and tidal wetland. Dominant species of the overstory canopy include *Quercus alba, Quercus rubra,* and *Pinus taeda,* while the understory is mainly composed of *C. florida, Ilex opaca,* and *Aralia spinosa.* Dyer *et al.* (2012) identified and georeferenced 452 adult flowering dogwood trees at the site and leaves were collected for genetic analyses. Of these, an average of 19 seeds were collected from each of 17 maternal individuals for genetic
analysis of pollen pools. DNA was extracted using DNEasy 96 Plant DNA kits (Qiagen, Germantown, MD), and individuals were genotyped at five polymorphic microsatellite loci. Molecular methods are detailed in Dyer et al. (2012).

Our empirical gravity models incorporated four at-site variables \( m \); Eqn 1, Table 3-1). Diameter at breast height \( (dbh) \) and total floral output \( (flor) \) were measured by hand from each maternal tree. Canopy openness above each tree was ascertained using hemispheric photography and qualified as both percent open canopy \( (pctsky) \) and evenness in the distribution of open canopy \( (clump) \). Previous work on \( C. florida \) showed that genetic structure of pollen pools can be significantly influenced by heterospecific canopy structure (see Sork et al. 2005, Dyer et al. 2012). We also measured a number of between-site variables \( e \); Eqn 1) to account for isolation by resistance (IBR). These variables were obtained from the Rice Center data repository and consisted of hyperspectral imagery classified as: open primary canopy due to fields \( (open) \), open primary canopy resulting from roads \( (roads) \), \( C. florida \) secondary canopy \( (cornus) \); derived from direct canopy measurement for all individuals under the primary canopy), pine primary canopy \( (pine) \), and deciduous primary canopy \( (decid) \). Between site ecological distances were estimated following gravity model specifications in Murphy et al. (2010) whereby we quantified the mean and variance in each feature measured along a straight-line transect between each pair of maternal individuals (Table 3-1). We explored the effect of increasing the width of between-site transects by calculating between-site variables at bandwidths of 1 (the minimum resolution of the data), 4, 6, 10, and 20 meters. Increasing bandwidth did not improve model fit (A1, Appendix), so all analyses were undertaken at the minimum resolution of 1 m. Spatial distance between maternal individuals was also included as the distance term \( d \); Eqn 1) to account for effects of isolation by distance (IBD). We tested all predictor variables for collinearity prior to inclusion in gravity models. Across transects, the largest correlation between alternative predictors was
$r^2 = 0.403$ for the estimation of the variance in road and pines, which we considered too small to cause collinearity problems in estimating regression model parameters. Models were constructed using the connectivity network found to be most informative in the simulation and built using a stepwise variable approach (both backward and forward). Model support was assessed using AIC and all models whose $\Delta$AIC < 2 were considered to be equally supported.

3.4 Results

3.4.1 Simulation Results

The simulations showed that pollen donor density and tree detectability both influenced the statistical characterization of pollen-mediated genetic connectivity. Pollen pool differentiation was asymptotically related to local density with differentiation among sampled pollen pools plateauing at high donor density (ANOVA; df=18,931; F=40.51; $p < 0.001$). High-density simulations ($\Phi_{FT}; \text{dad}=10\% = 0.138$) had significantly more spatial structure in sampled pollen pool genotypes than low-density populations ($\Phi_{FT}; \text{dad}=0.1\% = 0.098$; Fig. 3-1). A series of TukeyHSD post-hoc tests supported the asymptotic relationship between differentiation and density (results not shown).

As anticipated, the underlying connectivity network had a large effect on the statistical resolution for gravity models to describe genetic connectivity. Models constructed from the Pollination Graph networks explained more of the available variance ($R^2$) in the simulation data than gravity models based on fully saturated networks (A2, Appendix). When considered simultaneously with pollen donor density, a factorial ANOVA showed both effects (pruning & density) to significantly influence model fit (density, $F_{1,1096} = 58.38$, $p < 0.001$; pruning, $F_{1,1096} = 2100.84$, $p < 0.001$), with no interaction between the main effects ($F_{1,1096} = 1.01$, $p = 0.31$). As connectivity based on Pollination Graph provided better explanation of genetic variation, we used these connectivity networks for the rest of our analyses.
Simulations that included heterogeneity in individual detectability, $H$, a surrogate for at-site factors that may contribute to differential pollination success, provided better inference (lower AIC and higher summed Akaike weights) than models including only spatial distance between individuals ($d$; Fig. 3-2). The relative information content ($\partial$AIC) was influenced by both donor density (high vs low; $F_{1,296} = 51.90, P < 0.001$) and relative ‘discoverability’ of maternal individuals ($H$, $F_{1,296} = 43.50, P < 0.001$), with no indication of interaction ($F_{1,296} = 1.94, P = 0.16$). Perhaps more biologically relevant for future work, at-site features such as those summarized by $H$ were more important (as measured by summed Akaike weights) than even spatial distance (IBD) in these simulation results (Fig. 3-2b).

### 3.4.2 Empirical Gravity Model with *Cornus florida*

The best-fit model (Table 3-2) included variables measured at both the maternal individual level ($v$: clump, flor) and the between-site level ($c$: open.var, decid.mn). Inter-maternal spatial distance, $d$ (IBD), did not significantly add to the description of pollen pool genetic structure (Table 3-2) after inclusion of local and intervening ecological variables. Two equally likely models ($\partial$AIC < 2) were also identified, both of which contained both at-site and between-site ecological variables. For the best fit model, the signs on the gravity model regression coefficients suggested that clumping of canopy openings above the target tree is associated with increases in genetic distance whereas floral output, variance in open canopy between trees, and average density of deciduous (non-*Quercus*) canopy are associated with decreases in genetic distance.

### 3.5 Discussion

We showed that accounting for heterogeneity in individual maternal traits resulted in a better overall characterization of landscape level pollen-mediated gene flow than models only
including parameters for spatial (IBD) and ecological (IBR) separation. Adopting a regression approach configured within a gravity model framework provided an effective and flexible method for identifying which sets of localized and intervening variables are most important in describing contemporary genetic connectivity. Perhaps most surprisingly, our results show that at-site variability and intervening ecological heterogeneity are more statistically important in describing landscape-level connectivity than physical separation.

3.5.1 Connectivity Networks Using Pollination Graphs
Connectivity networks conditioned on genetic covariance provided significantly better-fit models than saturated connectivity networks (A2, Appendix). Gene flow often occurs at spatially restricted scales where proximate neighbors have a higher probability of mating than more distant pairs. Building a connectivity network using conditional genetic covariance is one approach for defining *model free* connectivity. We showed that this method outperformed saturated connectivity networks that were over parameterized and had a detrimental influence on the error variance (Jaquiery et al. 2011), reducing overall model performance. Other pruning methods have been suggested; including defining connections based on hypothesized dispersal distance (e.g., expert opinion; Murphy et al. 2010), arbitrarily specified limits to genetic distance (Rozenfeld et al. 2008), or algorithmic methods not based on biological inferences (e.g., Urban and Keitt 2001, Naujokaitis-Lewis et al. 2013). However, dispersal capacity is often unknown and is one of the underlying parameters being investigated. Using information about dispersal capacity to construct a gravity network that estimates connectivity may be tautological. In contrast, pruning by conditional genetic covariance provides an objective and unsupervised statistical approach to pruning connectivity networks, which in these studies led to stronger inference in both simulated and empirical data.
3.5.2 Effects of At-Site Features

When pollen donors are less dense on the landscape, pollinators must travel farther when foraging, reducing overall pollen pool genetic structure (Fig. 3-1). In high-density populations, overall population-level differentiation is increased, making the spatial granularity of structure much more fine-scaled. There are two interesting consequences of these results. First, our results seem to contradict Austerlitz and Smouse (2001) who suggested that local density can be largely ignored as long as one quantifies average dispersal and inter-mother distances. Variance decomposition based on estimations of continuous and anisotropic dispersal functions serve to simplify the derivations of idealized parameters such as $\Phi_{GT}$. However, they do so at the expense of actually understanding how localized heterogeneity influences connectivity, a goal of landscape genetic studies and a critical feature to consider when examining contemporary gene flow. Second, and perhaps more importantly, our simulations suggested that the dynamics in low density populations (forest trees), may be governed by a different set of processes than those that occur at higher densities (herbs and smaller plants). While both existing theoretical (Levin and Kerster 1969) and empirical studies (Dick et al. 2003, Gonzales et al. 2006) support our findings, subsequent work needs to provide more direct estimates of local density if we are to see if these are general trends across wide categories of plant species.

In addition to density, our simulation showed that individual variation in discoverability, a surrogate we used for reproductive output, created much better models of gene flow than those with distance (IBD) alone. Our parameterization of $H$ was somewhat arbitrary, though within the range of variability that could be measured on features such as nectar volume, floral architecture, color, display, and scent (Huber et al. 2005, Nattero et al. 2011, Ye et al. 2011, Karron and Mitchell 2012) as well as localized features of the immediate neighborhood within which the individual is located (e.g., local density, spatial arrangement of inflorescences on an
individual, or individuals on the landscape; Loveless and Hamrick 1984, Cartar and Real 1997, Ishii et al. 2008, Elliott and Irwin 2009, Streinzer et al. 2009). The statistical signature of $H$ in pollen pool genetic structure is weakened by reductions in density in part due to increased pollination distances. Thus a potential interaction between pollen donor density and estimation of local variability is important to consider, since they can both produce similar patterns of spatial genetic structure. We suggest that analyses of genetic connectivity pay particularly close attention to aspects of study design such that the effects of these confounding factors can be either isolated or factored out.

3.5.3 Empirical Gravity Model: *Cornus florida*

We found that pollen pool structure in *C. florida* was best described by models containing both at-site and between-site landscape features. Differentiation among pollen pools decreased with increasing floral display and clumping in the opening of the primary canopy over individuals. Large floral display is known to increase the visitation rate of pollinators across a wide range of plant species (e.g., Thompson 2001, Grindeland et al. 2005), presumably by increasing attractiveness to pollinators. Likewise, gaps in the primary canopy can influence pollinator behavior by altering understory plant growth and resource allocation (McConnaughay and Coleman 1999), changing temperature and moisture profiles of understory environments (Herrera 1995), and altering visibility of floral displays to pollinators overhead (Walters and Stiles 1996).

Our best-fit models also suggested that two aspects of the intervening landscape further influenced pollen flow: deciduous forest canopy and open canopy. Dyer *et al.* (2012), using a univariate correlative approach considering only intervening landscape variables, identified both open canopy and *C. florida* canopy as landscape features correlated with Pollination Graph structure. It is not surprising that a set of individual univariate correlations of between-site
variables and a model built from a multiple regression using both at-site and between-site variables may find a partially non-overlapping set of predictors. The inclusion of additional factors in a gravity model context, particularly at-site features, provided a statistically more robust characterization of gene flow than those that include only IBD and IBR terms.

Theory predicts that IBD should be pervasive in plants due to spatially restricted seed dispersal and mating. Over the past decade there has been a huge increase in the use of dispersal kernels to describe landscape-level pollen movement (e.g., Dick et al. 2003, Robledo-Arnuncio and Gil 2005, Fortuna et al. 2008), and even those that include features beyond spatial separation (e.g., Kamm et al. 2010, Moran and Clark 2011) do not incorporate at-site variability. A potential challenge associated with using dispersal kernels is that they homogenize over landscape heterogeneity, which our results suggest may be biologically informative. By definition, dispersal kernels are both continuous in space (even Euclidean) and isotropic (though some extensions to this have been suggested). Both of our simulation and empirical results suggested that heterogeneity in landscape features and variation in individual plant characteristics can add significantly to our understanding of what influences pollen-mediated gene flow. Whereas dispersal kernels gain generality by averaging out this heterogeneity, network approaches can be used to implicitly embrace landscape heterogeneity to gain further insight into the underlying processes influencing contemporary gene flow.

3.5.4 Conclusions
The major emphasis in what we currently refer to as landscape genetics, an incompletely defined extension of population genetics, is to identify how intervening landscape heterogeneity shapes genetic connectivity and potentially future population genetic structure. Here we show that at-site variables and individual plant characteristics are also important determinants of gene flow in plant-pollinator systems. The inclusion of these factors increases the statistical power of models
describing gene flow providing a more robust understanding of the factors influencing realized connectivity. We suspect that future work that specifically incorporates at-site features will continue to provide more robust statistical characterization and better overall biological inferences than simplified approaches based on IBD and IBR alone.

3.6 Acknowledgements
This work was completed during a Distributed Graduate Seminar conducted through the National Center for Ecological Analysis and Synthesis, and funded by the American Genetics Association and the Canadian Institute for Ecology and Evolution. Portions of the empirical data set were supported by NSF DEB-0640803 awarded to RJD and some field measurements were collected by C. Meadows as a component of her master’s thesis. We thank Nusha Keyghobadi and Jeremie Fant for comments on an earlier version of the manuscript as well as the comments from three anonymous reviewers that greatly improved this manuscript. This manuscript is the Virginia Commonwealth University Rice Center Contribution #42.

3.7 Author Contributions
The collaboration amongst authors on this manuscript was initiated as a group project in a Landscape Genetics Distributed Graduate Seminar in 2012. RJD conceived the study, provided the empirical data and wrote the simulation code. MD and RJD analyzed the data. All authors contributed to the design of the study and writing of the manuscript.
Table 3-1: Independent predictor variables used to build gravity models to explain connectivity by pollen flow in *C. florida*. Parameter: indicates whether the parameter in the gravity model (Eqn 1) represents spatial distance between sites (*d*), at-site variables (*m*), or between-site ecological variables (*e*).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Code</th>
<th>Variable</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>d</em></td>
<td><em>d</em></td>
<td>Euclidean distance</td>
<td>UTM coordinates</td>
<td>Distance between sampling locations in meters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diameter at breast height</td>
<td>field measurements</td>
<td>Diameter of maternal tree trunk at breast height in centimeters</td>
</tr>
<tr>
<td><em>m</em></td>
<td><em>dbh</em></td>
<td>diameter at breast height</td>
<td>field measurements</td>
<td>Diameter of maternal tree trunk at breast height in centimeters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>canopy openness</td>
<td>canopy photographs</td>
<td>Percent of open sky above the maternal tree</td>
</tr>
<tr>
<td></td>
<td><em>clump</em></td>
<td>canopy clumping</td>
<td>canopy photographs</td>
<td>Degree of clumping of canopy above maternal tree</td>
</tr>
<tr>
<td></td>
<td><em>flor</em></td>
<td>floral output</td>
<td>field measurements</td>
<td>Total number of inflorescences per maternal tree</td>
</tr>
<tr>
<td><em>e</em></td>
<td><em>open</em></td>
<td>open fields</td>
<td>hyperspectral imagery</td>
<td>Variance of probability of open canopy occurrence due to fields along transect between maternal trees</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deciduous primary canopy</td>
<td>hyperspectral imagery</td>
<td>Mean probability of mixed hardwood canopy occurrence in the forest overstory along transects between maternal trees</td>
</tr>
<tr>
<td></td>
<td><em>pine</em></td>
<td>pine primary canopy</td>
<td>hyperspectral imagery</td>
<td>Variance of probability of conifer canopy occurrence in overstory along transects between maternal trees</td>
</tr>
<tr>
<td></td>
<td><em>roads</em></td>
<td>roads</td>
<td>LIDAR</td>
<td>Mean probability of open corridor occurrence due to roads roads along transects between maternal trees</td>
</tr>
<tr>
<td></td>
<td><em>cornus</em></td>
<td>cornus canopy</td>
<td>field locations of dogwoods</td>
<td>Mean occurrence of <em>Cornus florida</em> canopy in understory along transects between maternal trees</td>
</tr>
</tbody>
</table>
Table 3-2: Relative model fit of gravity models constructed with spatial distance and at-site and between-site parameters and fit to *C. florida* connectivity data. Parameter prefixes indicate predictor variable as Euclidean distance (*d*), variable measured at the location of the individual tree (*m*), or features of the intervening landscape (*e*).

