8-2019

Biological Invasions on a Large Scale: Investigating the Spread of Baby's Breath (Gypsophila paniculata) Across North America

Sarah K. Lamar
Grand Valley State University

Biological Invasions on a Large Scale: Investigating the Spread of Baby's Breath (*Gypsophila paniculata*) Across North America

Sarah Katherine Lamar

A Thesis Submitted to the Graduate Faculty of

GRAND VALLEY STATE UNIVERSITY

In

Partial Fulfillment of the Requirements

For the Degree of

Master of Science

Biology Department

August 2019

DEDICATION

This thesis is dedicated to my late grandfather, Willard G. Shimmel, who filled my childhood summers with birds, snakes, toads, and countless walks in the woods and garden. He understood that nature has value simply because it *is*, not because of anything it can give us.

Thank you for everything I can never repay.

ACKNOWLEDGMENTS

The contribution of Dr. Charlyn Partridge to this thesis and my professional and academic development cannot be overstated. Thank you for your patience, humor, and guidance at every turn. Additional thanks to my committee members, Dr. Tim Evans and Dr. Jennifer Moore, for their input, direction, and support. Thank you to Dr. Eric Snyder, Dr. Megan Woller-Skar, Dr. Georgette Sass, and Dr. Coeli Fitzpatrick for the guidance and mentorship they have provided me at various points throughout my education. To the faculty and staff at the Robert B. Annis Water Resources Institute – thank you for your help and laughter along the way; the institute has been a wonderful place to learn and grow over the past two years. Special thanks to my cohort and lab mates for their support, particularly Hailee Leimbach-Maus for her help within the lab and her friendship outside of it. Thank you to my mother, Joanne, for listening to me practice my presentations over the phone and telling me they were great, even when they weren't. Further thanks to my brother, William, for his unwavering support, and to my father for his encouragement. Finally, thank you to my wonderful partner, friends, and family; life is a group effort.

ABSTRACT

Invasive species are a serious threat to biodiversity worldwide. While the impacts of invasive species increase annually, many gaps in our understanding of how these species invade, adapt, and thrive in the novel ecosystems into which they are introduced remain. This thesis aimed to add to our knowledge of invasion science, using the perennial forb *Gypsophila paniculata* as a study system. *Gypsophila paniculata* is a shrub native to the Eurasian steppe that was introduced into North America in the late 1800's. After introduction, *G. paniculata* quickly spread and now occupies diverse ecosystems across N. America. In chapter II of this thesis, I assessed relationships among *G. paniculata* growing in seven locations across its introduced range and current invasion status using historical herbarium records. Genetic relationships were analyzed using microsatellite analyses, which suggested the presence of two genetic clusters; when herbarium records were grouped according to these clusters, two distinct expansion phases became visible, suggesting the presence of at least two invasion events. In chapter III, I analyzed two populations of *G. paniculata* growing in distinct environments (Chelan, Washington and Petoskey, Michigan) for phenotypic and gene expression differences that may confer potential adaptation to unique environmental stressors. Results revealed that seeds collected from Washington germinated significantly quicker than seeds collected in Michigan (pairwise log-rank test, $p < 0.0001$). When grown in a common garden, seeds collected in Washington had higher levels of emergence (two-sided proportion test, $p=0.00018$). No significant differences in tissue allocation between populations were observed (ANOVA, $p = 0.0645$); however, family effects were visible (ANOVA, $p=0.0301$), though whether they are a function of maternal investment or evidence of genetic differences is unclear. Finally, results of RNA-seq transcriptome analyses revealed 1,149 genes differentially expressed among all tissue types

(root, stem, and leaf); when considered according to tissue type and growing location, overrepresentations of genes related to circadian rhythm, stress responses, and nutrient deprivation were observed among the genes that were differentially expressed. These results not only add to our understanding of the North American invasion of *Gypsophila paniculata*, but also increase our understanding of how invasive species may be able to cope with the novel environments they encounter in their introduced range.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER.FIGURE PAGE

ABBREVIATIONS

AD-MI = Arcadia Dunes, Michigan

BSA = bovine serum albumin

Ca = calcium

CEC = cation exchange capacity

CH-WA (CHWA) = Chelan, Washington

cm = centimeter

DNA = deoxyribonucleic acid

g = gram

GPS = global positioning system

ha = hectare

K = potassium

km = kilometer

KR-ND = Knife River Historic Indian Villages, North Dakota

m = meter

µM = micromolar

meq = milliequivalents

mg = milligram

mL = milliliter

mM = millimolar

M = molar

Mg = magnesium

ng = nanogram

NOAA = National Oceanic and Atmospheric Administration

nSSR = nuclear simple sequence repeats (nuclear microsatellites)

OB-WA = Osborne Bay, Washington

OM = organic matter

OT-MN = Otter Tail, Minnesota

PO_4= phosphorus

ppm = parts per million

PS-MI (PSMI) = Petoskey, Michigan

RNA = ribonucleic acid

RNA-seq = ribonucleic acid sequencing

SBD-MI = Sleeping Bear Dunes National Lakeshore, Michigan

CHAPTER I

INTRODUCTION

Biological invasions in the post-neolithic era are characterized by the introduction and establishment of a species outside of its native distribution, well beyond the borders of what could feasibly be attributed to natural range expansion and having some negative impact on this introduced range (Pyšek 1995). These invasions are often human-mediated, and as the footprint of human travel around the globe increases, so do the number of invasion events (Richardson and Pyšek 2006). Invasive species can have serious impacts on their new environments, impairing ecosystem services and the biotic communities built upon them (Pejchar and Mooney 2009; Vilà et al. 2011). Invasive species cost the United States (U.S.) upwards of $120 billion to mitigate on an annual basis, and nearly 45% of the species listed on the Endangered Species Act are threatened by either direct or indirect competition with invaders (Pimentel et al. 2005). Because of the growing and significant impacts of invasive species, they continue to be a focus of biological research.

The invasion history of most species follows a three stage pathway: 1) a lag phase, where numbers are relatively low and the introduced species has not yet become invasive, 2) an exponential growth phase, where the species' numbers or range rapidly expand, and 3) a plateau, which is reached when the new range becomes saturated (Larkin 2012). During the lag phase of a potential invasion, complete eradication of an invasive species may be achieved without a total understanding of a species' life history (Allendorf and Lundquist 2003). However, once the growth phase of an invasion has been reached, management goals shift from eradication to control, and information about a species' life history becomes increasingly important to guide management efforts (Allendorf and Lundquist 2003).

While the impacts of invasive fauna cannot be overlooked, plant species are some of the most problematic invaders in the U.S, where non-native plants now number almost ¼ of the total plant species (Pimentel et al. 2001; Pimentel et al. 2005). Many of these invasive plant species were brought over for agricultural or ornamental purposes and eventually escaped domestication (Pimentel et al. 2005). Invasive plants can dramatically alter their ecosystems, pushing out native species and dominating environments (Richardson and Pyšek 2006). Some invasive plant species have been shown to shift regional fire regimes (Balch et al. 2013), alter wetland plant communities (Zedler and Kercher 2004), and even change the properties of the soil in which they grow (Weidenhamer and Callaway 2010). Unfortunately, many plant invasions go unchecked until they require immediate intervention to mitigate.

While no life history trait is an absolute predictor of invasive potential, some traits have been consistently linked with successful plant invasions. Richardson and Pyšek (2006) found that small seed sizes, large seed crops, short germination periods, and relatively fast growth rates correlate positively with invasion success. Small seeds are typically associated with large seed crops (Kawano 1981; Greene and Johnson 1994), ease of dispersal (Harper et al. 1970), high initial germinability (Grime et al. 1997), and fast relative growth rate (Maranon and Grubb 1993). Larger seed crops provide an opportunity for rapid population expansion, and short germination periods with fast relative growth rates allow invasive species to begin to accumulate resources before slower native plant species have established (Rejmánek and Richardson 1996). While these traits may help facilitate invasion success, they do not guarantee it. However, knowing how an invasive species has been able to succeed in its introduced environment is important for the development of focused and effective management of problematic invaders.

Similarly, knowledge of a species' invasion history can inform managers about probable response times to an invasion, the number of distinct invasion events that have occurred, the size of founding populations, and the potential for those invasive populations to adapt (Lockwood et al. 2005; Rejmánek et al. 2005). Typically, as the size and number of founding populations of an invasive species increases, so does the genetic diversity of that species (Roman and Darling 2007). Because the rate of response to natural selection is directly related to the amount of additive genetic variance present within a population, knowledge of a species' genetic diversity is valuable to managers (Sakai et al. 2001). Populations of invasive species with high levels of genetic diversity and additive genetic variance are more likely to successfully respond to the selective pressures of their new environments.