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>R²</th>
<th>AIC</th>
<th>∆AIC</th>
<th>wᵢ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d + m(\text{clump}) + m(\text{flor}) + m(dbh) + m(pctsky) + e(open) + e(roads) + e(\text{cornus}) + e(\text{decid}) + e(\text{pine})$</td>
<td>0.57</td>
<td>71.6</td>
<td>23.5</td>
<td>4.00E-06</td>
</tr>
<tr>
<td>$d + m(\text{clump}) + m(\text{flor}) + m(pctsky) + e(open) + e(roads) + e(\text{cornus}) + e(\text{decid}) + e(\text{pine})$</td>
<td>0.57</td>
<td>68.1</td>
<td>20.0</td>
<td>2.00E-05</td>
</tr>
<tr>
<td>$d + m(\text{clump}) + m(\text{flor}) + e(open) + e(roads) + e(\text{cornus}) + e(\text{decid}) + e(\text{pine})$</td>
<td>0.56</td>
<td>67.1</td>
<td>19.0</td>
<td>3.30E-05</td>
</tr>
<tr>
<td>$m(\text{clump}) + m(\text{flor}) + e(open) + e(roads) + e(\text{cornus}) + e(\text{decid}) + e(\text{pine})$</td>
<td>0.56</td>
<td>61.9</td>
<td>13.8</td>
<td>4.50E-04</td>
</tr>
<tr>
<td>$m(\text{clump}) + m(\text{flor}) + e(open) + e(roads) + e(\text{decid}) + e(\text{pine})$</td>
<td>0.55</td>
<td>58.1</td>
<td>10.0</td>
<td>0.003</td>
</tr>
<tr>
<td>$m(\text{clump}) + m(\text{flor}) + e(open) + e(roads) + e(\text{decid})$</td>
<td>0.54</td>
<td>53.1</td>
<td>4.99</td>
<td>0.037</td>
</tr>
<tr>
<td>$m(\text{clump}) + m(\text{flor}) + e(open) + e(\text{decid})$</td>
<td>0.56</td>
<td>48.1</td>
<td>0</td>
<td>0.45</td>
</tr>
<tr>
<td>$m(\text{clump}) + m(\text{flor}) + e(open)$</td>
<td>0.57</td>
<td>49.0</td>
<td>0.91</td>
<td>0.292</td>
</tr>
<tr>
<td>$m(\text{flor}) + e(open)$</td>
<td>0.58</td>
<td>49.5</td>
<td>1.43</td>
<td>0.222</td>
</tr>
<tr>
<td>$d$</td>
<td>0.50</td>
<td>61.4</td>
<td>13.3</td>
<td>5.90E-04</td>
</tr>
</tbody>
</table>
Figure 3-1: Pollen pool differentiation ($\Phi_{FT}$) of populations simulated at different pollen donor densities. Notches represent 95% confidence intervals on the mean of 50 replicate simulations for each density.
Figure 3-2: The influence of tree detectability (simulated as $H$) in gravity models for high and low simulated pollen donor densities. (a) Differences in model fit for IBD model ($w$ alone as a predictor) minus that of the local discoverability ($H$) for low and high densities (0.1% and 10% site occupancy) pollen donor populations. Positive $\Delta$AIC indicates that the model containing $H$ was selected as the preferred model in the model selection process. (b) The relative importance of independent predictor variables in the gravity models for both low and high-density populations as quantified by Akaike weights. Relative variable importance was calculated by summing the Akaike model weights of all models that included that parameter.
Chapter 4

4 Highly Polymorphic Microsatellite Markers in *Pulsatilla vulgaris* (Ranunculaceae) Using Next Generation Sequencing


Author contributions: MFD conducted most lab work, developed and tested the markers from 454 sequence data, and wrote the manuscript. RG and RH helped with marker optimization. YR and HHW collected the *P. vulgaris* samples from the field, and YR prepared DNA samples for 454-sequencing.

4.1 Abstract

We developed novel microsatellite markers for the perennial plant *Pulsatilla vulgaris*, to investigate the effects of fragmentation on gene flow in this imperiled species. We identified microsatellites and developed primers based on 454 shotgun sequences. We identified 14 markers that were polymorphic and produced clean bands. Of these, eight could be analyzed as diploids. Genotyping of 97 individuals across two populations revealed these markers to be highly polymorphic with seven to 17 alleles per locus and observed heterozygosity from 0.41 to 0.83. The markers are highly informative and will be used to test if the reintroduction of shepherding in southern Germany improves genetic connectivity among fragmented populations of *P. vulgaris*. The combination of diploid and tetraploid markers presented here will be useful in resolving the polyploidization history of this and related species.
4.2 Introduction

Pulsatilla vulgaris (L.) Mill (Ranunculaceae) is an early-flowering perennial herb of conservation concern and a flagship and specialist species of calcareous grasslands across central Europe, ranging from France in the south to Sweden at its northern limit (Wells and Barling 1971, Pfeifer et al. 2002). Over the last century, P. vulgaris has witnessed rapid decline and local extirpation across its range, and is listed as “near threatened” by IUCN (2014). Its decline is linked to the abandonment of traditional grazing practices, which has resulted in the severe loss and fragmentation of calcareous grasslands following afforestation (Butaye et al. 2005), and increased above-ground competition from coarse grasses (Walker and Pinches 2011). Knowledge of the landscape-scale distribution of genetic variation is required to create effective management plans for fragmented populations, and evaluations of genetic diversity and inbreeding will allow the identification of populations that are at highest risk of extirpation. This, however, requires genetic markers with high resolution such as microsatellites. No such markers are yet available for P. vulgaris, and any potential genetic analyses are complicated as this species is suspected to be an allotetraploid (2n = 32; Bocher 1932). We therefore de novo developed microsatellite markers for this calcareous plant species, specifically selecting loci that can be analyzed as diploid or double-diploids in downstream genetic analyses.

4.3 Methods and Results

We extracted genomic DNA from homogenized leaf tissue from two populations in the Franconian Alb, Germany (A25: 48°57’38.7”N, 10°56’32.2”E; A03: 48°59’51.7”N, 11°3’35.4”E), using the QIAGEN DNeasy Plant Mini kit (QIAGEN, Mississauga, ON, Canada) following the manufacturer’s protocol. A voucher from the study area will be deposited at the TRTE Herbarium at the University of Toronto Mississauga. DNA from five individuals was
mixed, and a 4 μl sample was sent to LGC Genomics in Berlin, Germany for 454 shotgun sequencing on 1/8 of a plate using GS-FLX Titanium (Roche, 454 Life Sciences, Brantford, CT, USA). In total, 92,833 reads were returned and assembled into 4,088 contigs. We used MSATCOMMANDER version 1.0.8 to identify di-, tri-, and tetra-nucleotide microsatellites to develop primers in regions flanking the identified microsatellites (Faircloth 2008). A total of 457 microsatellites were identified and 76 suitable primer pairs could be designed. We amplified 75 of these pairs by PCR in an initial screen of 10 individuals, and of these, 18 showed clear peaks and 14 of these were polymorphic (Table 4-1). Eight markers showed evidence of disomic inheritance (e.g. had a maximum of two alleles per individual) and were retained for further analysis (Table 4-2).

We amplified these eight microsatellites in two multiplex reactions in 97 individuals with the QIAGEN Multiplex PCR Kit (QIAGEN, Mississauga, ON, Canada) in 10 μl final reaction volumes with 0.2 μM of each primer, 4.6 μl of Multiplex Mix, 1.2 μl of DNAse-free water, and 5-30 ng of template DNA. Thermocycling conditions for Multiplex A (Table 4-1) strictly followed the manufacturer’s protocol. Multiplex B was amplified using a touchdown procedure with an initial denaturation at 95 °C for 15 minutes; 10 cycles of denaturation at 95 °C for 60 seconds, annealing at 65-55 °C for 60 seconds, and extension at 72 °C for 90 seconds; 25 cycles of 95 °C for 60 seconds, 55 °C for 60 seconds, and 72 °C for 90 seconds; and a final extension at 72 °C for 10 minutes. PCR products were diluted 20x and 1-2 μl of product was loaded to an ABI3730xl Capillary Sequencer (Applied Biosystems, Burlington, ON, Canada) using GeneScan 500 LIZ (Life Technologies, Burlington, ON, Canada) size standard for fragment analysis at the Centre for Applied Genomics (The Hospital for Sick Children, Toronto, Canada). Genotyping was performed in GENEMARKER 2.4.0 (SoftGenetics, State College, PA, USA).
All diploid markers except for PV65 amplified a maximum of two alleles per individual. PV65 amplified a maximum of four alleles per individual in two sets of size-separated fragments (PV65a and b, Table 4-1). Loci showed high levels of polymorphism with seven to 17 alleles per locus, and observed heterozygosity from 0.41 to 0.83 per locus (Table 4-2). Exact tests performed in GENEPOP 4.2 (Rousset 2008) revealed three loci with significant departures from Hardy-Weinberg equilibrium (HWE), but departures were not consistent across populations, except for locus PV64. It is possible that selfing in one or both populations, or the presence of population substructure (i.e. the Wahlund effect; Wahlund 1928) led to this homozygote excess. Test for departures from linkage equilibrium expectations in GENEPOP showed that all pairs of loci were unlinked (results not shown).

We used open pollinated progeny arrays to confirm that the identified “diploid” loci followed a disomic pattern of inheritance by amplifying the eight loci in ten seeds collected from each of 15 mother plants. For PV65, all offspring genotypes had maternal contributions from both of the size-separated amplicons, suggesting that PV65a and PV65b represent independent genomic complements with disomic inheritance. For the remaining loci, disomic inheritance was confirmed by the observation of a maximum of two alleles for all offspring, with each offspring sharing an allele with the maternal genotype.

4.4 Conclusions

These new microsatellite markers are highly informative and will be used to quantify gene flow across fragmented populations of *P. vulgaris* in southern Germany. We will test whether the reintroduction of shepherding is a suitable conservation measure to improve genetic connectivity among populations of this species. The combination of diploid and tetraploid markers presented here will be useful in clarifying the polyploidization history of this and related species.
4.5 Acknowledgements

We thank Nimesh Patel, Amaneet Lochab, and Qasim Muhammad for assistance with DNA extraction and genotyping. We also thank Juergen Boehmer and Mr. Dadrich for help in the field. Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC CGS-D and Michael Smith Foreign Study Supplement to MFD, and Discovery Grant to HHW) and the Government of Central Franconia, Bavaria, German.
Table 4-1: Characteristics of 18 *de novo* developed microsatellite markers for *Pulsatilla vulgaris*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer Sequence (5' to 3')</th>
<th>Repeat Motif</th>
<th>$T_m$ (°C)</th>
<th>Multiplex</th>
<th>Dye†</th>
<th>Size range (bp)</th>
<th>Na</th>
<th>Marker Ploidy</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV2</td>
<td>F: GTTGCGATGATCAGAAGTGC</td>
<td>[AAAT]$_6$</td>
<td>55</td>
<td>A</td>
<td>HEX</td>
<td>410-426</td>
<td>4</td>
<td>diploid</td>
<td>KP885677</td>
</tr>
<tr>
<td></td>
<td>R: AGAACTCTCCAGAACAAGGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV7</td>
<td>F: ACCGCAACATAGCAAACAC</td>
<td>[AG]$_{10}$</td>
<td>TD 65-55</td>
<td>B</td>
<td>6-FAM</td>
<td>326-362</td>
<td>8</td>
<td>diploid</td>
<td>KP885678</td>
</tr>
<tr>
<td></td>
<td>R: ACCCACCACAACCTTGAGAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV9</td>
<td>F: GAACTAACTGCTTGCGTC</td>
<td>[AG]$_{11}$</td>
<td>55</td>
<td>single</td>
<td>HEX</td>
<td>283-309</td>
<td>12</td>
<td>tetraploid</td>
<td>KP885679</td>
</tr>
<tr>
<td></td>
<td>R: GCAAGACAAAGTCCACTCTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV16</td>
<td>F: TTGTTGGGTCTGTGAGAAG</td>
<td>[AC]$_{12}$</td>
<td>55</td>
<td>single</td>
<td>6-FAM</td>
<td>330-362</td>
<td>11</td>
<td>tetraploid</td>
<td>KP885680</td>
</tr>
<tr>
<td></td>
<td>R: ATCACTTGTAGCCTCCGGTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV18</td>
<td>F: CCATGACACCTTGTCACGAG</td>
<td>[AC]$_{10}$</td>
<td>55</td>
<td>single</td>
<td>6-FAM</td>
<td>415-451</td>
<td>11</td>
<td>tetraploid</td>
<td>KP885681</td>
</tr>
<tr>
<td></td>
<td>R: AGGACCCAAGATCTCAAGCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV27</td>
<td>F: AACCCCTGCCACACCAACTTG</td>
<td>[AG]$_{8}$</td>
<td>55</td>
<td>A</td>
<td>6-FAM</td>
<td>392-454</td>
<td>5</td>
<td>diploid</td>
<td>KP885682</td>
</tr>
<tr>
<td></td>
<td>R: AACTTTACTCTGGGCGGGAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV32</td>
<td>F: CATGCTTTTGATACCTGCTG</td>
<td>[AAG]$_9$</td>
<td>55</td>
<td>single</td>
<td>6-FAM</td>
<td>356-422</td>
<td>8</td>
<td>tetraploid</td>
<td>KP885683</td>
</tr>
<tr>
<td></td>
<td>R: AGACCTTTTTGACCTCTGCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV33</td>
<td>F: AGCCCTTGTGTATTTTGGGC</td>
<td>[AC]$_{9}$</td>
<td>TD 65-55</td>
<td>B</td>
<td>6-FAM</td>
<td>442-452</td>
<td>8</td>
<td>diploid</td>
<td>KP885684</td>
</tr>
<tr>
<td></td>
<td>R: GCTCACCTTGACCAACTCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV44</td>
<td>F: GTATGTGTGTGCCAAAGGTC</td>
<td>[ATC]$_{10}$</td>
<td>55</td>
<td>single</td>
<td>HEX</td>
<td>429</td>
<td>1</td>
<td>-</td>
<td>KR109213</td>
</tr>
<tr>
<td></td>
<td>R: TGCTAAGAGTGGCATGCCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV48</td>
<td>F: CCGGGGTGAATCTGATGCTC</td>
<td>[AG]$_{10}$</td>
<td>55</td>
<td>single</td>
<td>HEX</td>
<td>235</td>
<td>1</td>
<td>-</td>
<td>KR109214</td>
</tr>
<tr>
<td></td>
<td>R: GCAAATGGACGCAGTTCCACATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV50</td>
<td>F: GATGTGATGAGGGTTTGGG</td>
<td>[AAT]$_{12}$</td>
<td>55</td>
<td>single</td>
<td>HEX</td>
<td>406-466</td>
<td>10</td>
<td>tetraploid</td>
<td>KP885685</td>
</tr>
<tr>
<td></td>
<td>R: TGCCACCTACTTTCCACACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV52</td>
<td>F: TTGTCAATGGTCGCAACG</td>
<td>[AG]$_{10}$</td>
<td>55</td>
<td>A</td>
<td>6-FAM</td>
<td>262-280</td>
<td>7</td>
<td>diploid</td>
<td>KP885686</td>
</tr>
<tr>
<td></td>
<td>R: GGGTCTCAAGATTATCGGCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV54</td>
<td>F: TACTCGGACCTGACAAAGCTC</td>
<td>[AT]$_{10}$</td>
<td>55</td>
<td>single</td>
<td>6-FAM</td>
<td>204</td>
<td>1</td>
<td>-</td>
<td>KR109215</td>
</tr>
<tr>
<td></td>
<td>F: TTGTGGAGTGGAGGAAGACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV56</td>
<td>F: CGTCACGTGTTGGCTAAATC</td>
<td>[AC]$_{11}$</td>
<td>TD 65-55</td>
<td>B</td>
<td>6-FAM</td>
<td>236-270</td>
<td>9</td>
<td>tetraploid</td>
<td>KP885687</td>
</tr>
<tr>
<td></td>
<td>R: TCCTTTGTACTCTCCTCAACGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-1 Continued

<table>
<thead>
<tr>
<th></th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Allele Size</th>
<th>Amplification Dye</th>
<th>Annealing Temperature</th>
<th>Ploidy</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV57</td>
<td>F: GTGCAAAATTACTCACACTGCAG</td>
<td>R: TGCTCGAAACCATAAGTCTGC</td>
<td>[AC]₁₀ 55</td>
<td>single</td>
<td>6-FAM 419-429</td>
<td>5</td>
<td>tetraploid</td>
</tr>
<tr>
<td>PV64</td>
<td>F: AGCAATCTCTGTCCGCTAG</td>
<td>R: GTGACTGCAGATGTGGTGTTGG</td>
<td>[AC]₁₃ TD 65-55</td>
<td>B</td>
<td>HEX 368-404</td>
<td>7</td>
<td>diploid</td>
</tr>
</tbody>
</table>

Abbreviations: T<sub>a</sub>, annealing temperature; N<sub>a</sub>, total number of alleles based on initial screen of 10 individuals

†Fluorescent dye on forward primers used for fragment analysis
Table 4-2: Results from HWE exact tests for diploid markers developed for *Pulsatilla vulgaris* in two populations, A03 and A25.

<table>
<thead>
<tr>
<th>Locus</th>
<th>A03 (n=41)</th>
<th>A25 (n=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N&lt;sub&gt;a&lt;/sub&gt;</td>
<td>H&lt;sub&gt;o&lt;/sub&gt;</td>
</tr>
<tr>
<td>PV2</td>
<td>8</td>
<td>0.67</td>
</tr>
<tr>
<td>PV7</td>
<td>16</td>
<td>0.59</td>
</tr>
<tr>
<td>PV27</td>
<td>17</td>
<td>0.83</td>
</tr>
<tr>
<td>PV33</td>
<td>7</td>
<td>0.59</td>
</tr>
<tr>
<td>PV52</td>
<td>13</td>
<td>0.75</td>
</tr>
<tr>
<td>PV56</td>
<td>13</td>
<td>0.55</td>
</tr>
<tr>
<td>PV64</td>
<td>11</td>
<td>0.41</td>
</tr>
<tr>
<td>PV65a</td>
<td>13</td>
<td>0.41</td>
</tr>
<tr>
<td>PV65b</td>
<td>13</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Abbreviations: N<sub>a</sub>, total number of alleles; H<sub>o</sub>, observed heterozygosity; H<sub>e</sub>, expected heterozygosity. P-values are corrected for multiple testing.
5 An Ecological Network Maintains Genetic Diversity and Fitness of a Flagship Wildflower

This paper will be submitted for publication as: DiLeo MF, Rico Y, Boehmer HJ, Wagner HH
“An ecological network maintains genetic diversity and fitness of a flagship wildflower”

Author contributions: MFD conducted the majority of the lab work, analysed the data, and wrote the manuscript. YR, HJB, and HHW collected *P. vulgaris* samples in the field, and helped to edit the manuscript. YR conducted a subset of DNA extractions.