Gypsophila paniculata (common baby's breath) is a perennial forb native to Eurasia that can now be found inhabiting diverse habitats across much of the U.S. and Canada (Darwent and Coupland 1966; EDDMapS 2019). *Gypsophila paniculata* is a prolific reproducer, a single plant producing over 13,000 seeds per year (Stevens 1957). It also has a taproot that can grow to be several meters deep, helping the invader to outcompete native species for limited resources (Darwent and Coupland 1966). *Gypsophila paniculata* often forms dense stands in the areas it invades; in one portion of Sleeping Bear Dunes National Lakeshore, an invaded region of Michigan, it makes up 75% of the present vegetation (Rice 2018). Though this species is the focus of management efforts throughout the U.S., information about its invasion history and the adaptations that may have helped it to become a successful invader are largely unknown.

Purpose

This thesis aims to fill important gaps in our knowledge of the invasion and life histories of *Gypsophila paniculata*. Using a combination of historical herbarium data and microsatellite analysis, Chapter II aims to reconstruct invasion curves of *G. paniculata* to approximate its invasion stage at different scales and to establish the genetic relationships among seven contemporary sampling locations of this species. Chapter III looks at gene expression between two sampling locations of *G. paniculata* (Michigan and Washington), distinct both geographically and in habitat type, to investigate potential adaptations that may have helped *G. paniculata* successfully invade these two environments. Chapter III also provides insight into above vs belowground tissue allocation during the first month of growth and the germination rate of seeds from these same two sampling locations.

Scope

This study details the invasion history of *Gypsophila paniculata* throughout North America. The genetic relationship among seven sampling locations was established (Chelan, WA; Osborne Bay, WA; Knife River Historic Indian Villages, ND; Otter Tail, MN; Petoskey, MI; Sleeping Bear Dunes National Lakeshore, MI; and Arcadia Dunes, MI). Gene expression, germination characteristics, and early tissue allocation were investigated between two of those sampling locations: Chelan, WA and Petoskey, MI. These locations were chosen for further analysis due to their long history of invasion (~100 years) and diverse ecotypes (Washington sagebrush steppe and Michigan dune shore).

Assumptions

1. I assumed tissue sampling was unbiased and representative of the sampling locations.

2. I assumed soil sampling was unbiased and representative of the sampling locations.
3. I assumed negligible amounts of instrumental sequencing error.
4. I assumed that herbarium samples not able to be visually confirmed to species were accurately recorded by their institutions.
5. I assumed microsatellite markers (nSSR) were neutral.
6. I assumed seeds that had not germinated at 14 days, and after two successive days of no germination across all seeds, were no longer viable and would not germinate.
7. I assumed daily randomization of petri dishes and growth cones in experiments buffered against biases in light, temperature, or moisture regimes.

Objectives

The objectives of chapter II were: 1) to recreate the invasion history of *Gypsophila paniculata* and investigate invasion stages at two scales (Michigan and North America), and 2) to establish population structure among seven sampling locations of *G. paniculata* spread across its North American range. The objectives of chapter III were 1) to investigate potential differences in gene expression between two sampling locations of *G. paniculata* growing in diverse, harsh habitats (Chelan, WA and Petoskey, MI), 2) to estimate germination rate, and 3) to investigate above vs belowground tissue allocation during the first month of seedling growth between Michigan and Washington sampling locations.

Significance

This thesis fills important gaps related to the invasion and life history of *Gypsophila paniculata*. This weed species is the focus of management efforts in Michigan's dune habitat, where it can have significant impacts on native species. A better understanding of the invasion history of *G. paniculata* allows us to consider the potential of this species to

continue to spread. Additionally, investigating life history traits of *G. paniculata* allows us to understand how invasive species are able to proliferate across diverse, novel habitats. In conclusion, this thesis improves our understanding of the invasion of *Gypsophila paniculata* in North America and makes contributions to the greater body of literature surrounding plant invasions and the ability of invasive species to invade, adapt, and thrive in novel ecosystems.

DEFINITIONS

Adventive Species

A species that has been introduced into an area with the aid of human activity, but may not yet be established in an area

Contig(s)

Overlapping DNA segment that represents a consensus region of DNA

Germination

Functionally defined in this study as emergence of the radicle (embryonic root)

Introduced Species (Alien Species)

A species that has reached a new area because of the activities of post-neolithic man or domesticated animals, beyond what could be attributed to natural range expansion; found outside control or captivity as a potentially self-sustaining population

Invasion Curve

A plot of the range or cost of management of an invasive species as a function of time

Invasive Species

An introduced species that is increasing in range or abundance and has negative impacts on its new environment

Microsatellite(s)

Short, tandem repeats throughout the genome, typically found in series with one another

Population (Genetic Population)

A group of individuals of a single species that share enough genetic information to be viewed as a single unit from a genetic standpoint; not related to geographic proximity.

Sampling Location

A distinct geographic area assumed to be contained where sampling of a species occurred.

CHAPTER II

Manuscript

Old Meets New: Combining herbarium databases with genetic methods to evaluate the invasion status of baby's breath (*Gypsophila paniculata*) in North America

ABSTRACT

Aim: This paper aims to inform our knowledge of the current population structure and invasion status of *Gypsophila paniculata* (common baby's breath) using a combination of contemporary genetic methods and historical herbarium data.

Taxon: *Gypsophila paniculata* (Angiosperms: Eudicot, Caryophyllaceae)

Location: Samples were collected from seven locations spanning a portion of the plant's North American range: Washington, North Dakota, Minnesota, and Michigan, United States.

Methods: To analyze contemporary population structure, individuals of *G. paniculata* from 7 distinct sampling locations were collected and genotyped at 14 microsatellite loci. Population structure was inferred using both Bayesian and multivariate methods. To investigate the invasion status of *G. paniculata*, public herbarium databases were searched for collections of the species. Records were combined, resulting in a database of 307 herbarium collections dating from the late 1800's to current day. Using this database, invasion curves were created at different spatial scales.

Results: Results of genetic analyses suggest the presence of at least two genetic clusters spanning the seven sampling locations. Sampling locations in Washington, North Dakota, Minnesota, and northwestern Michigan form one genetic cluster, distinct from the two more southern sampling locations in Michigan, which form a second cluster with increased relative genetic diversity. Invasion curves created for these two clusters infer different time periods of invasion. An invasion curve created for North America suggests the range of *G. paniculate* may still be expanding.

Main conclusions: *Gypsophila paniculata* has likely undergone at least two distinct invasions in North America, and its range may still be expanding. Restricted genetic diversity seen across a wide geographic area could be due to a limited number of seed distributors present during the early period of this horticultural import's invasion.

Keywords: Baby's breath, genetic structure, *Gypsophila paniculata*, herbarium data, invasion history, invasive species

INTRODUCTION

Biological invasions are a growing concern in the era of global trade and transport. In the United States alone, there have been over 50,000 introductions of plant, animal, and microbe species into environments beyond their native range (Pimentel et al. 2005). These introductions can have dramatic impacts on native flora and fauna; roughly 42% of species listed under the U.S. Endangered Species Act are threatened by competition with invasives (Wilcove et al. 1998). Of particular concern among invasive species are invasive weeds, a group that currently spreads across the United States at a rate of 700,000 ha/year (Pimentel et al. 2005). This rapid consumption of land by non-native species makes managing invasive weeds a priority for the preservation of native ecosystems and the native biota that inhabit them.

Many plant and animal species that are transported into new environments will not become problematic invaders, defined as species not native to an area whose range or abundance is increasing regardless of habitat (Pyšek 1995; Williamson and Fitter 1996). Non-native species that go on to become invasive in their new environments can face many barriers to successful introduction, such as surviving transport, reproducing as a relatively small founding population, responding to potentially novel environmental stressors, and overcoming a "lag phase" (Larkin, 2012; Williamson & Fitter, 1996). This lag phase is characterized by a period of slow growth after initial introduction that, if overcome, can lead to a period of rapid population expansion before eventually plateauing as the new range is saturated (Mack et al., 2000). Despite the many potential barriers species face on the road to becoming invasive, the impacts of these events are a growing cause for concern.

As the number of global invasion events increases, so does the importance of developing and implementing cost effective methods for studying invasion events. Invasion curves are one

such tool used to assess an invasive species' status and rate of spread (see Antunes & Schamp, 2017; Shih & Finkelstein, 2008). Invasion curves can offer researchers important insight to a species' lag time after introduction into new environments, providing valuable information associated with response time, geographic barriers to spread, and the efficacy of existing management strategies (Crooks 2007; Antunes and Schamp 2017). Because they are crafted using historical data, such as herbarium records, invasion curves are both cost effective and capable of offering important glimpses into the often-unnoticed lag phase of an invasion (Antunes and Schamp 2017). Invasion curves have been used to recognize potential refuges for weed species (e.g. Lavoie, Jodoin, & De Merlis, 2007), identify major drivers of invasive species spread (e.g. Fuentes, Ugarte, Kühn, & Klotz, 2008; Petr Pyšek, Jarošík, Müllerová, Pergl, & Wild, 2008), and even help assess the efficacy of potential biocontrol agents (e.g. Boag & Eckert, 2013).

While invasion curves are useful for addressing many questions managers and researchers may have, they are limited by the constraints associated with herbarium records and survey data. To overcome these constraints, genetic analyses may be used to provide information concerning contemporary gene flow, adaptive potential, relatedness among invasive populations, and possible resistance to control efforts (e.g. Abdelkrim, Pascal, Calmet, & Samadi, 2005; Zalewski et al., 2010). Genetic analyses of invasive species have been used to identify potential barriers to migration (Haynes et al. 2009) and estimate the number of likely invasion events a species may have undergone (Meimberg et al. 2010) While this information can help improve our understanding of invasion science as a whole, it also has immediate benefits to managers. Because distinct genetic populations have different potential evolutionary trajectories,

understanding the genetic structure of populations is critical for effective management (Moritz 1994; Palsbøll et al. 2007).

Gypsophila paniculata (common baby's breath) is a perennial forb native to the Eurasian steppe region (Darwent and Coupland 1966; Darwent 1975). *Gypsophila paniculata* is characterized by a taproot that can reach several meters deep, which is thought to help the plant to out-compete natives for limited resources in harsh environments (Darwent and Coupland 1966). Though it does not produce floral primordia until at least its second year, *G. paniculata* can yield almost 14,000 seeds per growing season (Stevens 1957; Darwent and Coupland 1966). These seeds are small (86mg/100 seeds) and primarily distributed by wind; when plants reach senescence, they break off above the caudex and form tumbleweeds that spread seeds as they roll (Stevens 1957; Darwent and Coupland 1966).

Populations of *G. paniculata* were established in North America by the late 1880's, likely having been introduced due to its popularity in the garden and floral industries (Darwent and Coupland 1966). According to the Early Detection and Distribution Mapping System, *G. paniculata* can now be found growing as an invasive species in 30 U.S. states (EDDMapS 2019). It has been listed as a Class C (widespread noxious weed) in Washington and California and is considered a priority invasive by Michigan Department of Natural Resources (Emery and Doran 2013; Michigan Department of Natural Resources 2015; Swearingen and Bargeron 2016). *Gypsophila paniculata* can form dense stands in the areas that it invades; in some parts of Sleeping Bear Dunes National Lakeshore, an invaded area in Michigan, *G. paniculata* forms as much as 75% of the vegetation present (Karamanski 2000; Rice 2018). These dense monocultures can have impacts on native plant, nematode, and arthropod communities,

potentially having ripple effects across the trophic system (Emery and Doran 2013; Reid and Emery 2018).

To help understand the invasion status of this problematic plant species, this study aims to (1) define the population structure of contemporary *G. paniculata* growing throughout a portion of its introduced range, and (2) create invasion curves of *G. paniculata* to assess its current invasion status at different spatial scales.

METHODS

Study Sites and Contemporary Sample Collection

To investigate contemporary population structure of *G. paniculata*, tissue samples from five locations across the United States were collected in June of 2018: Petoskey, MI; Knife River Indian Villages National Historic Site, ND; Ottertail, MN; Chelan, WA; and Osborne Bay, WA (Figure 1, Table 1). Samples from two additional locations in Sleeping Bear Dunes National Lakeshore, MI and Arcadia Dunes, MI were collected in the summer of 2016 (Table 1) (Leimbach-Maus et al. 2018). Leaf tissue was collected from 15-30 individuals per location (5-10 leaves per plant). Tissue samples were placed inside coin envelopes and stored in silica until DNA extraction. Individuals were collected for sampling by identifying a plant of any size separated from other sampled individuals by at least 2 meters, in efforts to minimize the likelihood of sampling closely related plants.

Microsatellite Analysis of Contemporary Samples

For each contemporary sample (n=145), DNA was extracted from 0.25 g of dried leaf tissue using a Qiagen DNeasy plant mini kit (QIAGEN, Hilde, Germany), modified to include an extra wash with AW2 buffer. Extracted DNA was cleaned twice using a Zymo OneStep PCR Inhibitor

Removal Column (Zymo, Irvine, CA). Samples were amplified at 14 nuclear microsatellite loci identified as polymorphic and specific to *G. paniculata* (Leimbach-Maus, Parks, & Partridge, 2018b). PCR was conducted using a 5' fluorescently-labelled primer (6-FAM, PET, NED, or VIC) (Applied Biosystems, Foster City, CA) and an unlabeled reverse primer. Reaction mixtures consisted of 1x KCl buffer, 2.0-2.5 mM $MgCl_2$, 300 µM dNTP, 0.08 mg/mL BSA, 0.4 µM forward primer, 0.4 µM reverse primer, 0.25 units Taq polymerase, and 50 ng DNA template. The thermal cycling profile consisted of 5 minutes of denaturation at 94˚C, followed by 35 cycles of 94˚C for 1 minute, 1 minute of annealing at 62˚ (with the exception of locus BB_2888, see Leimbach-Maus et al. 2018b), 1 minute of extension at 72˚C, and a final elongation step of 10 minutes at 72˚C. PCR products were visualized on a 2% agarose gel using GelRed™ (Biotium, Freemont, CA) before multiplexing with consideration to dye color and allele size. Genescan 500 LIZ size standard (Thermo Fisher Scientific, Waltham, MA) was added to multiplexed product with Hi-Di™ Formamide (Thermo Fisher Scientific, Waltham, MA) to aid in denaturing. Fragment analysis was conducted on an ABI3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). Individuals were genotyped using the automatic binning procedure on GENEMAPPER v5 (Applied Biosystems, Foster City, CA) before being visually verified to reduce error. A subsample of 20 individuals were genotyped twice to ensure consistent allele scoring.

The presence of null alleles was investigated using MICRO-CHECKER v2.2.3; using this method, none were found (Van Oosterhout et al. 2004). Data were screened using the 'STRATAG' package in the R statistical program v3.4.3 (Archer et al. 2016; R Development Core Team 2017) for any individual that was missing greater than 20% of loci and any locus that was missing greater than 10% of individuals; on this basis, no data were removed.

Measures of Genetic Diversity and Structure in Contemporary Populations

Linkage disequilibrium and a test for Hardy-Weinberg equilibrium were calculated using GENEPOP v4.6 with 1,000 batches of 1,000 Markov chain Monte Carlo iterations (Raymond and Rousset 1995; Rousset 2008). There was no significant deviation from linkage equilibrium across populations and no data were removed on this basis. Expected versus observed heterozygosity, number of private alleles, and Weir and Cockerham's population pairwise F_{ST} values were conducted using GENALEX v6.502 in Microsoft Excel (Weir and Cockerham 1983; Peakall and Smouse 2006; Peakall and Smouse 2012). Inbreeding coefficient (F_{IS}) values were calculated in GENEPOP.

A principal coordinate analysis (PCoA) was conducted using a genetic distance matrix in GENALEX (Peakall and Smouse 2006; Peakall and Smouse 2012). Population clustering was analyzed with Bayesian methods in STRUCTURE v2.3.2(Pritchard et al. 2000) using an admixture model, both with and without *a priori* location information, and a burn-in length of 100,000 with 1,000,000 MCMC replicates after burn-in. Ten iterations were run for each *K* value (1-9). The number of genetic clusters was determined using the Evanno ΔK method (Evanno et al. 2005). Because ΔK is based on a rate of change, it does not evaluate *K*=1 and can be biased towards *K*=2 (Dupuis et al. 2017). Considering this, we also used discriminant analysis of principal components (DAPC) to support our STRUCTURE findings (Jombart et al. 2010). DAPC separates variance into within-group and between-group categories and works to maximize cluster discrimination; this analysis was conducted using the package 'adegenet' v2.1.1 in R (Jombart et al. 2010). Because retaining too many principal components (PC's) can lead to instability in cluster membership properties, a cross-validation was performed to inform the analysis of the optimal number of PC's. After cross-validation, 16 of 28 PC's and all eigenvalues

were retained. An analysis of molecular variance (AMOVA) was run using 9,999 permutations in GENALEX to test how much variance could be explained by between-population and within-population variation; regions were defined according to genetic cluster identified by STRUCTURE analysis.