5.1 Abstract

Ecological networks have been proposed as an efficient way to reconnect communities in fragmented landscapes. Yet few studies have evaluated if they are successful at enhancing actual functional connectivity of focal species, or if this enhanced connectivity is enough to maintain genetic diversity and fitness of plant populations. Here we test the efficacy of an ecological network implemented in southern Germany in 1989 to reconnect calcareous grassland fragments through rotational shepherding. We genotyped 1,449 individuals from 57 populations using seven microsatellite markers and measured fitness-related traits in 10 populations of *Pulsatilla vulgaris*, a flagship species of calcareous grasslands in Europe. We tested if the ecological network explained functional connectivity in *P. vulgaris* and if higher connectivity translated to higher genetic diversity and fitness of populations. We found that genetic distance between populations correlated best with shepherding distance, but that geographic distance may also play a role at the level of individual grazing routes. Populations that were well-connected within the shepherding network had significantly higher genetic diversity than populations that were more isolated or ungrazed, and genetic diversity was significantly positively correlated with both seed set and seed weight. Together our results suggest that the ecological shepherding network is
an effective conservation management measure to maintain functional connectivity, genetic diversity, and fitness at the landscape scale for a calcareous grassland specialist. Ungrazed populations and those with low connectivity within the network suffer from reduced genetic diversity, and these populations would benefit from inclusion, or better integration into the ecological network, respectively. Our study has empirically demonstrated the complete pathway of predicted positive associations between connectivity, genetic diversity, and fitness at the landscape scale, and provides a framework for testing the efficacy of ecological networks for focal species using genetic tools.

5.2 Introduction

Habitat loss and fragmentation are major threats to the persistence of populations across nearly all taxonomic groups (Fischer and Lindenmayer 2007). Together, these processes can lead to reductions in the effective population size, dispersal, and subsequent gene flow among previously contiguous patches, enhancing the effects of genetic drift and accelerating the loss of genetic diversity (Reed and Frankham 2003, Frankham 2005). Ultimately, small and isolated populations are more prone to inbreeding depression and suffer a reduced potential for adaptation (Frankham 2005). Preserving genetic diversity is not only important for the population or species at-hand, but also plays a critical role in the functioning of communities and ecosystems, with positive influences on species diversity, disease dynamics, food-web dynamics, and ecosystem cycling (Johnson et al. 2006, Hughes et al. 2008, Lamy et al. 2013).

Ecological networks (i.e. sets of connected suitable habitats that allows persistence of a viable metapopulation) are increasingly being implemented to restore dispersal linkages between remnant patches in fragmented landscapes (Boitani et al. 2007, Whitelaw and Eagles 2007, Baguette et al. 2013, Maiorano et al. 2015). Dispersal linkages are often structural features of the
landscape, such as habitat corridors or stepping-stones, that are created, protected, or maintained under the assumption that they directly support the movement of organisms between fragments (known as functional connectivity). For plants, an alternative strategy to reconnect populations exists where vectors of seed or pollen dispersal, instead of structural aspects of the landscape, are restored for focal species. For example, rotational shepherding has the potential to disperse seeds of wildflowers over very long distances (e.g. >100km; Fischer et al. 1996, Manzano and Malo 2006) and has been used as a strategy to reconnect grassland populations in central Europe (Butaye et al. 2005, Auffret et al. 2012). Both strategies are based on the assumption that individual corridors, whether physical or not, increase the functional demographic or genetic connectivity of populations or communities. There is an extensive literature that appears to support this notion for individual corridors, showing that they can be successful at preserving functional connectivity of plants and animals (reviewed in Gilbert-Norton et al. 2010). However, ecological networks consist of tens, or hundreds of individual corridors and linkages and it is unclear if these same benefits endure at larger spatial scales.

The spatial extent of ecological networks and their large number of corridors often precludes the collection of base-line connectivity data to help inform decisions about the suitability of linkages for protection. Most often, ecological networks are designed solely based on the structural connectivity of the landscape (Cushman et al. 2009). More recently, conservation managers have considered potential functional connectivity when designing networks, using dispersal thresholds of focal species to select linkages for protection (e.g. Bruinderink et al. 2003, Cushman et al. 2009, Carroll et al. 2012). However, rarely do we test if ecological networks are successful at maintaining actual functional connectivity (i.e. realized dispersal or gene flow, Calabrese and Fagan 2004, but see Melles et al. 2012), or test if it is enough to maintain genetic diversity. Consequently, there is a great need to evaluate the utility of
ecological networks for focal species and communities (Boitani et al. 2007). Neutral genetic diversity on its own may not be a suitable conservation target as the argument to conserve diversity is made on the basis of its presumed association with fitness; for selectively neutral markers this evidence is equivocal across the plant world (Leimu et al. 2006). Thus a better evaluation of the outcomes of conservation planning involves applied targets such as fitness components or a demonstrated link between diversity and fitness.

Here we test the efficacy of an ecological network implemented in southern Germany in 1989 to reconnect abandoned calcareous grassland fragments through rotational shepherding. We measure genetic differentiation, genetic diversity and fitness-related traits in the perennial wildflower *Pulsatilla vulgaris* – a flagship and specialist species of calcareous grasslands in Europe and one of high conservation concern (IUCN 2014). We test the hypothesis that the shepherding network maintains seed dispersal and thus gene flow of populations, and that this enhanced gene flow translates to higher genetic diversity and fitness of populations. Previous ecological research in this system showed that shepherding connectivity is associated with increased species richness (Wagner et al. 2013), colonization rates (Rico et al. 2012) and patch occupancy (Rico et al. 2014a) of characteristic calcareous grassland plants. At the molecular level, research on a single species, *Dianthus carthusianorum*, indicated that shepherding decreased genetic divergence among connected populations (Rico et al. 2014b), and increased within-population neutral genetic diversity (Rico et al. 2014a). The question that remains is whether such an increase in genetic diversity is enough make an impact on the fitness of populations, specifically so for species of concern such as *P. vulgaris*. We ask: (1) Does the shepherding network explain gene flow among *P. vulgaris* populations, as quantified by pairwise genetic differentiation? (2) Does the potential enhanced gene flow provided by the shepherding network translate to higher genetic diversity in connected populations? (3) Does higher genetic
diversity translate to higher fitness in this system? This system is ideal as it allows us to compare genetic diversity of patches in the network to ungrazed controls, but also allows a comparison across patches with varying degrees of connectivity within the network to control for possible direct effects of grazing on growth and reproduction.

5.3 Methods
5.3.1 Data Collection

This study was conducted in the Franconian Alb, Germany in a 10 x 15 kilometer region, containing 96 calcareous grassland fragments embedded in a matrix of agriculture, forest, and settlements. Although semi-natural, calcareous grasslands are of high conservation value as they represent one of the of most biodiverse ecosystems in central and northern Europe (WallisDeVries et al. 2002, Butaye et al. 2005). Abandonment of traditional grazing practices over the past century has led to a significant loss of calcareous grasslands, and previously contiguous patches have been fragmented by forest succession and urban development (Dolek and Geyer 2002). In 1989, an ecological network was implemented to reconnect fragmented patches of grasslands via rotational shepherding. The ecological network consists of three non-overlapping shepherding routes (Fig. 5-1). Prior to the implementation of the network, a baseline survey was conducted to record all vascular plants in previously abandoned grassland patches (Boehmer et al. 1990) and in 2009 this survey was repeated (Wagner et al. 2013).

*Pulsatilla vulgaris* (Ranunculaceae) is an early-flowering perennial herb of conservation concern designated as “near threatened” by IUCN (IUCN, 2015) and a flagship species of calcareous grasslands across central Europe. It is hermaphroditic and mainly outcrossing, and typically produces between one and three purple flowers in early spring (March-April) which each yield 40-100 seeds (Wells and Barling 1971). Seeds have long, feathery styles and although
they appear to be adapted for wind dispersal, those carried by wind rarely make it further than 20 cm from the plant (Wells and Barling 1971). Previous work in the study area showed that patch occupancy in 2009 was best explained by shepherding connectivity, suggesting that sheep are important dispersal vectors for *P. vulgaris* seeds (Rico et al. 2014a).

From April to May 2009, we collected leaf material of flowering plants (*n*=1,449) from all patches (*n*= 57) containing *P. vulgaris*. All flowering individuals were sampled in populations with fewer than 40 individuals. For populations that exceeded 40 individuals, we collected leaves from 30-40 flowering plants from across the population. Of the 57 containing *P. vulgaris*, 19 are ‘core areas’ that have been consistently grazed over the last 200+ years, and the remaining 38 occur in ‘previously abandoned’ patches that have been abandoned prior to 1960, and since the implementation of the management program have been either consistently grazed (every year, 3-5 times/season, *n*=17), intermittently grazed (grazed only within the first few years after 1989, *n*=6), or have remained ungrazed (n=15). Unlike Rico *et al.* (2012), we classified patches that are grazed only late in the season (August and onwards) as ungrazed, since *P. vulgaris* is an early-flowering species and we expect seed dispersal to occur shortly after seeds ripen in mid-May (Wells and Barling 1971). Population size class was estimated for each patch using four categories: 1-3 individuals, 4-39 individuals, 40-99 individuals, and ≥100, following the scheme of Rico *et al.* (2014a). For those populations where we collected fitness data (see below), census population size was recorded in 2013 as the number of flowering individuals.

We extracted genomic DNA from dried leaves using QIAGEN DNeasy Plant Mini Kit (QIAGEN, Mississauga, ON, Canada) following the manufacturer’s protocol. *P. vulgaris* is an allotetraploid (*2n=4x=32; Wells and Barling 1971*), and to simplify analysis we used microsatellite markers developed for this species that could be analyzed as diploid (DiLeo et al.
For each individual, we amplified seven species-specific microsatellites in two multiplex reactions (Multiplex A: PV2, PV27, PV65a PV65b; Multiplex B: PV7, PV33, PV56), conducted fragment analysis and genotyping, following the protocols described in DiLeo et al. (2015). In June 2013 we collected mature seed heads to measure fitness-related traits from 7-10 individuals in 10 populations (see A3, Appendix). We chose a combination of populations that differed in size and isolation, with wide coverage across the study region. To control for variation in seed production due to flowering phenology, flower buds were marked in populations within the same two-day time period during peak flowering in April, and seeds were collected from the marked flowers once matured. In the laboratory, developed and undeveloped seeds were counted, and the developed seeds were weighed per seed head. Developed seeds were easy to visually distinguish from undeveloped seeds based on size of both the seed and style.

5.3.2 Genetic Analysis of Functional Connectivity

Microsatellite loci were tested for departures from Hardy-Weinberg Equilibrium (HWE) and linkage equilibrium using exact tests in GENEPOP 4.2 (Raymond and Rousset 1995). We used Mantel and partial Mantel tests to test the correlation and partial correlations between inter-patch genetic distance and two ecological distance matrices representing alternative hypotheses of potential functional connectivity:

(1) IBD: a matrix of inter-patch Euclidean geographic distances, assuming that geographically close patches will experience more gene flow than far patches.

(2) IBR: a matrix of inter-patch sheep grazing distances, calculated as the number of patches traversed by sheep along the grazing route (Rico et al. 2014a). This hypothesis assumes that seed-mediated gene flow is determined by sheep as dispersal vectors.
Population genetic distance was calculated as $D_c$, Cavalli-Sforza and Edwards’ chord distance (1967) using the adegenet library (Jombart 2008) in R. We conducted four tests for consistently grazed patches: (i) Mantel tests of genetic distance with $IBD$, and (ii) genetic distance with $IBR$, (iii) partial Mantel tests of genetic distance with $IBD$, controlling for $IBR$, and (iv) genetic distance with $IBR$, controlling for $IBD$. We conducted all tests with data from the three grazing routes pooled to increase sample size and statistical power, and separately for each of the three routes. We additionally tested for IBD in ungrazed patches. We permuted genetic distance matrices 999 times to test for significance of Mantel correlation coefficients.

5.3.3 Genetic Diversity

We evaluated the effects of grazing treatment (ungrazed, intermittently grazed, consistently grazed) on genetic diversity and the inbreeding coefficient, $F_{is}$ with ANOVAs followed by Tukey HSD. We additionally tested for effects of population size class by including it as a factor in a two-way ANOVA, however observations from population size class ≥100 were excluded as only consistently grazed patches were found in this category. $F_{is}$ was calculated using GENEPOP 4.2 (Raymond and Rousset 1995).

Genetic diversity was measured as the mean number of alleles per population ($A_r$) using the gstudio package (Dyer 2014) in R 3.02 (R Core Team 2013). We used rarefaction with a sample size of nine and 999 permutations to control for differences in sample size between populations. Allelic richness was chosen over heterozygosity because it is more sensitive to recent demographic change (Allendorf 1986) and is better indicator of long-term adaptive potential of populations, even when based on neutral markers (Caballero and Garcia-Dorado 2013, Vilas et al. 2015). However, for comparison, we also present results of analyses using three alternative measures of genetic diversity - the effective number of alleles ($A_e$; correlation
with $A_r$: Pearson $r=0.95$), expected heterozygosity ($H_e$: correlation with $A_r$: Pearson $r=0.93$), and observed heterozygosity ($H_o$: correlation with $A_r$: Pearson $r=0.50$). Populations with less than nine individuals were excluded from analysis, as this was the minimum sample size of populations where we measured fitness-related traits.

We tested the hypothesis that populations that are well-connected within the grazing network have higher genetic diversity than more isolated populations using linear mixed effect models. Population connectivity was calculated for each patch within the shepherding network using Hanski’s $S_i$ index (Hanski 1994). The $S_i$ index calculates patch connectivity by summing distances between focal patch $i$ and all seed source patches $j$ using the equation:

$$S_i = \sum_{i \neq j} \exp(-\alpha d_{ij})$$

where $d_{ij}$ is the distance between patch $i$ and patch $j$, and $\alpha$ is a constant scaling parameter accounting for dispersal capacity, which we fitted through optimization (see below). We calculated the distance parameter ($d_{ij}$) based on the two hypotheses of functional connectivity described above for the Mantel tests: IBD, and IBR. For $S_{iIBD}$, $d_{ij}$ was calculated as the Euclidean geographic distance between patches. For $S_{iIBR}$, $d_{ij}$ was calculated as the number of patches traversed by sheep to get from $j$ to focal patch $i$, and pairs of patches from different grazing routes were given a value of 100 assuming that gene flow outside of the grazing routes is rare (Rico et al. 2012). Grassland patches in the network that did not contain $P. vulgaris$ populations (Fig. 5-1) were counted as steps traversed by sheep in the calculation of $S_{iIBR}$. Intermittently grazed patches were excluded from this analysis as they were grazed for only a couple of years after the implementation of the ecological network, and we expect they do not significantly contribute to functional connectivity in $P. vulgaris$. When intermittently grazed patches were
included as part of the network we found a similar ranking of models but lower R^2 values (results not shown). We optimized α for both \( S_{IBD} \) and \( S_{IBR} \) by testing values between 0.01-10 at increments of 0.01 and choosing the values that gave the highest R^2 in univariate regressions with \( A_r \) resulting in α=0.04 for \( S_{IBD} \) and 0.03 for \( S_{IBR} \). We constructed a set of linear mixed effect models to quantify the relationship between genetic diversity \( (A_r) \) and three predictor variables as fixed effects: \( S_{IBD}, \ S_{IBR}, \) and \( \text{population size class} \), and grazing route (herd 1, 2, or 3) as a random effect. Both \( S_{IBD} \) and \( S_{IBR} \) were square root-transformed. Census population sizes were only available for twelve populations and thus could not be included as a continuous variable without sacrificing power. We found that treating population size as a continuous variable for those twelve populations did not change results of model selection (see A4-6, Appendix). Population size classes were included as ranks, which helped to stabilize the distribution of residuals compared to treating them as categorical. We tested all possible sub-models of \( A_r \sim S_{IBD} + S_{IBR} + \text{population size class} \) and used AICc for model selection. We assessed the relative importance of each predictor by summing the Akaike weights \( (w_i) \) of all models that include that predictor. We used the protocol of Nakagawa & Schielzeth (2013) to calculate marginal R^2, which represents variance explained by the fixed effects.

### 5.3.4 Fitness

We evaluated the relationship between genetic diversity \( (A_r) \) and two fitness-related traits \( (seed \ set, \ \text{and seed mass}) \) in separate linear mixed effect models for the 10 populations where we collected seeds. \( Seed \ set \) was measured as the proportion of developed seeds per seed head, and \( seed \ mass \) was measured as the mean mass of developed seeds per seed head. \( A_r \) was included as a fixed effect and was exponentially transformed to linearize the relationship with \( seed \ set \) and \( seed \ mass \). Population was included as a random effect in the models to control for the non-independence of seed data collected from multiple mother plants within the same patch. Models
were estimated in \textit{nlme} (Pinheiro \textit{et al}. 2014) using maximum likelihood. We used likelihood ratio tests to determine the significance of the fixed effect, and report marginal $R^2$ (variance explained by fixed effects, Nakagawa and Schielzeth 2013).

Fitness related traits can be influenced not only by genetic diversity, but by population size, either directly (e.g. Allee effects; Lande 1988, Reed 2005), or indirectly by increasing genetic diversity (Leimu \textit{et al}. 2006). We used Pearson’s partial correlations to tease apart the effects of genetic diversity and population size on seed set in \textit{P. vulgaris} populations. Ideally, a full path analysis could be conducted to quantify the interactions between these variables, but our low sample size ($n=10$) precluded such a test. We calculated Pearson’s product moment correlations between mean seed set per population and \(A_r\), mean seed set and population size, and partial correlations between mean seed set and \(A_r\) controlling for population size, and mean seed set and population size controlling for \(A_r\), using the \texttt{ppcor 1.0} (Kim 2012) in \texttt{R}. Population size was defined as the number of flowering individuals per population. \(A_r\) was exponentially transformed and we took the logarithm of population size to linearize relationships.

5.4 Results

5.4.1 Genetic Analysis of Functional Connectivity

Markers PV7, PV27, PV65\texttext{a} and PV56 showed departures from HWE expectations, but there was no consistent pattern across populations (A7, Appendix) and all markers were retained. Marker pairs were found to be unlinked (results not shown).

We found significant Mantel correlations between genetic distance and both ecological distance matrices, \textit{IBD} and \textit{IBR} for pooled data and for each grazing route individually (Table 5-1). The relationship between genetic distance and IBR remained significant for pooled data after controlling for the effects of IBD in a partial Mantel test, and the reverse between genetic
distance and IBD controlling for IBR was not significant (Table 5-1). We found the same result at the level of individual grazing route for herd 2, but partial Mantel correlations were not significant for the remaining two grazing routes (Table 5-1). We found no significant Mantel correlation between genetic distance and IBD for ungrazed populations (Mantel r=0.095, p=0.45).

5.4.2 Effects of Shepherding Connectivity on Genetic Diversity

Intermittently and consistently grazed populations had significantly higher genetic diversity ($A_r$) than ungrazed populations (ANOVA, $F_{(2,40)}=5.49$, $p=0.008$; Fig. 5-2). The same results were found for both $Ae$ (ANOVA, $F_{(2,40)}=6.57$, $p=0.003$; A8, Appendix) and $H_e$ (ANOVA, $F_{(2,40)}=5.415$, $p=0.008$; A8, Appendix), but we found no significant differences in $H_o$ among populations with different grazing treatments (ANOVA, $F_{(2,40)}=0.277$, $p=0.76$; A8, Appendix). In a two-way ANOVA using populations with sizes 4-39 and 40-99, grazing treatment but not population size class had a significant effect on $A_r$ (ANOVA, grazing treatment: $F_{(2,25)}=3.90$, $p=0.03$, population size class: $F_{(1,25)}=2.8$, $p=0.11$). We found the same result for both $Ae$ (ANOVA, grazing treatment: $F_{(2,25)}=4.89$, $p=0.02$, population size class: $F_{(1,25)}=1.18$, $p=0.29$) and $H_e$ (ANOVA, grazing treatment: $F_{(2,25)}=3.47$, $p=0.05$, population size class: $F_{(1,25)}=1.55$, $p=0.22$) but not $H_o$ (ANOVA, grazing treatment: $F_{(2,25)}=0.16$, $p=0.85$, population size class: $F_{(1,25)}=0.22$, $p=0.64$). Neither grazing treatment nor populations size class had a significant effect on the inbreeding coefficient $F_{is}$ (ANOVA, grazing treatment: $F_{(2,25)}=0.075$, $p=0.93$, population size class: $F_{(1,25)}=0.001$, $p=0.99$).

Genetic diversity ($A_r$) was best fit by a model including connectivity by shepherding ($Si_{IBR}$) and population size as predictors in a linear mixed effect model, controlling for grazing route as a random effect (Table 5-2). The same model was chosen for $Ae$ and $H_e$ (A9, Appendix).
When $H_o$ was used as the response variable, the best model included $Si_{IBR}$ alone as a fixed predictor (A9, Appendix). In univariate models with $A_r$, variance explained was highest for $Si_{IBR}$, followed by population size class, and $Si_{IBD}$ (Table 5-2; Fig. 5-3). Summing Akaike weights across models suggested that $Si_{IBR}$ was the most important predictor variable ($w_s = 0.97$) followed by *population size class* ($w_s = 0.94$) and $Si_{IBD}$ ($w_s = 0.20$).

### 5.4.3 Effect of Genetic Diversity on Fitness

*Seed set* and allelic richness ($A_r$) showed a significant positive linear association (Likelihood ratio test: $\chi^2(1) = 9.2, p = 0.002$; marginal $R^2 = 0.12$; Fig. 5-4a). *Seed mass* also showed a significant positive association with $A_r$ (Likelihood ratio test: $\chi^2(1) = 11.76, p < 0.0001$; marginal $R^2 = 0.23$; Fig. 5-4b). *Seed set* and *seed mass* also showed a significant positive linear association with the effective number of alleles ($A_e$) and expected heterozygosity ($H_e$) but not observed heterozygosity ($H_o$; A10, Appendix).

Mean *seed set* per population showed strong positive Pearson correlations with both $A_r$ ($r=0.82, p=0.003$), and *population size* ($r=0.71, p=0.02$). The partial correlation between mean *seed set* and $A_r$ remained significant after controlling for *population size* ($r=0.64, p=0.03$), but the partial correlation between mean *seed set* and *population size* controlling for $A_r$ was not significant ($r=0.32, p=0.37$).

### 5.5 Discussion

Here we demonstrate the utility of an ecological network to maintain functional connectivity, genetic diversity, and fitness of populations of an imperiled wildflower, *Pulsatilla vulgaris* in calcareous grasslands in Germany. Our genetic data were best fit by a model of connectivity that incorporated the number of patches traversed by sheep in the shepherding network, suggesting that connectivity by shepherding, and not the geographic distance separating grassland patches
determines functional connectivity in *P. vulgaris*. We further show that enhanced connectivity in the shepherding network translated to enhanced genetic diversity, and that populations with higher genetic diversity produced more and heavier seeds. Together these results suggest that shepherding is an effective landscape management measure to sustain functional connectivity among fragmented populations of *P. vulgaris*, but populations that have low connectivity within the network still suffer from reduced genetic diversity.

### 5.5.1 Functional Connectivity

Our results suggest that the shepherding network affects genetic connectivity of *P. vulgaris* populations. Population genetic distances showed a significant correlation with shepherding distance, remaining significant even after controlling for IBD when analyzing the pooled data (Table 5-1). In contrast, genetic distance was no longer significantly correlated with IBD after controlling for the effects of shepherding distance (IBR), implying that geographic distance between patches does not influence gene flow at the landscape scale. This highlights the importance of considering the potential vectors of dispersal and their effect on functional connectivity when designing ecological networks. Considering patch proximity alone may not be enough to ensure dispersal and subsequent gene flow between fragmented populations (Kamm et al. 2010, Dyer et al. 2012). However, when analyzed at the level of individual grazing routes, the partial Mantel correlations of IBR and IBD with genetic distance were not significant for two of the three routes (Table 5-1). This might suggest that both grazing distance and inter-patch geographic distance affect gene flow at this scale, and that the two processes are confounded.

Previous work in this system showed that the rotational sheep grazing network predicted demographic connectivity for *P. vulgaris* and other typical grassland species at the community (Rico et al. 2012) and species level (Rico et al. 2014a), and at the genetic level for another
grassland plant, *Dianthus carthusianorum* (Rico et al. 2014b). The current study suggests that the same trend holds true at the genetic level for *P. vulgaris*, where both genetic distance and genetic diversity were best fit by a model including connectivity via shepherding, but not geographic distance at the landscape scale (Tables 5-1 & 5-2). The effects of fragmentation can manifest at different spatial and temporal scales depending on the population outcome measured (e.g. patch occupancy or abundance versus genetic diversity; Takkis et al. 2013, Jackson and Fahrig 2014) and the importance of considering both demographic and genetic factors for conservation planning is increasingly recognized (Luque et al. 2012, Landguth et al. 2014). Demographic outcomes such as patch occupancy and abundance are mediated by the processes of recruitment and colonization, which in plants occurs through the dispersal of seed. In contrast, genetic differentiation and genetic diversity of local populations is the product of gene flow over many generations, and is thought to mainly occur through pollination for most plant species (Ellstrand 1992, Ennos 1994). However, in some systems the contribution of seed to overall genetic connectivity far exceeds that of pollen, particularly when seeds are dispersed by animal vectors (Bacles et al. 2006, Manzano and Malo 2006). This may explain why we find such a strong effect of shepherding, especially on genetic diversity. The ecological network investigated in this study is different from many others, as it directly provides the functional vectors of seed dispersal (i.e. sheep) rather than protecting physical aspects of the landscape that have the potential to support connectivity by either seed or pollen. However, we used codominant genetic markers that carry the signal of both seed- and pollen-mediated gene flow, so future work is required to resolve the contribution of pollination to connectivity in this system. Ultimately, our results combined with the work of Rico *et al*. (2012, 2014a), Rico *et al*. (2014b), and Wagner *et al*. (2013), imply concordance across functional level (population and community) and data type (demographic and genetic), with multiple lines of evidence supporting a positive role of the
shepherding network in maintaining functional connectivity and diversity of calcareous grassland plants.

5.5.2 Effect of Shepherding Connectivity on Genetic Diversity

We found that intermittently and consistently grazed populations had significantly higher genetic diversity than ungrazed populations (Fig. 5-2), even after controlling for population size. This suggests that populations of *P. vulgaris* that are incorporated into the ecological network are better off. However, it is unclear if the observed effect is due to the enhanced connectivity provided by the ecological network or is a direct result of the grazing process. For example, grazing can influence plant reproductive success by altering vegetation height and habitat quality (Jacquemyn et al. 2003, de Vere et al. 2009), flowering phenology (Lennartsson et al. 2012) or richness and abundance of pollinators (Kormann et al. 2015). When we restricted analyses to only those populations within the ecological network to control for any direct reproductive effects of grazing, we found that both shepherding connectivity (*S_{IBR}* and population size were significant predictors of genetic diversity (Table 5-2; Fig. 5-3). This suggests that the functional connectivity provided by the ecological network maintains genetic diversity of populations, but importantly, populations that are small and have low connectivity within the network still experience a reduction in allelic richness. These small and more isolated populations may be useful targets for connectivity network optimization and restoration (Mijangos et al. 2015).

Interestingly we found that both the inbreeding coefficient (*F_{is}* and observed heterozygosity (*H_o*) were unaffected by population size and grazing treatment. This may suggest that genetic drift rather than direct or bi-parental inbreeding is driving the loss of genetic diversity observed in small and more isolated populations (Leimu et al. 2006).
5.5.3 Effect of Genetic Diversity on Fitness

We found that higher genetic diversity translated to an increase in both mean seed set and seed mass of populations (Fig. 5-4), suggesting a beneficial role of the ecological network on population fitness. Strong positive correlations between genetic diversity and fitness have been observed in a number of plant species (reviewed in Reed and Frankham 2003, Leimu et al. 2006), and theory predicts that populations with higher fitness have a higher probability of persistence (Frankham 2005). However, our results should be interpreted with caution given our low sample size (n=10 populations) and that we measured fitness-related traits in a single year for a long-lived species. Given these caveats, it is surprising to see such a strong effect of genetic diversity on seed set, with 64% of variation explained.

The lack of variation in $F_{is}$ and observed heterozygosity ($H_o$) across populations was surprising given that we found a strong correlation between allelic richness and both measured fitness-related traits. This suggests that inbreeding (which should decrease homozygosity, i.e. $1-H_o$) is not the reason for the low reproductive success in populations with low diversity. One explanation for this result is that the increase in fitness does not have a genetic basis but is rather the result of an Allee effect, where density-dependent mating success or facilitation in large populations leads to higher seed production (Lande 1988). This has been demonstrated in a number of plant systems (reviewed in Reed 2005), including *P. vulgaris* where Hensen *et al.* (2005) found strong positive correlations between population size and both seed set and seed mass in central Germany. However, we found that after controlling for the effects of genetic diversity, the partial correlation between seed set and population size was no longer significant. This suggests that population size does not directly influence fitness, but that large populations have higher genetic diversity, which in turn have higher seed set. An alternative explanation is that pollination plays an important role in maintaining both genetic diversity and fitness of *P.*
vulgaris populations. For example, Breed et al. (2012) found that the diversity of pollen received by flowers was more important than inbreeding in determining progeny fitness in *Eucalyptus socialis*. Sampling more diverse pollen sources opens up the opportunity for female selection of more compatible pollen grains or ‘good genes’, resulting in increased fitness that is independent of inbreeding (Yasui 1998, Armbruster and Rogers 2004). Populations with high genetic diversity will provide more diverse pollen, and likewise populations that have higher connectivity will receive pollen from more diverse sources. Given the high richness and abundance of pollinators in calcareous grasslands (WallisDeVries et al. 2002), pollen-mediated gene flow likely contributes a great deal to patterns of genetic variation in *P. vulgaris* populations. Future work is required to clarify the role of pollination in this system.

### 5.5.4 Conclusions

Ecological networks have been implemented worldwide to try to mitigate the negative effects of habitat loss and fragmentation. Despite their growing popularity as a conservation tool, their ability to enhance functional connectivity enough to maintain genetic diversity and fitness of populations has not been empirically demonstrated at the landscape scale. Here we show that an ecological network in southern Germany has been successful at maintaining functional connectivity of a flagship wildflower, and that this enhanced connectivity translated to enhanced genetic diversity and fitness of populations. Our results suggest that incorporating ungrazed populations into the existing network may be a suitable conservation measure to boost genetic diversity and fitness, and that populations that are more isolated within the network might benefit from further optimization of the network topology. This is the first study, to our knowledge, that has empirically demonstrated the complete pathway of predicted positive associations between connectivity, genetic diversity, and fitness in the context of an ecological network, and provides a framework for testing the efficacy of ecological networks for focal species using genetic tools.
Taken together with previous work in this system which showed that shepherding maintains species richness (Wagner et al. 2013) and demographic connectivity (Rico et al. 2012, 2014a) at the community level, these results suggest a positive role of ecological shepherding networks for protecting two of the three levels of biodiversity recommended for conservation by the International Union for Conservation of Nature (IUCN; McNeely et al. 1990) - species and genetic diversity – for calcareous grassland plants.

5.6 Acknowledgements
Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC CGS-D and Michael Smith Foreign Study Supplement to M.F.D, and Discovery Grant to H.H.W.) and the Government of Central Franconia, Bavaria, Germany. We thank M-J. Fortin for comments on an earlier version of this manuscript, and R. Holderegger and M.T.J. Johnson for helpful discussions. We thank K. Dadrich, D. Baumgartner, B. Raab, S. Haacke H. Lehnert for support in the field, A. Lochab, and M. Liu for help in the lab, and the shepherds E. Beil, E. Neulinger, A. Grimm for detailed information on management strategies and their implementation.
Table 5-1: Results from Mantel tests between genetic distance and geographic distance (IBD) among *P. vuglaris* populations, and genetic distance and shepherding distance (IBR), and partial Mantel tests between genetic distance and geographic distance controlling for shepherding distance (IBD|IBR), and genetic distance and shepherding distance controlling for geographic distance (IBR|IBD). Mantel $r$ and p-values are shown for tests conducted on each herd independently and for data pooled across the three herds.