Invasion Curves

To create invasion curves for *G. paniculata* population clusters, public herbarium databases were searched for specimen records of this species; species identification was visually confirmed when possible. Records that did not include location data (either GPS, county (U.S.) or regional municipality (Canada)) and year were discarded, resulting in 307 records from 65 North American institutions (Table S1). All locality information was standardized to the county scale to reduce the risk of redundant specimen collection while maintaining adequate resolution (Antunes and Schamp 2017). Earliest samples were found in the late 1890's-early 1900s in California, Michigan, Minnesota, and New York and this is consistent with the earliest times in which *G. paniculata* seeds were first being sold in the United States (1886), based on a search of the Henry G. Gilbert Nursery and Seed Trade Catalog Collection from the Biodiversity Heritage Library (https://www.biodiversitylibrary.org/).

To examine the invasion status of contemporary populations belonging to genetic clusters identified from our population genetics analysis, herbarium records were grouped according to desired spatial scales (cumulative North America, current location of genetic cluster 1, and current location of genetic cluster 2 in contemporary samples). Only specimen records for the first collection of *G. paniculata* in each county or regional municipality were kept. Cumulative records for all of North America had 184 unique municipalities represented, while records from the geographic area of both genetic clusters had fewer unique localities (cluster 1 = 42, cluster 2

=16) and required log transformation for better visualization. Data were plotted as the cumulative number of localities invaded over time using the statistical program R v6.0.

RESULTS

Measures of Genetic Diversity and Structure in Contemporary Populations

The five western populations and northernmost Michigan population (i.e. genetic cluster 1) showed lower levels of genetic diversity compared with the two more southern populations in Michigan (i.e. genetic cluster 2) (Table 2). Pairwise comparisons yielded significant F_{ST} values between all populations; however, SBD-MI, AD-MI, and KR-ND showed comparatively high pairwise F_{ST} values compared to other populations (Table 3). F_{ST} values between CH-WA, OB-WA, and PS-MI were relatively low compared to other sample locations in this study, suggesting more limited genetic differentiation among these populations (Table 3).

Results of Bayesian clustering analysis suggest two population clusters (*K*=2), both from ΔK and Ln Pr (X|K) (Figure S1). Analysis was conducted both with and without prior sampling location; there was no observable difference between the two (without priors shown in Figure 2). Cluster 1 is comprised of sampling locations in North Dakota, Minnesota, Washington, and the northernmost site in Michigan; cluster 2 is comprised of the two more southern sites in Michigan (Figure 2). Overall, there is little admixture between the two clusters, with only few individuals in AD-MI showing any signs of genetic mixing.

Contemporary population structure was further analyzed with a PCoA based on a genotypic distance matrix. Population division along the primary principal coordinate accounted for 27.22% of variation present. Along this coordinate, the trends seen in STRUCTURE analysis were supported, with populations SBD-MI and AD-MI separating out from the remaining five populations (Figure 3). The secondary principal component suggests further separation may exist

between SBD-MI and AD-MI (9.80% of variation present) if *K* is forced to 3. The grouping of CH-WA, OB-WA, OT-MN, KR-ND, and PS-MI into the same cluster is supported by this analysis.

DAPC's Bayesian Information Criterion suggested either 2 or 3 genetic clusters (Figure S2). Sampling locations in Arcadia Dunes, MI and Sleeping Bear Dunes, MI separated into distinct populations when *K* was pushed to 3, in order to investigate all cluster possibilities (Figure 4a). Individual membership to clusters is detailed in Figure 4c, which shows that cluster 1 is 82% comprised of individuals from SBD-MI, cluster 2 is 93% comprised of individuals from AD-MI, and cluster 3 has a relatively even contribution of individuals from CH-WA, OB-WA, KR-ND, OT-MN, and PS-MI. When individual distribution is viewed along the primary discriminant function, overlap between clusters 1 (SBD-MI) and 2 (AD-MI) is clearly visible (Figure 4b), while cluster 3 shows little to no overlap with clusters 1 or 2.

AMOVA results show that a significant amount of variation could be explained by differences among populations within regions (ϕ_{PR}= 0.229, p <0.001) and by differences between our first region (CH-WA, OB-WA, KR-ND, OT-MN, PS-MI) and second region (SBD-MI and AD-MI) (ϕ_{RT} = 0.246, p <0.001). However, most variation present was found within populations (ϕ_{PT} = 0.419, p<0.001).

Invasion Curves

Invasion curves created using herbarium specimen records, standardized to the scale of local municipality, were used to visualize the invasion stage (i.e. lag phase, expansion phase, or plateau phase) of *G. paniculata* at various spatial scales (Figure 5). Records for North America slowly accumulate during the early periods of invasion (1890's) until roughly the 1940's, after which the number of records being collected in new localities begin to accumulate rapidly

(Figure 5b). This likely represents the shift from the initial lag phase of invasion to the expansion phase. With no clear plateau being reached, the expansion phase of *G. paniculata* across the entirety of North America appears to continue. Considering herbarium specimen records collected in the geographic area of each contemporary genetic cluster, initial collection for cluster 1 (WA, ND, MN, and PS-MI) is noted in the late 1890's, but few additional records were archived until the mid 1920's, when herbarium data for *G. paniculata* suggest an expansion of this population (Figure 5c). A plateau can be seen beginning in the mid 1990's when the curve of the line begins to taper. Specimen records from the same location as contemporary genetic cluster 2 (Figure 5d) are comprised of collections from mid-southwest Michigan (defined as south of the Leelanau Peninsula, based on results from this study and a previous study conducted by Leimbach-Maus et al., 2018a). Rapid expansion began shortly after its first collection in the late 1940's, with the spread beginning to plateau around 1970. No discernable lag period is noted in the collection data for this cluster.

DISCUSSION

Our data from populations of *G. paniculata* growing across a portion of its introduced range in North America reveal the presence of at least two distinct genetic clusters in contemporary populations. The northernmost sampling location in Michigan (PS-MI) clustered with the four sampling locations located across North Dakota, Minnesota, and Washington, and separately from the two southernmost sampling locations in Michigan. When further structuring was explored, the two MI locations (AD-MI and SBD-MI) separated out into their own genetic clusters, though overlap was clearly visible when viewing discriminant functions. The two more southern sampling locations in Michigan also had higher levels of genetic diversity than the other five sampling locations.

There are likely multiple factors contributing to the genetic patterns that we observe across these populations of invasive baby's breath. The increased levels of genetic diversity observed in the SBD-MI and AD-MI populations compared with the other sampled locations could be due to a combination of population size and connectivity. Populations located in SBD-MI tend to be much larger than other locations sampled in this study. Larger populations tend to be more robust to the effects genetic drift and can help resist the effects of inbreeding, helping to retain diversity within these populations (see Ellstrand & Elam, 2003). Another possible reason for the patterns found here is that sampling locations spread across the western U.S. are more isolated than the two southernmost Michigan locations, which may be contributing to lower levels of genetic diversity among these areas. Some sample locations (CH-WA, OT-MN) occur in relatively fragmented or space-limited environments, which may result in a lack of gene flow to other populations of *G. paniculata* growing nearby or prevent its spread altogether. The close geographic proximity between SBD-MI and AD-MI could also be maintaining some gene flow between these populations. However, many of our other sample locations with limited genetic diversity (OB-WA, PS-MI, KR-ND) were part of a contiguous landscape that was not obviously limiting to expansion.

One potential explanation for the distinct genetic clustering we observed with our data is that the populations of SBD-MI and AD-MI that were established in the 1940's could have been founded by individuals from the existing PS-MI population. SBD-MI and AD-MI could then have significantly diverged from the initial source over the past 50 years. However, this scenario seems unlikely. Our data show that SBD-MI and AD-MI have higher levels of genetic variation compared to PS-MI and a number of private alleles were found in both SBD-MI and AD-MI that are not present in PS-MI. Additionally, chloroplast microsatellite data from a previous study

(Hailee B Leimbach-Maus et al. 2018) show that the SBD-MI and AD-MI populations have distinct DNA haplotypes compared to the PS-MI population and other more northern Michigan populations not included in this study. The combination of these data suggest that SBD-MI and AD-MI are likely not the result of serial founding events from the source population of PS-MI.