<table>
<thead>
<tr>
<th>Model</th>
<th>Pooled (n=32)</th>
<th>Herd 1 (n=9)</th>
<th>Herd 2 (n=11)</th>
<th>Herd 3 (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>p-value</td>
<td>$r$</td>
<td>p-value</td>
</tr>
<tr>
<td>IBD</td>
<td>0.36</td>
<td>0.006</td>
<td>0.36</td>
<td>0.03</td>
</tr>
<tr>
<td>IBR</td>
<td>0.20</td>
<td>0.002</td>
<td>0.39</td>
<td>0.03</td>
</tr>
<tr>
<td>IBD</td>
<td>IBR</td>
<td>0.04</td>
<td>0.99</td>
<td>0.04</td>
</tr>
<tr>
<td>IBR</td>
<td>IBD</td>
<td>0.12</td>
<td>0.001</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table 5-2 Results and model fit of linear mixed effect models testing the effect of geographic connectivity ($S_{IBD}$), shepherding connectivity ($S_{IBR}$) and population size on genetic diversity ($A_r$) of *P. vulgaris* populations. All models included sheep herd (herd 1, herd 2, or herd 3) as a random effect. Model weights ($w_i$) and marginal $R^2$ of fixed effects are shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>$AIC_c$</th>
<th>$\Delta AIC$</th>
<th>$w_i$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{IBR} +$ population size class</td>
<td>20.20</td>
<td>0</td>
<td>0.726</td>
<td>0.61</td>
</tr>
<tr>
<td>$S_{IBD} + S_{IBR} +$ population size class</td>
<td>23.10</td>
<td>2.96</td>
<td>0.165</td>
<td>0.62</td>
</tr>
<tr>
<td>$S_{IBR}$ population size class</td>
<td>25.10</td>
<td>4.98</td>
<td>0.060</td>
<td>0.47</td>
</tr>
<tr>
<td>$S_{IBD} +$ population size class</td>
<td>27.60</td>
<td>7.48</td>
<td>0.017</td>
<td>0.14</td>
</tr>
<tr>
<td>$S_{IBD} +$ population size class</td>
<td>27.90</td>
<td>7.74</td>
<td>0.015</td>
<td>0.18</td>
</tr>
<tr>
<td>$S_{IBD} + S_{IBR}$</td>
<td>28.10</td>
<td>7.90</td>
<td>0.014</td>
<td>0.47</td>
</tr>
<tr>
<td>$S_{IBD}$</td>
<td>33.60</td>
<td>13.45</td>
<td>0.001</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Figure 5-1: Sampled populations and grazing treatment of *P. vulgaris* in the study region in the Franconian Alb, Germany. Grazing routes connecting consistently grazed populations are indicated by black lines and are labeled with the associated herd. Forest cover in the study regions is shown in grey.
Figure 5-2: Boxplots showing differences in genetic diversity ($A_r$) among grazing treatment of $P. vuglaris$ populations. Bold lines in boxes show the median, lower and upper sections of the boxes represent the 25$^{th}$ and 75$^{th}$, respectively, and whiskers show 1.5 interquartile distance from the median. Consistently and intermittently grazed populations had significantly higher genetic diversity than ungrazed populations (ANOVA, $F_{(2,40)}=5.49, p=0.008$).
Figure 5-3: Scatterplots showing the relationships between genetic diversity ($A_r$) and a) connectivity index based on among-population shepherding distance ($S_{IBD}$), b) connectivity index based on among-population geographic distance ($S_{IBR}$), and c) population size class (2 = 4-39 individuals, 3 = 40-99 individuals, 4 = ≥100 individuals). Solid lines show predicted values for the fixed effect ($S_{IBD}$, $S_{IBR}$, or Population Size Class) from linear mixed effect models controlling for sheep herd as a random effect. See Table 5-2 for results of model selection on these data.
Figure 5-4: Scatterplots showing the relationships between a) seed set and genetic diversity ($A_r$), and b) mean seed mass and genetic diversity. Each point represents a single individual in one of ten populations. Solid lines show predicted values for the fixed effect ($A_r$) from linear mixed effect models controlling for population as a random effect, and dotted lines represent standard error of relationships. Genetic diversity was exponentially transformed to linearize relationships.
Chapter 6

6 Multi-Scale Drivers of Contemporary Pollen Flow in the Insect-Pollinated Herb, *Pulsatilla vulgaris*

Michelle F. DiLeo, Rolf Holderegger, and Helene H. Wagner

Author contributions: MFD conducted the lab work, analysed the data, and wrote the manuscript. RH contributed to the conceptual design of the paper. HHW collected the *P. vulgaris* samples in the field and helped edit the manuscript.

6.1 Abstract

Contemporary pollen flow is commonly assumed to be a function of inter-mate distance. Yet it is increasingly recognized that geographic distance alone captures only a fraction of variation in mating success, and is a particularly poor predictor of the tail end of dispersal kernels (i.e. long-distance pollen flow events). This suggests that other factors may be at play, and that patterns of local and long-distance pollen flow cannot be modeled using a single function. Here we quantify the effects of floral resources and landscape composition on contemporary pollen flow in the insect-pollinated herb, *Pulsatilla vulgaris* over multiple spatial scales. We reconstruct realized pollen flow and pollen immigration for seven populations, and tested the hypothesis that within-population mating outcomes are related to resources measured locally, and that among-population pollen flow is related to features measured at larger spatial scales. We found that within-population pollen flow was explained by features of individual plants and patches; mean pollination distances and the proportion of selfed seeds per mother decreased as a function of floral density around maternal plants, and larger populations had significantly lower selfing rates. In contrast, among-population contemporary pollen flow (interpreted as pollen immigration rates) was related to features measured at the level of patches and surrounding landscape; pollen immigration significantly increased with population size and the amount of seminatural
landscape surrounding populations measured at a large spatial scale (1000 m radius), and significantly decreased with the amount of forest measured at an intermediate spatial scale (250 m radius). Together our results suggest that within- and among-population contemporary pollen flow may be governed by different underlying processes, possibly related to differences in foraging range of bee species that contribute to pollination at each scale.

6.2 Introduction

Understanding the spatial scale of gene flow is essential for the management of species living in fragmented landscapes. In plants, contemporary pollen flow is commonly modeled as a functional of inter-mate distance, with pollen dispersal declining exponentially within a short distance of mother plants. However, parentage-based reconstruction of realized pollen flow has uncovered pervasive dispersal over surprisingly large distances for animal-pollinated species (e.g. Robledo-Arnuncio and Gil 2005, Ahmed et al. 2009, Ismail et al. 2012). This suggests that geographic distance alone explains only a fraction of inter-individual and inter-population variation in reproductive success, and indeed often fails to sufficiently capture variation at the tail end of dispersal kernels (i.e. long-distance dispersal events; Ashley 2010, Robledo-Arnuncio et al. 2014). Thus there is a critical need for researchers to take a more pollinator-eyed view of contemporary gene flow in plants by: (1) considering additional factors that may alter attractiveness and detectability of plants to pollinators beyond the effects of distance, and, (2) recognizing that within and among-population gene flow may depend on non-overlapping sets of pollinators that respond to these features at different spatial scales.

Attributes of individual plants, patches, and local landscapes can modify pollinator behaviour and thus impact mating success and the scale of pollen flow of plant populations (Loveless and Hamrick 1984, Hegland 2014, Schueepp et al. 2014). At the level of individual
plants, scent or the size of floral displays can impact spatial patterns of contemporary pollen flow beyond the effects of distance by influencing the attractiveness of individuals to pollinators (e.g. Jakobsson et al. 2009, Dileo et al. 2014). Likewise, larger plant patches or populations may experience higher rates of pollen immigration (i.e. among-population gene flow) by attracting pollinators from further away; although the opposite can be true when pollinators must move more often to forage between small populations that have few floral resources (i.e. the fragmentation paradox; Kramer et al. 2008). Landscape context is also expected to play a critical role in determining both patterns of within-population contemporary pollen flow and rates of pollen immigration. Populations that are embedded-in, or surrounded by suitable pollinator nesting habitat may be exposed to higher abundances and diversity of pollinators and thus may achieve higher visitation rates, lower selfing, and higher seed set than populations surrounded by unsuitable habitat (Klein et al. 2003). Landscape features can additionally impact rates of pollen immigration by modifying the detectability of populations to pollinators (e.g. canopy gaps; Walters and Stiles 1996), and by providing corridors or barriers to dispersal (Kamm et al. 2010, Krewenka et al. 2011, Lander et al. 2011).

It is well known from the pollinator literature that the spatial scale at which insect pollinators perceive the landscape varies across guilds; social and large-bodied bees have larger foraging ranges and tend to respond to the landscape at much larger spatial scales than small-bodied and solitary bees (Steffan-Dewenter et al. 2002, Westphal et al. 2003, Westphal et al. 2006, Benjamin et al. 2014). For plant species that are pollinated by a variety of insects, this may imply that patterns of within- and among-population pollen flow depend on different sets of pollinators who may respond to the landscape in different ways and at different scales. This is rarely considered in plant dispersal models and may offer an explanation as to why the tail end of dispersal kernels are particularly difficult to fit (Ashley 2010).
Here we test the hypothesis that landscape effects on contemporary pollen flow are scale dependent for the generalist-pollinated herb, *Pulsatilla vulgaris*. The main pollinators of *P. vulgaris* are small-bodied and solitary or semi-social members of the bee genera *Lasioglossum*, *Osmia* and *Andrena*, although flowers are also visited by large bodied and social species such as *Bombus* and *Apis* (Kratochwil 1988, Fay and Barlow 2014). We hypothesize that local pollination patterns (e.g. selfing rates, within-population pollen flow distances) will be determined by its main small-bodied pollinators and thus will be related to landscape and floral features measured at small spatial scales. In contrast, we hypothesize that rates of pollen immigration will depend on its large-bodied pollinators and will thus be related to landscape features measured at larger spatial scales. Although seed flow is an important form of landscape-scale gene flow in this species (DiLeo et al. unpublished), virtually nothing is known about the scale of pollination. Using progeny arrays we track realized pollen flow within populations and measure contemporary pollen immigration. We ask: (1) Do within-population patterns of pollen flow follow a uniform distribution with respect to the availability of potential fathers? (2) Are measures of within-population pollination (e.g. individual variation in selfing, correlated paternity, mean outcrossing distance) explained by local factors such as floral density and maternal plant isolation? and, (3) Are population-level measures of selfing and pollen immigration explained by landscape context of populations, and if so what is the most relevant spatial scale for each response?

### 6.3 Methods

#### 6.3.1 Study Species

*Pulsatilla vulgaris* (Ranunculaceae) is a perennial, hermaphroditic herb of conservation concern and a flagship species of calcareous grasslands across central Europe. Among other calcareous grassland species in our study region in the Franconian Alb, *P. vulgaris* is the first to flower,
with flowers opening when temperatures reach above 12-15°C (usually around March-April) and lasting 4-6 weeks (Kratochwil 1988, Hensen et al. 2005). Plants normally produce 1-3 hermaphroditic, purple flowers, although up to 16 flowers per plant have been observed in our study region. *P. vulgaris* is self-compatible, but has been previously characterised as mainly outcrossing due to the absence of spontaneous self pollination in a pollinator exclusion experiment (Wells and Barling 1971). Flowers are protogynous and exhibit spatial separation of stamens and pistils (herkogamy), however separation of the male and female reproductive surfaces in both space and time are incomplete (Jonsson et al. 1991). Thus, self-fertilization is possible through both autogamy (i.e. self-fertilization occurring within the same flower) and geitonogamy (i.e. fertilization occurring between two flowers belonging to a single individual), and seed set from selfed pollen has been reported between 18-50% from crossing experiments in wild populations (Lindell 1998, Warden 2001). *P. vulgaris* flowers are pollinated by a variety of bees; predominately by members of Hymenoptera Apoidea (Kratochwil 1988). A study in south-western Germany found two species, *Lasioglossum lineare* and *Andrena bicolor* to be the most important pollinators of *P. vulgaris* (Kratochwil 1988), whereas a study in the U.K. found the most important pollinator to be *Osmia bicolor* (Fay and Barlow 2014). In our study region in south-central Germany, *O. bicolor*, *A. bicolor*, and multiple species of *Lasioglossum* (*L. calceatum, L. fulvicorne, L. pauxillum, L. morio*) and *Bombus* (*B. pascuorum, B. pratorum*) have been observed visiting flowers of *P. vulgaris* (K. Weber, personal obs.).

### 6.3.2 Study Area and Sampling

The study area is a 10 x 15 km region in the southern Franconian Alb, Germany (Fig. 6-1). The region is characterized by a series of valleys and plateaus ranging in elevation from 410-610 m (Wagner et al. 2013). The plateaus contain a mix of agricultural fields, forest, grassland and settlements. Calcareous grassland habitat is found on shallow soils on the plateaus, or more
typically on steep, eroded slopes at the margin of the plateaus and valleys. From April-May 2009, leaf material was collected from flowering *P. vulgaris* individuals (*n*=1,327) from all known populations of *P. vulgaris* in the study region (*n*=57). We recorded the number of flowers per plant and GPS coordinates were recorded to an accuracy of 1m with a handheld Trimble GeoExplorer XSeries (Trimble, Concord, ON, Canada). For paternity analysis, mature seeds were collected from 6-9 mothers from each of seven populations varying in size, isolation, and landscape context (Fig. 6-1). We collected leaf material from all potential fathers from populations where seeds were sampled, and also from all other populations in the study region with population sizes of less than 40. For populations that exceeded 40 individuals, leaf material was collected from 30-40 plants. Thus we do not have complete sampling of all potential fathers within the study region, but will be able to confidently assign fathers within populations and we assume that unassigned seeds are the product of pollen immigration.

6.3.3 Genotyping

We extracted genomic DNA from leaves and seed embryos using QIAGEN DNeasy Plant Mini Kit following the manufacturer’s protocol (QIAGEN, Mississauga, ON, Canada). Samples were amplified and genotyped at seven species-specific and highly polymorphic microsatellite markers (pv2, pv7, pv27, pv33, pv56, pv65a, pv65b) using published conditions (DiLeo et al. 2015). Markers were checked for departures from Hardy-Weinberg equilibrium and linkage equilibrium in GENEPOP 4.2 (Raymond and Rousset 1995).

6.3.4 Reconstructing Contemporary Pollen Flow

We conducted a paternity analysis for seeds using COLONY2 (Wang 2004, Wang and Santure 2009). This program reconstructs full and half-sib families using genetic information and has been found to assign a greater proportion of correct paternities than the commonly used
categorical assignment program CERVUS, especially when there is incomplete sampling of potential fathers (Walling et al. 2010). COLONY2 allows for errors in genotyping, and we set stochastic error rates individually per locus based on the number of mother-offspring genotype mismatches observed in our progeny arrays (pv2=0.01, pv7=0.05, pv27=0.01, pv33=0.05, pv56=0.03, pv65a=0.04, pv65b=0.01). To test the sensitivity of COLONY2 to assign paternities based on the pool of potential fathers included, we conducted the parentage analysis at three spatial scales. First, we conducted separate parentage analyses for each of the seven populations where we have seeds, including potential fathers only within each population. Second, we again conducted separate analyses for each population, but this time included potential fathers from all population within a 1 km radius. Third, we conducted a single analysis including all of the seven populations and potential fathers from all 57 populations within the study region. The assignments did not substantially differ between the three runs, and thus we only present results of the third analysis.

Based on the paternity assignments, we categorized each seed as either selfed, outcrossed from within the population (hereinafter “within”), outcrossed from outside the population (i.e. pollen immigrants) or unassigned. COLONY2 gives individual probabilities of assignments and we used 90% probability as our cut-off for assigned seeds. Seeds that had greater than 90% probability of being assigned to an unsampled father were assumed to be the result of a pollen immigration event. Some seeds were assigned multiple potential fathers, each with probabilities less than 90%. If the multiple potential fathers were from within the population of the sampled mother and had a joint probability of greater than 90%, the seed was categorized as “within”. Likewise, if the multiple fathers were from outside the population with a combined probability of greater than 90%, the seed was categorized as a pollen immigrant. If the multiple potential
fathers were from a combination of within versus outside, the seed was categorized as unassigned.

To evaluate the spatial scale of pollination and test if pollination occurs randomly with respect to available fathers within populations, we compared observed pollination distances to those expected under a uniform distribution using Wilcoxon rank sum tests. Only outcrossed pollination events that occurred within populations were included. Observed pollination distances were weighted by the probability of COLONY2 assignment, which allowed for information from multiple identified fathers to be incorporated. We generated expected pollination distances under a uniform distribution (i.e. all distances equally likely) by calculating geographic distance between focal mother plants and all potential fathers within each population.

6.3.5 Family-Level Analysis

We used mixed models to evaluate the role of local conspecific floral density and isolation of mother plants on mating parameters. Mating parameters of interest were calculated per mother and included: mean pollination distance of within-population outcrossed seeds, correlated paternity of outcrossed seeds ($r_p$), and the proportion of selfed seeds. Correlated paternity measures the proportion of seed pairs, sampled from a single mother plant, that share the same father (Ritland 1989), and was calculated using KINDIST in the POLDISP 1.0 software package (Robledo-Arnuncio et al. 2007). Local floral density was measured as the number of flowers within five metres of the mother plant, and mother isolation was measured as the mean distance of the mother plant to all other plants within the population (mean neighbour distance). The two predictors were tested for collinearity and were found to be moderately correlated with a variance inflation factor of 1.7 and Pearson $r$ of -0.65. For all models, we included both floral density and mean neighbour distance as fixed effects, and population as a random effect. We
used linear mixed effect models to test fixed and random effects on both mean pollination distance (square root transformed) and correlated paternity. Linear mixed models were estimated using maximum likelihood for model selection, and final models were estimated using restricted maximum likelihood (REML). We used a generalized linear mixed model (GLMM) with binomial error distribution to test the fixed and random effects on the proportion of selfed seeds per mother, and included an additional observation level random effect (OLRE) to control for overdispersion (Harrison 2014). For all models, floral density and mean neighbour distance were square root transformed to linearize relationships. We tested all combinations of predictors and chose the best model based on AICc. We assessed significance of the best model with likelihood ratio tests (LR test), and calculated variance explained by fixed effects (marginal $R^2$) using the protocol of Nakagawa and Schielzeth (2013).

We tested for a relationship between the proportion of seed set and the proportion of selfed seeds (fixed effect) with GLMM with binomial error distribution and both population level and observation level random effects. The proportion of set seeds represents the number of filled seeds divided by the total number of seeds per mother. Filled and unfilled seeds were easy to distinguish by both the size of the seed and length of the feathery style.

6.3.6 Population-Level Analysis
We used logistic regression with binomial distributed error to evaluate the effect of population size and landscape composition on the proportion of selfed and outside sired seeds per population. Population size was counted as the number of flowering plants per population. We were most interested in two types of landscape features: seminatural habitat and forest. Seminatural habitat (includes permanent grasslands, groves, mires, barren land, and orchards) provides suitable nesting and foraging habitat for wild bees, and has been shown to strongly
predict abundance and diversity of bees in European landscapes (Steffan-Dewenter et al. 2002). Thus we use the amount of seminatural habitat as a proxy for local pollinator abundance, and expect the proportion of selfed seeds to be lower, and pollen immigration to be higher, in populations surrounded by more seminatural habitat. Forest can alter the detectability of plant populations to pollinators and has also been found to act as a barrier to dispersal for bumblebees (Kreyer et al. 2004). Thus we expected the proportion of selfed seeds to be highest, and the proportion of pollen immigration to be lowest in populations surrounded by more forested landscape.

We further expected the proportion of selfed seeds and pollen immigration to be related to landscape composition at different spatial scales. The main pollinators of *P. vulgaris* are small-bodied bees, with limited dispersal ability and thus likely contribute to local pollination patterns (i.e. selfing), but not to among-population pollen flow (i.e. pollen immigration). In contrast, large bee visitors to *P. vulgaris* such as those in the genera *Bombus* and *Apis* are able to disperse long distances and likely contribute to patterns of among-population gene flow. Small-bodied and solitary bees tend to perceive the landscape at much finer spatial scales than large-bodied and social bees with large foraging ranges (Steffan-Dewenter et al. 2002, Westphal et al. 2003, Westphal et al. 2006, Knight et al. 2009, Benjamin et al. 2014). Thus we expected that the proportion of selfed seeds would correlate most strongly to landscape measured at small spatial scales and that the proportion of pollen immigration will correlate most strongly to landscape measured at large spatial scales. To explore these potential scale-dependent effects, we quantified landscape composition at five nested spatial scales, at buffers of 50, 100, 250, 500, and 1000m around population centroids. We quantified the total amount (in hectares) of forest and seminatural habitat within each buffer using digital land use maps (Tatsachehliche Nutzung aus ALKIS 2008–2009; Bayerische Vermessungsverwaltung, Munich, Germany). Seminatural
landscape included permanent grasslands, groves, mires, barren land, and orchards following the classification of Steffan-Dewenter et al. (2002).

Logistic regressions were implemented as generalized linear models (GLM) with binomial error distribution and logit-links in R 3.2.3 (R Core Development Team, 2015). We tested the effect of each predictor (population size, amount of forest, amount of seminatural habitat) on the proportion of selfed and immingrant seeds separately. Spearman rank correlations between the predictor variables population size and amount of forest ranged from 0.54-0.77, and ranged from 0.05-0.58 between population size and amount of seminatural habitat (A11, Appendix). We tested each model for overdispersion by dividing the residual deviance by the residual degrees of freedom. All models that included the proportion of selfed seeds as the response variable exhibited overdispersion, and thus we corrected standard errors using quasi-binomial models. We assessed model fit by conducting Chi-square analysis of deviance, or F-tests for models exhibiting overdispersion. We selected the best landscape scale (50m, 100m, 250m, 500m, 1000m) for each landscape predictor/response pair by choosing the model with the lowest residual deviance.

If biparental inbreeding is high within parental populations, it is possible that seeds could be incorrectly assigned as selfed when in reality the father is a close relative to the known mother. To investigate if seeds were more likely to be assigned as selfed in inbred populations, we tested for relationships between the proportions of selfed seeds and inbreeding ($F_{is}$) and observed heterozygosity ($H_o$) of the parental generation, per population using logistic regressions.
6.4 Results

6.4.1 Reconstructing Contemporary Pollen Flow

Seeds from four mothers in population A03, from one mother in A41, and from two mothers in G05a were aborted and were thus excluded from analysis. We also excluded samples that failed to amplify at more than two loci. In total, 394 seeds were genotyped and included in the paternity analysis and final sample sizes per population are shown in Table 6-1. COLONY2 assigned 382 seeds to a single father (sampled or unsampled) with greater than 90% probability, and twelve seeds were assigned multiple potential fathers with individual probabilities less than 90%. Ten of these twelve seeds were categorized as either within-population pollen flow or pollen immigration events based on the rules described in the methods. In total 201 seeds were selfed, 129 seeds were sired by fathers from within the same population as the sampled mother, 62 were sired by fathers from outside of the population of the sampled mother (i.e. pollen immigration), and two seeds were unassigned and excluded from further analysis.

Wilcoxon rank sum tests revealed significant differences between observed and expected within-population pollination distances in four populations (A03: W=283, \(p=0.018\); A21: W=11, \(p=0.03\); A25: W=339.5, \(p<0.001\); A45: W=373, \(p<0.001\)) and no significant differences in three populations (A26: W=897, \(p=0.18\); A41: W=1201, \(p=0.65\); G05a: W=370.5, \(p=0.07\); Fig. 6-2).

6.4.2 Family-Level Analysis

An initial screening of the data showed that floral density and mean neighbour distance affected mean pollination distance in opposite directions in A25 compared to the rest of the populations. Removing A25 from the linear mixed effect models for mean pollination distance did not alter the outcome of model selection, but did strengthen relationships with the predictor variables and thus we present results with A25 excluded here. The best model explaining mean pollination
distance showed that mean pollination distance significantly decreased with increasing local floral density around maternal plants ($R^2=0.51$; Likelihood ratio test, $\chi^2(1)=16.7$, $p<0.001$; Table 6-2a; Fig. 6-3a). The next best model had a ΔAICc less than two, and included a negative effect of floral density and positive effect of mean neighbour distance (Table 6-2a; Fig. 6-3c). Tests for the significance of individual fixed factors found that floral density had a significant effect on mean pollination distance (Likelihood ratio test, $\chi^2(1)=5.84$, $p=0.01$), but mean neighbour distance had a non-significant effect (Likelihood ratio test, $\chi^2(1)=1.2$, $p=0.26$). A simple linear regression including only population A25 showed that mean neighbour distance had a significant negative relationship with mean pollination distance ($F_{(1,5)}=8.64$, $p=0.03$, $R^2=0.56$). There was a non-significant positive relationship between local floral density and mean pollination distance in A25 ($F_{(1,5)}=3.97$, $p=0.1$, $R^2=0.33$).

The null model including only random population effects was selected as the best model explaining correlated paternity (Table 6-2b). The next best model had a ΔAICc of less than two and included a positive effect of mean neighbour distance, however the model was not significant and explained very little variance ($R^2=0.02$; Likelihood ratio test, $\chi^2(1)=0.75$, $p=0.38$; Table 6-2b).

The best model explaining the proportion of selfed seeds per mother showed that maternal plants surrounded by high local floral density had a significantly lower proportion of selfed seeds ($R^2=0.2$; Likelihood ratio test, $\chi^2(1)=7.1$, $p=0.007$; Table 6-2c; Fig. 6-3b). The next best model had a ΔAICc of less than two and included both a negative effect of floral density and positive effect of mean neighbour distance (Table 6-2c; Fig. 6-3d). Tests for the significance of individual fixed factors found that floral density had a significant effect on selfing (Likelihood
ratio test, $\chi^2_{(1)}=5.5, p=0.02$), but mean neighbor distance had a non-significant effect (Likelihood ratio test, $\chi^2_{(1)}=1.4, p=0.24$).

We found no significant association between the proportion of seed set and proportion of selfed seeds per mother (deviance=339.3, $R^2=0.008$; Likelihood ratio test: $\chi^2_{(1)}=1.7, p=0.19$).

6.4.3 Population-Level Analysis

The proportion of selfed seeds significantly decreased, and the proportion of pollen immigration significantly increased as a function of population size (as assessed by the number of flowering plants; Fig. 6-4; Table 6-3). Of the landscape models, forest measured at 50 m and seminatural habitat measured at 1000 m had the lowest residual deviance and were thus identified as the best predictors of the proportion of selfed seeds. However, none of the landscape predictors measured at any scale showed a significant relationship with the proportion of selfed seeds as assessed by analysis of deviance with F-tests (Table 6-3, Fig. 6-5). The proportion of pollen immigration significantly decreased with increasing forest surrounding populations, with the strongest association with forest measured within a radius of 250 m (Table 6-3, Fig. 6-5c). Pollen immigration significantly increased with increasing seminatural habitat surrounding populations, with the strongest association measured within a radius of 1000 m (Table 6-3, Fig. 6-5d).

The proportion of selfed seeds did not show a significant relationship with either inbreeding ($F_{is}$) or observed heterozygosity ($H_o$) of the parental generation, indicating that seeds were not more likely to be assigned as selfed in inbred populations (GLM, $F_{is}$: Deviance=57.3, Likelihood ratio test $F_{(1)}=0.87, p=0.43$; $H_o$: Deviance=65.2, Likelihood ratio test $F_{(1)}=0.87, p=0.39$).
6.5 Discussion

Here we quantified the multi-scale determinants of contemporary pollen flow in *Pulsatilla vulgaris* across seven populations varying in size and landscape context. Overall, we found that patterns of within-population pollen flow were related to floral resources measured at the scale of individuals and patches, whereas among-population pollen flow (i.e. pollen immigration rates) could be explained by features measured at the patch and landscape-scale. Within-population mean pollination distance and the proportion of selfed seeds per mother decreased with increasing floral density around maternal plants. Likewise, large populations had significantly lower overall proportions of selfed seeds compared to small populations. In contrast, pollen immigration showed a significant positive relationship with population size, and was related to landscape composition measured at intermediate and large spatial scales. Together our results suggest that within- and among-population patterns of contemporary pollen flow may be governed by different underlying processes and highlights the importance of using multi-scale approaches when modeling dispersal in generalist-pollinated plants.

6.5.1 Pollen Flow as a Multi-Scale Process

We found that attributes of individual floral neighbourhoods, patches, and local landscape all played a role in predicting aspects of contemporary pollination in *P. vulgaris*. Patterns of within-population pollination (i.e. selfing rates, within-population pollination distances) could be explained by local and patch-scale factors, whereas pollen immigration was explained by factors measured at the patch and landscape scale. We think a likely explanation for these scale-dependent effects is that different sets of pollinators are responsible for within- versus among-population contemporary pollen flow. The main pollinators of *P. vulgaris* are small-bodied bees (4-10 mm; Meyer 2007) with restricted dispersal ability (180-600 m maximum foraging range estimated for bees of similar size; Gathmann and Tscharntke 2002, Zurbuchen et al. 2010), and
are therefore likely to contribute to patterns of local pollination, but unlikely to move pollen between populations. In contrast, less frequent large-bodied visitors of *P. vulgaris* such as *Bombus pascuorum* and *Apis mellifera* have estimated foraging ranges well into the thousands of metres (Westphal et al. 2006, Knight et al. 2009, Zurbuchen et al. 2010), and are thus likely to be more important for the maintenance of among-population pollen flow. This scale-of-effect based on foraging range and sociality has been demonstrated a number of times in the pollination literature, with small and/or solitary bees perceiving landscape at relatively small spatial scales compared to large social bees (Steffan-Dewenter et al. 2002, Westphal et al. 2006, Benjamin et al. 2014).

6.5.2 Landscape Effects on Contemporary Pollen Flow

Our results further suggest that pollinators contributing to local and among-population pollination may respond to different environmental cues; we found that selfing rates were related to availability of floral resources (e.g. population size), whereas pollen immigration related to both floral resources and landscape context. Landscape can influence contemporary pollination in a variety of ways. For example, seminatural habitat provides important nesting sites for bees in European landscapes and is a good predictor of the richness and abundance of bee communities (Steffan-Dewenter et al. 2002, Hopfenmueller et al. 2014). In addition to providing suitable habitat, landscape can affect immigration rates by altering the detectability of populations to pollinators, or by promoting or restricting pollinator movement (Kamm et al. 2010, Lander et al. 2011). Our data suggest any one of these scenarios is possible for *P. vulgaris*, and they are not necessarily mutually exclusive. For example, the positive relationship that we found between pollen immigration rates and amount of seminatural habitat (Fig. 6-5d) may suggest that *P. vulgaris* populations surrounded by more bee habitat simply encounter more pollinators. Alternatively, seminatural habitat tends to be open, and populations surrounded by
more open landscapes may be more easily detectable. Kamm et al. (2010) found that open areas
promoted contemporary pollen flow among individuals of the insect-pollinated forest tree Sorbus
domestica in Switzerland, suggesting that open landscape can additionally act to enhance
dispersal of pollinators. Closed canopies, on the other hand, may make populations harder to find
by overhead bees or act as a barrier to dispersal (Kreyer et al. 2004), which fits well with our
observed result of decreased pollen immigration in populations surrounded by more forest (Fig.
6-5c). However, it is important to note that the rank correlation between forest cover within a
radius of 250 m and the number of flowering individual was quite high (\( \rho = -0.77 \); Appendix). Thus there is potential confound between forest cover and population size and
additional populations are required to test the independent contribution of each predictor to
patterns of pollen immigration.

In contrast to our findings on pollen immigration, we were surprised to find no
relationship between the amount of seminatural habitat and the proportion of selfed seeds per
population. This suggest that seminatural habitat is not a good proxy for the abundance of P.
vulgaris' main small-bodied pollinators, and that they are instead limited by the amount of floral
resources in the landscape. Alternatively our categorical definition of seminatural habitat may
have been too simplistic to capture the true suitability of habitat for nesting of bees.
Microclimate, habitat shape and edge, and habitat quality can alter the suitability of nesting sites
and thus may be better predictors of local abundance of small, solitary bees (Williams and
Kremen 2007, Hopfenmueller et al. 2014, Somme et al. 2014). One study in the U.K. found that
seminatural habitat showed a positive relationship with bee abundance, but only when included
as an interaction with floral resources, suggesting that both nesting and foraging resources
together are important (Nayak et al. 2015). Our sample size of seven populations precluded any
exploration into additive or interactive effects, so it is possible that we overlooked an important
complexity in the landscape drivers of local pollination. Nonetheless, the availability and density of local floral resources appear to be of high importance to explaining pollination success (i.e. outcrossing) at both the level of individuals and populations.

6.5.3 Consequences for P. vulgaris Populations

We found surprisingly high selfing rates across populations, with 51% of all genotyped seeds assigned as selfed, and up to 85% in population A21 (Table 6-1). This is despite the fact that P. vulgaris has potential adaptations to promote outcrossing through the spatial and temporal separation (although incomplete) of male and female parts of hermaphroditic flowers (i.e. herkogamy and dichogamy; Jonsson et al. 1991). Such high selfing indicates that populations may be pollen limited, particularly so in small populations. Although we did not find a significant relationship between seed set and selfing rates, maternal plants that had fewer selfed seeds tended to set a higher proportion of seeds (A12, Appendix). This provides some evidence that pollen limitation might be occurring at least at the level of individual plants.

When outcrossing did occur, the scale of within-population pollen flow was limited with most pollination events occurring within 10 m of the maternal plant (Fig. 6-2; A13, Appendix). At the level of individual plants, mothers with high local floral density experienced lower selfing but this came at the price of reduced outcrossed pollination distances (Fig. 6-3). Thus beyond selfing, there is an additional risk of biparental inbreeding if near-neighbours are close relatives. The fitness consequences of this potential high risk of inbreeding in the offspring generation are unclear. One experimental crossing study showed that selfed seeds had lower germination rates than outcrossed seeds in a single population of P. vulgaris in Sweden (Lindell 1998). Previous work in our study region showed that populations of P. vulgaris with low genetic diversity produced fewer and lighter seeds (DiLeo et al. unpublished). However, the parental generation
studied here showed very little inbreeding (\(F_{is}\) from -0.08-0.09), suggesting that the fitness consequences of selfing and neighbour mating has not been transmitted to the level of populations. \(P. vulgaris\) has low recruitment in the wild and is fairly long-lived (Fay and Barlow 2014), which indicates that any genetic changes to populations will occur relatively slowly. Thus the high selfing and limited gene flow we found here could represent a preview of changes to come. But as a perennial species, \(P. vulgaris\) may also be less vulnerable to the negative effects of pollen limitation, with selfing providing reproductive assurance in years when pollinators are scarce (Ashman et al. 2004, Knight et al. 2005).