A more likely explanation for the distinct patterns observed among our populations could be a signature of *G. paniculata*'s horticultural past. The earliest occurrences of *G. paniculata* populations across several different regions in the U.S. coincides with its initial introduction to N. America though seed sales. Based upon seed catalogs from the Biodiversity Heritage Library, *G. paniculata* was promoted as a garden ornamental as early as 1856 in the *Farmer's Promotion Book* (Reinhold 1856). By 1868 at least two seed distributors (J.M. Thorburn & Co, NY and Hovey & Nichols, Chicago) were selling *G. paniculata* in their catalogs in New York and Chicago; the earliest herbarium records of *G. paniculata* collected in the United States were from CA (1907), MN (1896), MI (1913), and NY (1894) (Table S1). We hypothesize that when *G. paniculata* initially invaded N. America in the late 1890's there may have been little standing genetic diversity present in the garden cultivars being grown at the time. Additionally, the limited number of overseas distributors of seeds may have been further restricting possible diversity. These potential limitations to genetic diversity during the early periods of invasion are likely why some of our populations cluster together, despite the large geographic distances between them. According to herbarium specimen records, populations of *G. paniculata* in SBD-MI and AD-MI were not established until the later 1940's, when *G. paniculata* had become a more popular garden ornamental. This increased popularity likely led to the number of seed distributors being greatly increased. We suggest then that the genetic patterns observed in this

study among populations of *G. paniculata* are a signature of the horticultural past that helped facilitate its invasion into N. America.

One confounding factor to our genetic analyses is that tissue from the SBD-MI and AD-MI populations was collected two-years prior to the other locations (2016 compared to 2018). However, a study by Leimbach-Maus et al. (2018a) examining the genetic structure of populations throughout west Michigan found that baby's breath populations north of the Leelanau Peninsula (i.e., PS-MI) cluster distinctly from both SBD-MI and AD-MI. This study, combined with the perennial growth habit of *G. paniculata*, supports the distinct clustering of PS-MI from the SBD-MI and AD-MI.

Invasion curves created at multiple spatial scales help assess the current invasion status of *G. paniculata* across its introduced range in North America. Herbarium specimen records compiled for North America indicate that *G. paniculata* has likely not yet reached a plateau phase, and its range could still be expanding. When this larger invasion is viewed at a finer spatial scale, additional trends become visible. Herbarium specimens collected from the geographic area currently inhabited by cluster 1 (Washington, North Dakota, Minnesota, and northwestern Michigan) show a lag period that ended in the 1920's as *G. paniculata* collection increased in new localities and its range began expanding. The invasion curve created for the geographic area currently occupied by cluster 2 (Michigan south of the Leelanau Peninsula) shows that the expansion phase was already in process during the first collection period or shortly after, with little lag phase observed. Whether this is because *G. paniculata* was present within the region prior to this period but not collected until the 1940's, or whether populations were not present in this area until the 1940's and began spreading rapidly shortly after introduction is unclear. Regardless, the expansion in this region was in progress in the mid

1940's, with a plateau in new localities invaded taking place around 1970. These distinct expansion phases could suggest at least two separate periods of invasion occurring across our sampled range, one expanding in the 1920's and another in the 1940's.

This combination of genetic and herbarium specimen data offers valuable insight into the invasion of a problematic weed across a large portion of its invaded range. Using genetic analyses, we were able to infer the likely number of distinct invasion events across a large geographic spread of invasive weed populations. Informed by these analyses, we were then able to construct possible invasion curves that reveal trends that would otherwise have been obscured in the large pool of available data. This combination of genetic analyses as *a priori* information for the construction of herbarium specimen-derived invasion curves proves a powerful method for extracting information on the invasion status of distinct invasion events, as well as maximizes the benefits of data maintained and made freely available by herbaria across N. America. In an era of increased invasions and dwindling conservation funding, the use of existing data in the most effective and informed way possible is paramount for the continued effective management of invasive species and increased understanding of invasion success.

In conclusion, this study offers insight into the population structure and invasion status of *Gypsophila paniculata* in its introduced N. American range. Our data suggest that the distinct population clusters observed in contemporary populations through genetic analyses are likely explained by the species' history as a horticultural species, a characteristic that facilitated its spread to the continent. When viewed in light of these genetic clusters, herbarium specimen data suggest the presence of at least two invasion events, evidenced by unique expansion phases across the species' range. Combining herbarium specimen records with genetic analyses of contemporary populations has provided a more complete understanding of the invasion history of

this species, and this type of work would serve as a useful tool for characterizing the invasion status of other invasive populations.

ACKNOWLEDGEMENTS

The authors would like Hailee Leimbach-Maus and Emma Rice for their help in the field. Additionally, the authors would like to thank the Environmental Protection Agency – Great Lakes Restoration Initiative (C.G.P., Grant #00E01934), the Michigan Botanical Foundation, and Grand Valley State University for financial support.

REFERENCES

Abdelkrim J, Pascal M, Calmet C, Samadi S. 2005. Importance of assessing population genetic structure before eradication on invasive species: Examples from insular Norway rat populations. Conservation Biology. 19(5): 1509–1518.

Antunes PM, Schamp B. 2017. Constructing standard invasion curves from herbarium data—toward increased predictability of plant invasions. Invasive Plant Science and Management. 10(04): 293–303.

Archer FI, Adams PE, Schneiders BB. 2016. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources. 17(1): 5–11.

Boag AE, Eckert CG. 2013. The effect of host abundance on the distribution and impact of biocontrol agents on purple loosestrife (*Lythrum salicaria*, Lythraceae). Écoscience. 20(1): 90–99.

Crooks JA. 2007. Lag times and exotic species: The ecology and management of biological invasions in slow-motion. Écoscience. 12(3): 316–329.

Darwent AL. 1975. The biology of Canadian weeds: 14. *Gypsophila paniculata* L. The Biology of Canadian Weeds. 55(14): 1049–1058.

Darwent AL, Coupland RT. 1966. Life History of *Gypsophila paniculata*. Weeds. 14(4):313.

Dupuis JR, Janes JK, Malenfant RM, Gorrell JC, Andrew RL, Miller JM, Cullingham CI. 2017. The K = 2 conundrum. Molecular Ecology. 26(14): 3594–3602.

EDDMapS. 2019. Early Detection & Distribution Mapping System. The University of Georgia - Center for Invasive Species and Ecosystem Health. Available online at http://www.eddmaps.org/.

Ellstrand NC, Elam DR. 2003. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annual Review of Ecology and Systematics. 24(1): 217–242.

Emery SM, Doran PJ. 2013. Presence and management of the invasive plant *Gypsophila paniculata* (baby's breath) on sand dunes alters arthropod abundance and community structure. Biological Conservation. 161: 174–181.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology. 14(8): 2611–2620.

Fuentes N, Ugarte E, Kühn I, Klotz S. 2008. Alien plants in Chile: Inferring invasion periods from herbarium records. Biological Invasions. 10(5): 649–657.

Haynes GD, Gilligan DM, Grewe P, Nicholas FW. 2009. Population genetics and management units of invasive common carp *Cyprinus carpio* in the Murray-Darling Basin, Australia. Journal of Fish Biology. 75(2): 295–320.

Jombart T, Collins C. 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics. 11(94).

Karamanski TJ. 2000. A nationalized lakeshore: the creation and administration of Sleeping Bear Dunes National Lakeshore.

Larkin DJ. 2012. Lengths and correlates of lag phases in upper-Midwest plant invasions. Biological Invasions. 14(4): 827–838.

Lavoie C, Jodoin Y, De Merlis AG. 2007. How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Quebec? A historical analysis using herbarium records. Journal of Biogeography. 34(10): 1751–1761.

Leimbach-Maus Hailee B, Parks SR, Partridge CG. 2018. Genetic structure of invasive baby's breath (*Gypsophila paniculata*) populations in a freshwater Michigan dune system. Pre-print. 8042(440): 83–87.