Around 16% of all seeds were assigned as pollen immigration events, with a range of 8-30% immigrants per population (Table 6-1). Even if relatively infrequent, long-distance dispersal can have large impacts on the genetic structure of populations (Nathan 2006), and thus could be the key to the maintenance of genetic diversity despite high selfing in \(P. vulgaris\) populations. However, we were unable to resolve the true scale of pollen-mediated gene flow, since incomplete sampling of potential sires precluded the reconstruction of realized pollen flow trajectories among populations. Previous work in this system showed that seed-mediated gene flow via an ecological shepherding network contributes significantly to the demographic and genetic connectivity of \(P. vulgaris\) populations at the landscape scale (Rico et al. 2012; DiLeo et al. unpublished). Thus future work is required to both quantify the true scale of pollen flow and disentangle the relative contribution of pollen and seed to genetic connectivity in this system.

6.5.4 Conclusions
Here we found that features of individual floral neighbourhoods, patches, and local landscape contribute to contemporary pollen flow in \(Pulsatilla vulgaris\). The availability of floral resources played a key role in explaining pollination success both within and among-populations and
landscape context was additionally found to be an important predictor of pollen immigration. Overall, small populations had the highest proportion of selfed seeds and lowest pollen immigration rates, suggesting that they are at greatest risk of inbreeding. Further work is required to resolve the scale of among-population pollen flow and determine the relative contribution of floral resources versus landscape context to pollen immigration rates. Our results highlight the importance of considering drivers of pollination at multiple spatial scales for generalist-pollinated species. Given that most studies of contemporary pollen flow focus on trees, our work provides important insight into pollen-mediated gene flow in an herbaceous species.
Table 6-1: Sample sizes for the seven *P. vulgaris* populations where paternity analysis was conducted. All flowering individuals were genotyped within the seven populations. Population size (number of flowering plants), and the number of seeds sampled across mothers and flowers per population are shown. The number of seeds assigned as selfed and pollen immigrants based on COLONY2 paternity analysis are shown.

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of Flowering Plants</th>
<th>Sample Sizes</th>
<th>Number Selfed</th>
<th>Number Immigrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mothers</td>
<td>Flowers</td>
<td>Seeds</td>
</tr>
<tr>
<td>A03</td>
<td>44</td>
<td>2</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>A21</td>
<td>21</td>
<td>6</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>A25</td>
<td>57</td>
<td>8</td>
<td>8</td>
<td>79</td>
</tr>
<tr>
<td>A26</td>
<td>22</td>
<td>7</td>
<td>7</td>
<td>60</td>
</tr>
<tr>
<td>A41</td>
<td>15</td>
<td>9</td>
<td>9</td>
<td>78</td>
</tr>
<tr>
<td>A45</td>
<td>22</td>
<td>6</td>
<td>6</td>
<td>55</td>
</tr>
<tr>
<td>G05a</td>
<td>46</td>
<td>6</td>
<td>6</td>
<td>31</td>
</tr>
</tbody>
</table>
Table 6-2: Results of model selection for linear mixed models estimating the relationship of mean pollination distance per mother plant with flower density around maternal plants and mean distance of maternal plants to all potential sires within the population (mean neighbour distance) (a), linear mixed models estimating the relationship of correlated paternity with flower density and mean neighbour distance (b), and generalized mixed models estimating the relationship of the proportion of selfed seeds per maternal plant with flower density and mean neighbour distance (c). Population was included as random effect in all models and an observation-level random effect was additionally included in models on the proportion of selfed seeds. Model weights ($w_i$) and marginal $R^2$ of fixed effects are shown.

<table>
<thead>
<tr>
<th>Response</th>
<th>Predictor(s)</th>
<th>Deviance</th>
<th>AICc</th>
<th>ΔAIC</th>
<th>$w_i$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) mean pollination distance</td>
<td>~ flower density</td>
<td>-</td>
<td>50.5</td>
<td>0</td>
<td>0.673</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance + flower density</td>
<td>-</td>
<td>52.5</td>
<td>1.9</td>
<td>0.259</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance</td>
<td>-</td>
<td>55.1</td>
<td>4.6</td>
<td>0.068</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>~ (1</td>
<td>Population)</td>
<td>-</td>
<td>64.4</td>
<td>13.9</td>
<td>0.001</td>
</tr>
<tr>
<td>b) correlated paternity</td>
<td>~ (1</td>
<td>Population)</td>
<td>-</td>
<td>40.9</td>
<td>0</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance</td>
<td>-</td>
<td>42.7</td>
<td>1.8</td>
<td>0.224</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>~ flower density</td>
<td>-</td>
<td>43.5</td>
<td>2.6</td>
<td>0.154</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance + flower density</td>
<td>-</td>
<td>45.5</td>
<td>4.6</td>
<td>0.057</td>
<td>0.02</td>
</tr>
<tr>
<td>c) proportion selfed</td>
<td>~ flower density</td>
<td>188.5</td>
<td>197.4</td>
<td>0</td>
<td>0.558</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance + flower density</td>
<td>187.1</td>
<td>198.5</td>
<td>1.1</td>
<td>0.319</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>~ mean neighbour distance</td>
<td>192.6</td>
<td>201.6</td>
<td>4.2</td>
<td>0.070</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>~ (1</td>
<td>Population) + (1</td>
<td>OLRE)</td>
<td>195.6</td>
<td>202.1</td>
<td>4.7</td>
</tr>
</tbody>
</table>
Table 6-3: Estimates for population-level logistic regressions, showing the relationship of the proportion of selfed seeds or proportion of immigrants per population with population size (number of flowering plants) and landscape composition (forest and seminatural habitat) measured at increasing radii from population centroids. Standard errors of model estimates are given (SE). The best fitting model for each predictor type was selected as the model with the lowest residual deviance. Model significance was assessed with analysis of deviance with F-tests (proportion selfed) or Chi-square tests (proportion immigrant) and significant models are marked with asterisks.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Proportion Selfed</th>
<th>Proportion Immigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>Deviance</td>
</tr>
<tr>
<td>Number of Flowering Plants</td>
<td>-0.04 (0.02)</td>
<td>30.14*</td>
</tr>
<tr>
<td></td>
<td>0.02 (0.008)</td>
<td>4.33**</td>
</tr>
</tbody>
</table>

Forest Measured at Scale:
- 50 m: 5.35 (2.85) 38.7 -2.33 (1.42) 9.23
- 100 m: 0.81 (0.52) 43.8 -0.61 (0.25) 5.58*
- 250 m: 0.08 (0.07) 51.7 -0.08 (0.03) 5.17**
- 500 m: 0.02 (0.02) 55.7 -0.02 (0.01) 5.29**
- 1000 m: 0.005 (0.01) 55.9 -0.01 (0.002) 6.98*

Seminatural Habitat Measured at Scale:
- 50 m: -1.10 (4.64) 64.5 -0.95 (1.87) 11.7
- 100 m: -0.44 (1.12) 63.7 0.50 (0.50) 10.9
- 250 m: 0.07 (0.16) 62.9 0.05 (0.06) 11.3
- 500 m: 0.002 (0.05) 65.2 0.03 (0.02) 9.1
- 1000 m: -0.01 (0.01) 52.0 0.01 (0.005) 5.9*

* P < 0.05; ** P < 0.01
Figure 6-1: Study area in the Franconian Alb, Germany, showing main land classification and all known *Pulsatilla vulgaris* populations in the region. Populations where seeds were sampled and paternity analysis was conducted are labeled and represented by filled black squares. Land cover classified as “other” includes settlements and agriculture.
Figure 6-2: Histograms showing observed (grey bars) and expected (white bars) within-population pollination distances. Density represents the relative frequency of observed or expected pollination distances, with the total height of bars for each category summing to one. Expected distances represent expectations under random mating, where mating with all potential sires, regardless of distance from maternal plant, is possible. Observed pollination distances were reconstructed with paternity analysis. Populations with asterisks showed significant differences in observed versus expected pollination with Wilcoxon rank sum tests.
Figure 6-3: Scatterplots showing the relationships of mean within-population pollination distance per mother and local floral density (a), mean pollination distance and mean neighbour distance (c), proportion of selfed seeds per mother and local floral density (b), and proportion of selfed seeds and mean neighbour distance (d). Each point represents a single maternal plant. Predictors and mean pollination distance were square root transformed to linearize relationships. Solid lines are predicted values of fixed effects from linear mixed models controlling for Population as random effect (a,c), and predicted values of fixed effects from generalized mixed models including Population and observation level random effects (b,d). See Table 6-2 for results of model selection on these data.
Figure 6-4: Scatterplot showing the relationship of population size (number of flowering plants) with the proportion of selfed seeds per population and the proportion of immigrants per population. Solid lines represent fitted values from logistic regression.
Figure 6-5: Scatterplots showing the relationship of the proportion selfed seeds per population with the amount of forest measured within in a 50 m radius (a) and seminatural habitat measured within a 1000 m radius (b), and the relationship of the proportion of immigrants per population with forest measured within 250 m (c) and seminatural habitat measured within 1000 m (d). The best-supported scale of each landscape predictor from logistic regression is shown (see Table 6-3 for model selection). Solid lines show predicted values of significant logistic regression models.
Chapter 7

7 Synthesis

7.1 Summary

My thesis quantifies the effects of landscape structure, individual plant characteristics, and directed seed dispersal on gene flow in plants, and tests if these features explain variation in genetic differentiation beyond the effects of isolation-by-distance. Using simulations, I found that inter-individual variation in attractiveness to pollinators explained significantly more variation in patterns of contemporary pollen flow than inter-mate distance. In an empirical study system of the specialist herb *Pulsatilla vulgaris* in the Franconian Alb, I found that connectivity within an ecological shepherding network explained more variation in genetic differentiation among populations than geographic distance, suggesting that sheep-mediated seed dispersal is an important component of genetic connectivity. In a subset of seven populations in the same system, I reconstructed realized pollen flow and found high rates of selfing, and that within-population pollination occurred at small spatial scales (<10 m from mother plant). I found that within-population variation in pollen flow (i.e. selfing rates) correlated with floral resources measured at the individual and patch scale, whereas among-population variation in pollen flow (i.e. pollen immigration rates) correlated with floral resources measured at the patch scale and landscape context measured at intermediate and large spatial scales. This suggests that within- and among-population contemporary pollen flow are governed by different underlying processes, possibly related to differences in foraging range of bee species that contribute to pollination at each scale. Together, my thesis highlights the importance of using multi-scale and multi-vector approaches when modeling gene flow in plants, and suggests that typical models based on distance alone may be insufficient to capture the complexity of pollination and seed dispersal.
7.2 Beyond Isolation-by-Distance

Gene flow among individuals and populations is commonly modeled as a function of geographic distance (i.e. isolation-by-distance; IBD; Wright 1943). However, it is increasingly recognized that geographic distance alone explains only a fraction of variation in inter-individual and inter-population gene flow in plants (McRae and Beier 2007, Ashley 2010, Robledo-Arnuncio et al. 2014). The relative permeability of the intervening landscape has been shown to influence gene flow beyond the effects of distance for a number of species, however this has mainly been tested in animals (Holderegger et al. 2010, Storfer et al. 2010). In a review of papers that tested landscape genetic hypotheses between 1998-2008, Storfer et al. (2010) found that only 14.5% were conducted in plants. Chapter 2 of my thesis confirms that this bias still exists today, where plants were found to be over-represented compared to other taxa in the group of papers testing hypotheses related to IBD, but under-represented in the group testing hypotheses related to landscape resistance (i.e. isolation-by-resistance, IBR; Fig. 2-2). My thesis helps to address this gap by testing the effects of landscape context on contemporary pollen flow in *P. vulgaris* (chapter 6), and the relative importance of intervening landscape, geographic distance, and at-site variables (i.e. individual plant characteristics) for the understory tree *Cornus florida* (chapter 3).

My literature review (chapter 2) uncovered another important gap in our current knowledge of how landscape affects genetic structure: most studies (mostly animal) focus on the effects of the intervening landscape on gene flow, but few consider the role of local habitat amount (i.e. suitable habitat surrounding populations in node-based studies, chapter 2). Habitat amount is a strong predictor of abundance and species richness across a number of taxonomic groups (Thornton et al. 2011, Fahrig 2013), including insect pollinators (Steffan-Dewenter et al. 2002). For animal-pollinated plants, the amount of suitable pollinator habitat surrounding populations has been shown to increase both visitation rates and seed set (Hadley et al. 2014,
Somme et al. 2014). However, because the source of pollen is unknown (i.e. selfed, outcrossed within populations, pollen immigration), these types of studies are unable to differentiate if the effects on plant fitness are due to pollen quantity limitation or pollen quality limitation (Aizen and Harder 2007). This is a missing link where parentage-based reconstruction of contemporary pollen flow can provide insight into the landscape determinants of within-versus among-population pollen flow and the consequences for genetic diversity and fitness of plant populations. Even at low rates (e.g. ~16% in *P. vulgaris* populations; chapter 6), pollen immigration can be an important source of diversity with strong implications for plant fitness (Breed et al. 2012), and these effects may go undetected in measures of seed set. Importantly, I showed that selfing and pollen immigration rates of *P. vulgaris* were correlated with different aspects of landscape structure, demonstrating the importance of differentiating between pollen sources.

Beyond landscape context, my thesis demonstrates the importance of including variation in individual plant traits in genetic models. The influence of floral traits on pollinator visitation and plant fitness have been well characterized (Barrett et al. 1994, Barrett and Harder 1996, Ishii et al. 2008, Mitchell et al. 2009), however they are rarely considered in models of contemporary pollen flow (Ashley 2010). Using simulations, the results of chapter 3 showed that inter-individual variation in attractiveness explained significantly more variation in genetic differentiation among pollen pools than geographic distance. Interestingly, the simulations showed that the relationship between pollen-pool differentiation and variation in inter-individual attractiveness was stronger in high density compared to low-density systems (Fig. 3-2), suggesting that these factors are particularly important to consider in herbaceous species. In chapter 6, I showed that this is true for at least one herb, *P. vulgaris*; I found that both selfing
rates and mean pollination distances of maternal plants showed a stronger relationship with local floral density than mean inter-mate distance.

7.3 The Importance of Considering Multiple Vectors

Most studies quantify gene flow in plants from the perspective of either pollen or seed, but not both, and typically, the focus is on pollen as it is assumed to contribute more strongly to genetic connectivity compared to seed flow (Petit et al. 2005). My thesis contributes to current knowledge by quantifying the landscape and ecological determinants of both pollen- and seed-mediated gene flow. Importantly, I showed that seed flow, which is often ignored, contributes substantially to genetic connectivity at the landscape scale, and may make up for the high rates of self-pollination and the restricted scale of within-population outcrossing in *P. vulgaris*.

Previous work in Europe has shown that sheep can act as a vector for long-distance seed dispersal for a variety of plant species (Fischer et al. 1996), and that grazing networks can maintain demographic connectivity for calcareous grassland plants (Auffret et al. 2012, Rico et al. 2012, Wagner et al. 2013). Rico et al. (2014b) further demonstrated that the shepherding network in the Franconian Alb maintains genetic connectivity for another wildflower, *Dianthus carthusianorum*. My thesis builds on this by showing that the benefits of enhanced connectivity provided by the shepherding network percolate up to the level of population fitness, and thus may have important impacts on the long-term persistence of populations.

It is now recognized that many, if not most, plant species interact with multiple vectors for both pollen and seed flow (Higgins et al. 2003). Vectors may have different foraging ranges and thus may contribute to dispersal at different spatial scales, and respond to landscape in different ways. This means that the typical dispersal kernel approach to modeling the decay of contemporary pollen flow or seed flow with distance from maternal plants may be inappropriate
if different sets of pollinators or seed dispersers contribute to local versus long-distance dispersal events. One study on the wind-pollinated tree *Populus trichocarpa* showed that modeling contemporary pollination as a two-component process (i.e. local pollination versus updrafts that move pollen long distances) significantly improved inference (Slavov et al. 2009). To my knowledge this two-component process has not been modeled in a plant-pollinator system. My results from chapter 6 suggest that a two-component process of pollen flow may exists for *P. vulgaris*, and is likely explained by the differences in foraging range by small-bodied bees responsible for within-population pollination (i.e. selfing rates) and large-bodied bees responsible for among-population pollination (i.e. outcrossing rates). This highlights the importance of using a multi-scale approach, and will be particularly relevant for plant species without specialized pollinators and who are visited by both small and large-bodied bees. Chapter 6 further highlights the importance of considering rare pollinators. Large bodied bees are thought to contribute very little to total pollen transfer in *P. vulgaris* (< 15% of all flower visitors; Kratochwil 1988), yet may be responsible for all among-population pollen flow events.

### 7.4 Conservation Implications for *P. vulgaris*

*Pulsatilla vulgaris* is a flagship species of high conservation concern in Europe, and has witnessed rapid decline and local extirpation across its range (1988). In the U.K., where country-wide data is available, only 33 of the 130 known populations of *P. vulgaris* survived from 1750 to 1960s (Wells 1968, Hensen et al. 2005, Walker and Pinches 2011, Fay and Barlow 2014), and an additional 16 populations were lost between 1968 and 2006 (Wells 1968). Similar declines have been reported for populations elsewhere in Europe, including Germany (IUCN 2016; Hensen et al. 2005). Thus there is a critical need to protect and effectively manage remaining populations. Low-intensity winter grazing has been suggested as a management option to reduce above ground competition and increase recruitment and population size of *P. vulgaris*.
populations in the U.K. (Walker and Pinches 2011). However, my results from chapter 5 suggest that rotational summer grazing during the period of seed shed might be a more suitable conservation measure for *P. vulgaris* as it provides the additional benefit of directed dispersal of seeds among grazed populations. Importantly, I found evidence that the benefits of the shepherding network translate to the level of diversity and fitness of populations: populations that were well connected within the network had higher genetic diversity than those that were ungrazed or more isolated within the network, and populations with higher diversity produced more and heavier seeds.

Preserving seed-mediated gene flow in *P. vulgaris* might be especially important given that contemporary pollination in the Franconian Alb seemed to occur mainly through selfing, and that when outcrossing did occur within populations it was mostly at limited spatial scales (<10 m from mother plant; chapter 6). On the other hand, even low levels of pollen immigration like the ones I found in chapter 6 (16% of all pollination events) may have important impacts on the genetic structure of plant populations (Walker and Pinches 2011). For ungrazed populations where seed-mediated gene flow is unlikely to contribute to genetic connectivity, it will be important to see if these low levels of pollen immigration are enough to maintain genetic diversity and thus long-term viability of populations. In the same study region, Rico (2012) showed that ungrazed populations of another grassland specialist, *Dianthus carthusianorum*, had significantly higher spatial genetic structure (i.e. relatedness among neighbours) than grazed populations. This suggests that the consequences of restricted within-population mating may be higher in ungrazed population; if near neighbours are more likely to be close relatives in ungrazed versus grazed populations, restricted pollen flow will lead to higher levels of inbreeding.
7.5 Limitations and Future Directions

An important limitation of this work on *P. vulgaris* is that it is observational and thus causation cannot be inferred. The field study provided biological realism with practical management implications, but not all potential confounding variables could be controlled for. For example, local environmental or site condition of sampled patches may influence plant vigour and flowering phenology, which in turn could impact spatial patterns of contemporary pollen flow (Lennartsson et al. 2012, Suni and Whiteley 2015). In addition, local environment can alter maternal plant resource allocation (i.e. maternal effects; Roach and Wulff 1987, Weiner et al. 1997), leading to variation in seed set and/or seed weight that does not have a genetic basis. Future work using controlled crosses or a common-garden is required to resolve the fitness consequences of reduced seed- and pollen-mediated gene flow for *P. vulgaris* populations.

Because I was unable to quantify the scale of among-population pollination for *P. vulgaris*, further work is required to resolve the contribution of pollen-mediated gene flow to connectivity at the landscape scale. Only with this information can we determine the *relative* importance of pollen- versus seed-mediated gene flow to the maintenance of genetic diversity and fitness of populations. Ideally, future work will incorporate adaptive genetic loci that show stronger relationships with fitness traits, and which would additionally provide insight into the balance of local adaptation and gene flow in this system.
References


atronitens (Aves: Pipridae), a white sand campina specialist, to understand Pleistocene landscape evolution in Amazonia. Biological Journal of the Linnean Society 110:60-76.


Herrera, C. M. 1995. Microclimate and individual variation in pollinators - flowering plants are more than their flowers. Ecology 76:1516-1524.


Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103-106.


Appendices

A 1: Boxplots showing correlation coefficients of gravity models estimated for *Cornus florida* as a function of bandwidth used to measure between-site ecological variables (Chapter 3). Measuring ecological variables within a 1 m buffer around lines connecting *C. florida* maternal plants gave the strongest fit and was used to estimate final gravity models. See Table 3-1 for description of between-site ecological variables measured.
A 2: Boxplots showing the distribution of variance explained, $R^2$, for gravity models fit to simulated data based on saturated and pruned (by Pollination Graph) graph topologies as function of simulated pollen donor density (Chapter 3). Both the effects of pollen donor density and graph topology significantly influenced variance explained.
A 3: Locations of ten populations where seeds were collected to measure fitness-related traits in Chapter 5. All calcareous grassland patches within the study region are shown and black lines represent the three non-overlapping shepherding routes that make up the ecological network. See Fig. 5-1 for grazing treatment of each patch and location of *P. vulgaris* populations.

Legend

- seeds collected
- seeds not collected
- Forest Cover
A 4: Supplementary Methods: The effect of connectivity and population size on genetic diversity ($A_r$) of *P. vulgaris* populations, with population size as a continuous variable (Chapter 5).

We tested the effects of including population size as a continuous variable in linear mixed effect models on genetic diversity. We had information on census population size for only twelve populations that had fewer than 40 individuals or where we collected fitness data. The remaining fourteen populations were placed in one of two categories of estimated population size: 40-99 individuals, and ≥100. We formatted population size as a continuous variable by using census population size for those twelve populations where this information was available, and gave a size of 75 and 100 for populations with estimated size of 40-99 and ≥100, respectively. As in the main text, we tested all possible sub-models of $A_r \sim S_{IBD} + S_{IBR} + population\ size$ and used AICc for model selection, but this time we treated population size as a continuous rather than categorical variable. As in the original analysis in the main text, $S_{IBR}$ and *population size* were selected as the most important predictors of genetic diversity (see A5-6, Appendix).
A 5: Results of linear mixed effect models testing the effect of connectivity based on geographic distance among populations ($S_{IBD}$), connectivity based on shepherding distance among populations ($S_{IBR}$), and population size (continuous) on genetic diversity ($A_r$) of *P. vulgaris* populations (Chapter 5). Models weights ($w_i$) are shown. Model selection based on $\text{AIC}_c$, shows that the same model was chosen regardless of whether population size was treated as a categorical or continuous variable (see Table 5-2).

<table>
<thead>
<tr>
<th>Model</th>
<th>$\text{AIC}_c$</th>
<th>$\Delta\text{AIC}$</th>
<th>$w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{IBR} + \text{population size}$</td>
<td>16.4</td>
<td>0</td>
<td>0.68</td>
</tr>
<tr>
<td>$S_{IBD} + S_{IBR} + \text{population size}$</td>
<td>18.2</td>
<td>1.87</td>
<td>0.27</td>
</tr>
<tr>
<td>$S_{IBR}$</td>
<td>22.4</td>
<td>5.99</td>
<td>0.034</td>
</tr>
<tr>
<td>$S_{IBD} + S_{IBR}$</td>
<td>25.0</td>
<td>8.65</td>
<td>0.009</td>
</tr>
<tr>
<td>$S_{IBD} + \text{population size}$</td>
<td>36.3</td>
<td>19.88</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$S_{IBD}$</td>
<td>36.8</td>
<td>20.38</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$\text{population size}$</td>
<td>37.6</td>
<td>21.24</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>
A 6: Scatterplot showing the relationship between population size and genetic diversity ($A_r$) across *P. vulgaris* populations (Chapter 5). Census population sizes were known for populations where seeds were collected and for those with fewer than 40 individuals. For populations over 40 individuals where census size was unknown, we set population size to be 100.
A 7: Sequentially bonferroni-corrected *p-values* for Hardy-Weinberg exact tests conducted within populations for each microsatellite marker (Chapter 5).

<table>
<thead>
<tr>
<th>Population</th>
<th>pv2</th>
<th>pv7</th>
<th>pv27</th>
<th>pv33</th>
<th>pv56</th>
<th>pv65a</th>
<th>pv65b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td>A04</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.04</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A05</td>
<td>1</td>
<td>0.09</td>
<td>0.24</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A06</td>
<td>1</td>
<td>0.06</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A13</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>0.09</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A21</td>
<td>1</td>
<td>0.77</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A23</td>
<td>1</td>
<td>0.12</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.83</td>
<td>1</td>
</tr>
<tr>
<td>A24</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.09</td>
<td>1</td>
</tr>
<tr>
<td>A25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.08</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A29</td>
<td>1</td>
<td>0.96</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A31</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A34</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>0.43</td>
<td>1</td>
</tr>
<tr>
<td>A36</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A37</td>
<td>1</td>
<td>1</td>
<td>0.91</td>
<td>1</td>
<td>0.94</td>
<td>0.22</td>
<td>1</td>
</tr>
<tr>
<td>A38</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A41</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>1</td>
<td>0.34</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A44</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A45</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>0.92</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A49</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.92</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.63</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C04</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.33</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C05</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.04</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E02</td>
<td>0.09</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G01</td>
<td>1</td>
<td>1</td>
<td>0.44</td>
<td>1</td>
<td>0.08</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G05a</td>
<td>0.43</td>
<td>0.01</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td>G100</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>G16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.46</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G30</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G31</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.03</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G37</td>
<td>1</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.42</td>
<td>1</td>
</tr>
<tr>
<td>G40</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>G47</td>
<td>0.38</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.11</td>
<td>1</td>
</tr>
<tr>
<td>G48</td>
<td>1</td>
<td>1</td>
<td>0.48</td>
<td>0.45</td>
<td>0.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G49</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G50</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G51</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GZ</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.06</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td>N03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>N05</td>
<td>1</td>
<td>&lt;0.001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>N07</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.68</td>
<td>1</td>
</tr>
<tr>
<td>N08</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.19</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>N09</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.69</td>
<td>1</td>
</tr>
</tbody>
</table>
A 8: Boxplots of alternate measure of genetic diversity, effective number of alleles (a), expected heterozygosity (b), and observed heterozygosity (c), across grazing treatments for *P. vulgaris* populations (Chapter 5). Measures of genetic diversity were averaged across seven microsatellite markers and individuals within populations. The effective number of alleles (Ae) was corrected for differences in population size using rarefaction.
A 9: Results of linear mixed effect models testing the effect of connectivity based on geographic distance among populations ($S_{IBD}$), connectivity based on shepherding distance among populations ($S_{IBR}$), and population size on three alternate measures of genetic diversity for *P. vulgaris* populations (Chapter 5). All models included sheep herd as a random effect. Models weights ($w_i$) and marginal $\text{R}^2$ of fixed effects are shown. Model selection based on AICc, shows that the same model was chosen for $A_e$, $H_e$, and $A_r$ (see Table 5-2). A different model was chosen for $H_o$, however, all models with $H_o$ as the response variable had poor fit as assessed by marginal $\text{R}^2$. Abbreviations: $A_e$, effective number of alleles; $H_e$, expected heterozygosity; $H_o$, observed heterozygosity.

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Model</th>
<th>AICc</th>
<th>ΔAIC</th>
<th>$w_i$</th>
<th>$\text{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_e$</td>
<td>$S_{IBR} + \text{population size class}$</td>
<td>13.10</td>
<td>0</td>
<td>0.59</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR} + \text{population size class}$</td>
<td>15.50</td>
<td>2.45</td>
<td>0.17</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>$S_{IBR}$</td>
<td>15.70</td>
<td>2.62</td>
<td>0.16</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR}$</td>
<td>18.50</td>
<td>5.2</td>
<td>0.04</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + \text{population size class}$</td>
<td>20.30</td>
<td>7.27</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>$\text{population size class}$</td>
<td>20.50</td>
<td>7.45</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD}$</td>
<td>23.20</td>
<td>10.18</td>
<td>&lt;0.001</td>
<td>0.07</td>
</tr>
<tr>
<td>$H_e$</td>
<td>$S_{IBR} + \text{population size class}$</td>
<td>-128.50</td>
<td>0</td>
<td>0.42</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>$S_{IBR}$</td>
<td>-127.90</td>
<td>0.59</td>
<td>0.31</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR} + \text{population size class}$</td>
<td>-126.20</td>
<td>2.26</td>
<td>0.13</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR}$</td>
<td>-125.50</td>
<td>2.93</td>
<td>0.10</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + \text{population size class}$</td>
<td>-121.80</td>
<td>6.70</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD}$</td>
<td>-121.20</td>
<td>7.25</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$\text{population size class}$</td>
<td>-121.10</td>
<td>7.34</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>$H_o$</td>
<td>$S_{IBR}$</td>
<td>-92.40</td>
<td>0</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD}$</td>
<td>-91.20</td>
<td>1.20</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>$\text{population size class}$</td>
<td>-90.50</td>
<td>1.90</td>
<td>0.10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR}$</td>
<td>-89.80</td>
<td>2.60</td>
<td>0.07</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>$S_{IBR} + \text{population size class}$</td>
<td>-89.40</td>
<td>3.00</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + \text{population size class}$</td>
<td>-88.30</td>
<td>4.10</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>$S_{IBD} + S_{IBR} + \text{population size class}$</td>
<td>-86.50</td>
<td>5.90</td>
<td>0.01</td>
<td>0.13</td>
</tr>
</tbody>
</table>
A 10: Results of linear mixed effect models showing the strength and significance of relationships between alternate genetic diversity metrics and fitness-related traits for ten populations of *P. vulgaris*. Seed set and seed mass were measured as a mean value per mother plant within ten populations and genetic diversity predictors were measured as a mean value per population. A random population effect was included in the mixed models to control for repeated measures of fitness-related traits per population. The significance of relationships was tested with likelihood ratio tests following a Chi square distribution. Marginal $R^2$ of fixed effects are shown. Abbreviations: $A_e$, mean effective number of alleles per population; $H_e$, mean expected heterozygosity per population; $H_o$ mean observed heterozygosity per population.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed set ~ $A_e$</td>
<td>0.12</td>
<td>9.20</td>
<td>0.002</td>
</tr>
<tr>
<td>Seed set ~ $H_e$</td>
<td>0.08</td>
<td>5.79</td>
<td>0.02</td>
</tr>
<tr>
<td>Seed set ~ $H_o$</td>
<td>0.006</td>
<td>0.39</td>
<td>0.5</td>
</tr>
<tr>
<td>Seed mass ~ $A_e$</td>
<td>0.25</td>
<td>13.60</td>
<td>0.0002</td>
</tr>
<tr>
<td>Seed mass ~ $H_e$</td>
<td>0.12</td>
<td>4.90</td>
<td>0.03</td>
</tr>
<tr>
<td>Seed mass ~ $H_o$</td>
<td>0.001</td>
<td>0.03</td>
<td>0.87</td>
</tr>
</tbody>
</table>
A 11: Spearman rank correlations (*rho*) among predictor variables for population-level analysis in Chapter 6. Population size (flowering plants) and landscape predictors measured within 50 – 1000 m buffers around populations are shown. Abbreviations: seminat, seminatural habitat.

<table>
<thead>
<tr>
<th></th>
<th>flowering plants</th>
<th>forest 50</th>
<th>forest 100</th>
<th>forest 250</th>
<th>forest 500</th>
<th>forest 1000</th>
<th>seminat 50</th>
<th>seminat 100</th>
<th>seminat 250</th>
<th>seminat 500</th>
<th>seminat 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>flowering plants</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest 50</td>
<td>-0.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest 100</td>
<td>-0.77</td>
<td>0.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest 250</td>
<td>-0.77</td>
<td>0.25</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest 500</td>
<td>-0.59</td>
<td>-0.11</td>
<td>0.93</td>
<td>0.93</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest 1000</td>
<td>-0.54</td>
<td>0.07</td>
<td>0.86</td>
<td>0.86</td>
<td>0.82</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>seminat 50</td>
<td>-0.05</td>
<td>-0.71</td>
<td>0.29</td>
<td>0.29</td>
<td>0.54</td>
<td>0.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>seminat 100</td>
<td>0.54</td>
<td>-0.75</td>
<td>-0.14</td>
<td>-0.14</td>
<td>0.18</td>
<td>-0.04</td>
<td>0.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>seminat 250</td>
<td>0.36</td>
<td>0.04</td>
<td>-0.46</td>
<td>-0.46</td>
<td>-0.43</td>
<td>-0.39</td>
<td>-0.21</td>
<td>0.46</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>seminat 500</td>
<td>0.31</td>
<td>0.36</td>
<td>-0.64</td>
<td>-0.64</td>
<td>-0.75</td>
<td>-0.61</td>
<td>-0.57</td>
<td>0</td>
<td>0.86</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>seminat 1000</td>
<td>0.58</td>
<td>0.07</td>
<td>-0.86</td>
<td>-0.86</td>
<td>-0.93</td>
<td>-0.46</td>
<td>0.11</td>
<td>0.64</td>
<td>0.82</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
A 12: Scatterplot showing the relationship between the proportion of seed set and proportion of selfed seeds per maternal plant across seven *P. vulgaris* populations (Chapter 6). A generalized mixed model including both population and individual-level random effects found a non-significant negative relationship (GLMM, estimate = -0.48 ± 0.36, deviance = 339.3, marginal $R^2 = 0.0008$; LR tests, $\chi^2_{(1)}=1.7 \ p=0.019$).
A 13: Within population pollen flow trajectories, reconstructed using paternity analysis (Chapter 6). Circles represent individual plants, and the size of circles represents the floral display size (number of inflorescences). Filled circles indicate maternal plants where seeds were sampled for paternity analysis. Lines between plants indicate realized pollen flow events, and the thickness of the lines represent the frequency of pollen flow events between mother and sire.
Copyright Acknowledgements

Chapter 2 is reproduced with permission from Springer International Publishing. The final publication is available at: http://dx.doi.org/10.1007/s40823-016-0013-x

Chapter 3 is reproduced with permission from John Wiley & Sons Inc. License number: 3894890659273.

Chapter 4 is published under a Creative Commons Attribution 3.0 license (CC-BY-NC-SA) and no special permission is required to reproduce for noncommercial reuse.