Leimbach-Maus Hailee B., Parks SR, Partridge CG. 2018. Microsatellite primer development for the invasive perennial herb *Gypsophila paniculata* (Caryophyllaceae). Applications in Plant Sciences. 6(12): e1203.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA. 2000. Biotic invasions: causes, epidemiology, global consequences, and control.

Meimberg H, Milan NF, Karatassiou M, Espeland EK, McKay JK, Rice KJ. 2010. Patterns of introduction and adaptation during the invasion of *Aegilops triuncialis* (Poaceae) into Californian serpentine soils. Molecular Ecology. 19(23): 5308–5319.

Michigan Department of Natural Resources. 2015. 2015 Michigan Invasive Species Grant Program Handbook.

Moritz C. 1994. Defining 'Evolutionarily Significant Units.' Trends in Ecology and Evolution. 9(10): 373–375.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 4(3): 535–538.

Palsbøll PJ, Bérubé M, Allendorf FW. 2007. Identification of management units using population genetic data. Trends in Ecology and Evolution. 22(1): 11–16.

Peakall R, Smouse PE. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 6(1): 288–295.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 28(19): 2537–2539.

Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics. 52(3): 273–288.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus data. Genetics. 155: 945–959.

Pyšek P. 1995. On the terminology used in plant invasion studies. In: Plant lnvasions - General Aspects and Special Problems. Amsterdam, The Netherlands: SPB Academic Publishing. p. 71–81.

Pyšek P, Jarošík V, Müllerová J, Pergl J, Wild J. 2008. Comparing the rate of invasion by *Heracleum mantegazzianum* at continental, regional, and local scales. Diversity and Distributions. 14(2): 355–363.

R Development Core Team. 2017. R: A language and environment for statistical computing.

Raymond M, Rousset F. 1995. An exact test for population differentiation. Evolution. 49(6): 1280–1283.

Reid ML, Emery SM. 2018. Scale-dependent effects of *Gypsophila paniculata* invasion and management on plant and soil nematode community diversity and heterogeneity. Biological Conservation. 224: 153–161.

Reinhold CG. 1856. The farmer's promotion book. Pittsburgh, PA: W.S. Haven.

Rice E. 2018. Assessment of invasive *Gypsophila paniculata* control methods in the northwest Michigan dunes. Masters Theses. 888.

Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources. 8(1): 103–106.

Shih JG, Finkelstein SA. 2008. Range dynamics and invasive tendencies in *Typha latifolia* and *Typha angustifolia* in eastern North. Wetlands. 28(1): 1–16.

Stevens OA. 1957. Weights of Seeds and Numbers Per Plant. Weeds. 5(1): 46–55.

Swearingen J, Bargeron C. 2016. Invasive plant atlas of the United States. University of Georgia Center for Invasive Species and Ecosystem Health. invasiveplantatlas.org.

Weir BS, Cockerham CC. 1983. Estimating F-statistics for the analysis of population structure. Evolution. 38(6): 1358.

Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E. 1998. Quantifying threats to imperiled species in the United States. BioScience. 48(8): 607–615.

Williamson M, Fitter A. 1996. The varying success of invaders. Ecology. 77(6): 1661–1666.

Zalewski A, Michalska-Parda A, Bartoszewicz M, Kozakiewicz M, Brzeziński M. 2010. Multiple introductions determine the genetic structure of an invasive species population: American mink *Neovison vison* in Poland. Biological Conservation. 143(6): 1355–1363.

TABLES

Table 1. Locations, dates, sample size, and geographic coordinates for contemporary samples of *Gypsophila paniculata*.

Sampling Location	Sampling Code	GPS Coordinates	Sampling Date	*n*
Chelan, WA	CH-WA	47.7421˚N 120.2177˚W	June 7-8, 2018	20
Osborne Bay, WA	OB-WA	47.9129˚N 119.0433˚W	June 7, 2018	16
Knife River Indian Villages National Historic Site, ND	KR-ND	47.3302˚N 101.3859˚W	June 6, 2018	14
Ottertail, MN	OT-MN	46.4627˚N 95.5733˚W	June 11, 2018	15
Petoskey State Park, MI	PS-MI	45.4037˚N 84.9121˚W	June 1, 2018	20
Dune Plateau, Sleeping Bear Dunes National Lakeshore, MI	SBD-MI	44.8731˚N 86.0585˚W	July, 2016	30
Arcadia Dunes, Sleeping Bear Dunes National Lakeshore, MI	AD-MI	44.5366˚N 86.2253˚W	July 8 and 15, 2016	30

Table 2. Genetic diversity measures for seven contemporary *Gypsophila paniculata* sampling locations sequenced at 14 microsatellite (nSSR) loci.

	Sampling Locations						
	CH-WA	OB-WA	KR-ND	OT-MN	PS-MI	SBD-MI	AD-MI
Loci							
BB_21680							
N	20	15	14	15	19	30	30
N_A	4	3	1	2	2	3	3
H_O	0.400	0.267	0.000	0.467	0.474	0.500	0.700
H_E	0.599	0.646	0.000	0.370	0.491	0.549	0.555
F_{IS}	0.3377	0.5957	-	-0.2727	0.0357	0.0909	-0.2661
BB_6627							
N	20	16	14	15	20	30	30
N_A	1	1	1	1	1	2	2
H_O	0.000	0.000	0.000	0.000	0.000	0.500	0.467
H_E	0.000	0.000	0.000	0.000	0.000	0.503	0.472
F_{IS}	-	-	-	-	-	0.0068	0.0122
BB_3968							
N	20	16	14	15	20	30	30
N_A	1	1	1	1	2	4	2
H_O	0.000	0.000	0.000	0.000	0.150	0.367	0.133
H_E	0.000	0.000	0.000	0.000	0.219	0.421	0.183
F_{IS}	-	-	-	-	-0.0556	0.1320	0.2750
BB_5151							
N	20	16	14	15	20	30	28

N_A	2	2	2	1	2	2	2
H_O	0.150	0.063	0.357	0.000	0.100	0.467	0.179
H_E	0.142	0.063	0.389	0.000	0.097	0.499	0.508
F_{IS}	-0.0556	0.0000	0.0845	-	-0.0270	0.0667	0.6530
BB_4443							
N	20	16	14	15	20	30	30
N_A	1	3	4	1	4	9	5
H_O	0.000	0.563	0.429	0.000	0.450	0.767	0.567
H_E	0.000	0.558	0.516	0.000	0.562	0.771	0.675
F_{IS}	-	-0.0075	0.1746	-	0.2028	0.0052	0.1623
BB_31555							
N	20	16	13	15	20	30	30
N_A	1	2	2	2	1	4	3
H_O	0.000	0.500	0.462	0.400	0.000	0.600	0.467
H_E	0.000	0.484	0.443	0.460	0.000	0.624	0.554
F_{IS}	-	-0.0345	-0.0435	0.1340	-	0.0396	0.1603
BB_14751							
N	20	16	14	15	20	30	30
N_A	5	4	3	2	3	8	6
H_O	0.750	0.563	0.500	0.333	0.500	0.633	0.467
H_E	0.726	0.619	0.521	0.370	0.472	0.782	0.631
F_{IS}	-0.0345	0.0940	0.0421	0.1026	-0.0615	0.1933	0.2632
BB_3335							
N	20	16	14	13	19	30	30
N_A	2	3	1	2	3	7	6

H_O	0.000	0.000	0.000	0.538	0.368	0.667	0.600
H_E	0.097	0.492	0.000	0.508	0.534	0.831	0.721
F_{IS}	1.0000	1.0000	-	-0.0633	0.3505	0.2000	0.1707
BB_4258							
N	20	16	14	15	20	30	30
N_A	2	2	1	1	1	2	2
H_O	0.350	0.063	0.000	0.000	0.000	0.033	0.300
H_E	0.450	0.063	0.000	0.000	0.000	0.033	0.345
F_{IS}	0.2267	0.0000	-	-	-	0.0000	0.1329
BB_3913							
N	20	15	13	15	20	30	30
N_A	2	3	2	2	3	4	2
H_O	0.050	0.200	0.077	0.333	0.150	0.667	0.467
H_E	0.050	0.191	0.077	0.287	0.145	0.588	0.452
F_{IS}	0.0000	-0.0500	0.0000	-0.1667	-0.0364	-0.1373	-0.0331
BB_2888							
N	20	16	14	15	20	30	30
N_A	2	2	2	2	2	5	5
H_O	0.550	0.375	0.286	0.267	0.450	0.833	0.667
H_E	0.481	0.484	0.254	0.405	0.512	0.807	0.599
F_{IS}	-0.1484	0.23080	-0.1304	0.3488	0.1231	-0.0335	-0.1154
BB_5567							
N	19	16	14	15	20	30	30
N_A	3	3	2	3	3	4	5
H_O	0.842	0.563	0.500	0.600	0.550	0.667	0.767
H_E	0.681	0.599	0.495	0.549	0.612	0.614	0.728

F_{IS}	-0.2441	0.0625	-0.0111	-0.0957	0.1030	-0.0872	-0.0545
BB_7213							
N	20	16	14	15	19	30	30
N_A	1	1	2	2	2	3	3
H_O	0.000	0.000	0.286	0.467	0.105	0.500	0.667
H_E	0.000	0.000	0.254	0.370	0.102	0.575	0.644
F_{IS}	-	-	-0.130	-0.2727	-0.0286	0.1317	-0.0366
BB_8681							
N	19	16	14	15	19	30	30
N_A	3	3	2	2	3	4	3
H_O	0.368	0.500	0.357	0.467	0.316	0.400	0.600
H_E	0.383	0.476	0.495	0.480	0.562	0.464	0.445
F_{IS}	0.0382	-0.0526	0.2857	0.0297	0.4447	0.1397	-0.3558

Notes: *N* number of individuals, N_A number of alleles per locus, H_O observed heterozygosity, H_E expected heterozygosity, F_{IS} inbreeding coefficient (Weir and Cockerham 1984). Sampling location codes: Chelan, WA (CH); Osborne Bay, WA (OB); Knife River Historic Indian Villages, ND (KR); Otter Tail, MN (OT); Petoskey State Park, MI (PS); Sleeping Bear Dunes National Lakeshore, MI (SBD); Arcadia Dunes, MI (AD).

Table 3. Population pairwise FST (Weir and Cockerham, 1984) for contemporary *Gypsophila paniculata* populations using microsatellite data calculated in GenAlEx 6.502 (Peakall and Smouse, 2006, 2012) running 9,999 permutations. Darker colors indicate increasing (higher) values; all values are significant with p-values <0.05.

Sampling location codes: Chelan, WA (CH-WA); Osborne Bay, WA (OB-WA); Knife River Historic Indian Villages, ND (KR-ND); Ottertail, MN (OT-MN); Petoskey State Park, MI (PS-MI); Sleeping Bear Dunes National Lakeshore, MI (SBD-MI); Arcadia Dunes, MI (AD-MI).

	CH-WA	OB-WA	KR-ND	OT-MN	PS-MI	SBD-MI	AD-MI
CH-WA	–						
OB-WA	0.077	–					
KR-ND	0.188	0.141	–				
OT-MN	0.124	0.104	0.194	–			
PS-MI	0.111	0.075	0.150	0.094	–		
SBD-MI	0.202	0.131	0.201	0.196	0.173	–	
AD-MI	0.192	0.153	0.188	0.168	0.160	0.070	–

Table S1. *Gypsophila paniculata* herbarium records used in this study.

Institution	Catalog #	Collection Date	GPS Coordinates	Location Information Provided	State/Province
Arizona State Univ. Vascular Plant Herbarium	ASU0080637	7/31/2013	41.301038, -105.570631	Laramie Basin	WY(USA)
B. A. Bennett Herbarium, Yukon Government	BABY-0160	7/9/1991	49.5833, -119.65	Summerland	BC(CA)
B. A. Bennett Herbarium, Yukon Government	BABY-6662	7/26/2008	49.18418, -119.535292	Osoyoos	BC(CA)
Boise District Bureau of Land Management	1461	10/12/1995		9 km NE of Weiser	ID(USA)
Boise State Univ., Snake River Plains Herbarium	49505	8/1/1972		Cardston	AB(CA)
Boise State Univ., Snake River Plains Herbarium	35193	7/1/2007	43.724433, -115.604067	Loftus Hot Springs	ID(USA)
Boise State Univ., Snake River Plains Herbarium	54162	8/23/2013	47.702411, -116.802719	Coeur d'Alene	ID(USA)
Brigham Young Univ., S.L Welsh Herbarium	BRYV0140072	6/22/2012	40.23994, -109.01077	Dinosaur, Rio Blanco	CO(USA)
Brigham Young Univ., S.L Welsh Herbarium	BRYV0092109	7/31/2011	46.00617, -112.61569	Silver Bow	MT(USA)
Brigham Young Univ., S.L Welsh Herbarium	BRYV0030863	8/15/2011	40.38927, -109.79833	Uintah	UT(USA)
Canadian Museum of Nature	CAN 450828	8/22/1980	43.533333, -79.633333	Mississauga Lorne Park	ON(CA)
Carnegie Museum of Nat. History Herbarium	CM195622	6/30/1956	44.686204, -85.512464	7.5mi SE of Traverse City	MI(USA)
Carnegie Museum of Nat. History Herbarium	CM462845	7/26/1967		Little Manistee River Crossing on Route 37	MI(USA)
Carnegie Museum of Nat. History Herbarium	CM195621	7/8/1966		10mi W of Coronport	SK(CA)
Central Michigan Univ.	CMC00019957	7/27/2015		Beaver Island, Whiskey Point lighthouse, St James	MI(USA)
Clemson Univ. Herbarium	6157	7/6/1928		Anderson	SC(USA)
Colorado State Univ. Herbarium	9072	8/20/1974	37.438, -105.7597		CO(USA)
Colorado State Univ. Herbarium	48075	7/22/1982	40.6796, -107.4408	Moffat County	CO(USA)
Colorado State Univ. Herbarium	71428	8/14/1984	40.9955, -104.9148	Weld County	CO(USA)
Colorado State Univ. Herbarium	72900	7/15/1989	40.5684, -105.0267	Fort Collins	CO(USA)
Consortium of California Herbaria	UC1714554	8/1907		Cisco, Placer	CA(USA)
Consortium of California Herbaria	UC455027	9/24/1909	35.30012, -120.66232	San Luis Obispo	CA(USA)
Consortium of California Herbaria	CASBOTBC388473	7/1912		Yrkeka, Siskiyou	CA(USA)
Consortium of California Herbaria	UCD98413	7/25/1950		Dorris, Siskiyou	CA(USA)
Consortium of California Herbaria	CDA3427	7/29/1953		McDoel, Tule Lake, Siskiyou	CA(USA)
Consortium of California Herbaria	CDA3425	7/16/1963	40.32005, -120.53503	Janesville, Lassen	CA(USA)
Consortium of California Herbaria	CASBOTBC388470	6/29/1967		SW part of Weed, Siskiyou	CA(USA)
Consortium of California Herbaria	CDA3428	6/23/1971		Benton Station, Mono County	CA(USA)
Consortium of California Herbaria	CDA3429	10/6/1971		1mi N of Janesville, Lassen	CA(USA)
Consortium of California Herbaria	CDA3426	5/15/1972		Orosi, Tulare	CA(USA)
Consortium of California Herbaria	UCSB39545	9/2/1981	34.42200, -119.79500	Santa Barbara	CA(USA)
Consortium of California Herbaria	CDA34391	6/17/1987		Janesville, Lassen	CA(USA)
Consortium of California Herbaria	CDA35529	8/15/1991		Stanislaus	CA(USA)
Consortium of California Herbaria	RSA719893	7/29/2006	40.31370, -120.53863	Janesville, Lassen	CA(USA)
Consortium of California Herbaria	RSA820288	5/8/2014	33.36120, -117.32250	Camp Pendelton North	CA(USA)
Eastern Michigan Herbarium	EMC010873	7/28/1976		Lapeer	MI(USA)
Eastern Michigan Herbarium	EMC010872	8/1894		Geneva	NY(USA)
Gouvernement du Québec	QUE0139003	7/7/1960		Rimouski	QC(CA)
Harvard Univ. Herbarium	691948	6/30/1938		Danbury, CT	CT(USA)
Harvard Univ. Herbarium	691945	8/10/1916		Westmore, Maine	ME(USA)
Harvard Univ. Herbarium	691946	7/18/1967		Burlington, VT	VT(USA)

Harvard Univ. Herbarium	691947	8/5/1967		Colchester, VT	VT(USA)
Hope College	HCHM01972	7/24/1978		West end of Crystal Lake, Benzie	MI(USA)
Illinois Natural History Survey	7546	6/17/1939		Starved Rock Park, La Salle	IL(USA)
Illinois Natural History Survey	93696	7/13/1963		Kankakee	IL(USA)
Illinois Natural History Survey	158788	6/18/1977		Mason County	IL(USA)
Illinois State Museum Herbarium Collection	14598	7/9/1940		Winnebago County	IL(USA)
Illinois State Museum Herbarium Collection	53098	6/23/1957		Mason County	IL(USA)
Illinois State Museum Herbarium Collection	57134	6/21/1959		Cook County	IL(USA)
iNaturalist Observations		4/12/2016	36.032, -90.44027	Greene	AR(USA)
Intermountain Herbarium	UTC00212261	9/12/1975		West end of Craig	CO(USA)
Intermountain Herbarium	UTC00110332	7/24/1958		Logan. Cache	UT(USA)
Intermountain Herbarium	UTC00240481	10/14/2004	37.9085, -111.3768	Garfield	UT(USA)
Kathryn Kalmbach Herbarium	KHD00013339	8/4/1975	39.740063, -105.512601	Clear Creek County	CO(USA)
Kathryn Kalmbach Herbarium	KHD00013340	7/18/1981	39.547561, -105.093572	Littleton	CO(USA)
Kathryn Kalmbach Herbarium	KHD00027068	8/12/2010	40.989279, -105.009321	Larimer County	CO(USA)
Klamath National Forest Herbarium		7/26/1978		Klamath Nat'l Forest, Siskiyou	CA(USA)
Louisiana State Univ., Shirley C. Tucker Herbarium	LSU00080268	8/4/1972		Custer	SD(USA)
Minot State Univ.	889	8/13/1963	48.2618, -101.4468	Burlington	ND(USA)
Missouri Botanical Garden	1663185	7/20/1987		Grand Junction, Mesa	CO(USA)
Missouri Botanical Garden	744953	8/17/1991	39.75, -105.66666		CO(USA)
Montana State Univ.	51309	7/26/1956	48.20178, -114.314	Kalispell	MT(USA)
Montana State Univ.	57531	7/20/1959	45.712572, -111.04224	Bozeman	MT(USA)
Montana State Univ.	60068	8/1/1960	45.981500, -112.519000	Deer Lodge, Silver Bow	MT(USA)
Montana State Univ.	63044	7/14/1967	46.988237, -114.18249	Missoula	MT(USA)
Montana State Univ.	63273	7/19/1967	48.77472, -104.56194	Plentywood	MT(USA)
Montana State Univ.	64029	7/19/1968	45.65579, -111.87232	Madison County	MT(USA)
Montana State Univ.	78364	7/19/1969	45.754509, -111.05906	Bozeman	MT(USA)
Montana State Univ.	65746	7/10/1970	48.7925, -105.42028	Scobey	MT(USA)
Montana State Univ.	65961	7/19/1971	47.71667, -104.15583	Sidney	MT(USA)
Montana State Univ.	125437	7/22/1999	47.574800, -112.338200	Teton County	MT(USA)
Montana State Univ.	78365	7/16/2001		Eddy Flat, Sanders	MT(USA)
Montana State Univ.	78564	6/24/2003	46.19389, -104.36944	Baker	MT(USA)
Montana State Univ.	79545	7/14/2005	46.596034, -112.02693	Helena	MT(USA)
Montana State Univ.	82151	7/31/2008	47.774443, -112.33899	Teton County	MT(USA)
Morton Arboretum	0013059MOR	7/5/1974		Kane	IL(USA)
Morton Arboretum	0013060MOR	7/12/1992		St. Joseph	IN(USA)
Muhlenberg College	MCA0012438	8/9/1963		Lehigh, West Bethlehem	PA(USA)
Muhlenberg College	MCA0012437	6/18/1964		Lehigh, West Bethlehem	PA(USA)
Muhlenberg College	MCA0012436	6/28/1964		Lehigh, West Bethlehem	PA(USA)
Murray State Univ. Herbarium	12357R	8/12/1972	46.699720, -92.001390	South Range	WI(USA)
National Museum of CA, Flora of New Brunswick	50157	8/5/2010	46.50, -66.75	Lawrence, New Brunswick	NB(CA)
Nevada Dept. of Agriculture Herbarium	NDOA0085	9/9/1967		Washoe, 6mi S of Reno	NV(USA)
Nevada Dept. of Agriculture Herbarium	NDOA0082	6/23/1976		Washoe, Stewart Indian Colony, Carson City	NV(USA)
New York Botanical Garden	446359	8/27/1982		Mono County	CA(USA)
New York Botanical Garden	1104462	7/1/2007	43.724433, -115.604067	Loftus Hot Springs	ID(USA)
New York Botanical Garden	88097	8/13/1997	42.65, -103.98	Bowen	NE(USA)
New York Botanical Garden	446361	7/9/1978		Mottsville Cemetery, Douglas County	NV(USA)

New York Botanical Garden	446357	8/7/1986		White Pine County	NV(USA)
New York Botanical Garden	446362	7/21/1973		Pine Valley Campground, Washington County	UT(USA)
New York Botanical Garden	446360	7/17/1984		Washington County	UT(USA)
New York Botanical Garden	446358	8/13/1991	42.8732, -109.8512	Pinedale	WY(USA)
New York Botanical Garden	1192083	7/20/2001	43.3064, -110.6775	Jackson	WY(USA)
Northern KY U, John W. Theiret Herbarium	31973000024234	7/10/1976		Emmett	MI(USA)
Northern KY U, John W. Theiret Herbarium	31973000024236	8/5/1967		Chittenden	VT(USA)
OAC Herbarium	41438	8/10/1967		4mi E of Okotoks	AB(CA)
OAC Herbarium	25058	8/6/1962		1mi E of Fishe	SK(CA)
OAC Herbarium	40625	8/29/1963		Regina	SK(CA)
OAC Herbarium		9/17/1965		Eastend	SK(CA)
OAC Herbarium	58551	8/12/1986		Regina	SK(CA)
Oregon State Univ.	OSC241930	8/23/2013		Kootenai	ID(USA)
Oregon State Univ.	OSC233030	7/31/2011		Silver Bow	MT(USA)
Oregon State Univ.	OSC90946	11/5/1956	42.225, -121.7806	Klamath Falls	OR(USA)
Oregon State Univ.	OSC130826	7/28/1969	43.5864, -119.0531	Burns	OR(USA)
Oregon State Univ.	OSC130826	7/1969	43.5864, -119.0531	Burns	OR(USA)
Oregon State Univ.	OSC215439	9/2/2005	44.1461, -121.3322	Bend	OR(USA)
Oregon State Univ.	OSC241888	7/2012	43.5438, -119.084	Hines	OR(USA)
Pacific Lutheran Univ.	963	7/23/1972		E of Parkland, Pierce	WA(USA)
Pacific Northwest National Library	PNNL00903	7/20/1984		Hanford, Benton	WA(USA)
Pacific Northwest National Library	PNNL00902	6/14/1993		Benton	WA(USA)
PNW Herbarium, Western Washington Univ.	8621	1963		Sand Hills Region, 75mi S of Fargo	ND(USA)
PNW Herbarium, Western Washington Univ.	15487	7/18/1971		Winthrop, Okanogan	WA(USA)
Portland State Univ.	16759	7/19/1974	44.056012, -121.31584	Bend	OR(USA)
R. L. McGregor Herbarium	238214	7/21/1989		W. Moosejaw	SK(CA)
Robert F. Hoover Herbarium, Cal Poly State Univ.	5327	8/27/1963		Siskiyou	CA(USA)
Robert F. Hoover Herbarium, Cal Poly State Univ.	59388	7/22/1966	34.60500, -120.41700	Santa Barbara	CA(USA)
Robert F. Hoover Herbarium, Cal Poly State Univ.	74231	7/25/1991		Albany	WY(USA)
Rocky Mountain Herbarium		8/19/2011	43.63266, -113.29578	Arco	ID(USA)
Rocky Mountain Herbarium	166182	7/18/1934	48.3818, -114.0832		MT(USA)
Rocky Mountain Herbarium	454655	8/1/1960	45.9815, -112.519	Butte	MT(USA)
Rocky Mountain Herbarium	100847	9/3/1924	44.0418, -103.1309	Green Valley	SD(USA)
Rocky Mountain Herbarium	118304	7/10/1929	44.0748, -103.2221	Rapid City	SD(USA)
Rocky Mountain Herbarium	268305	7/12/1962		Platte	WY(USA)
Rocky Mountain Herbarium	322704	7/9/1978	44.4646, -105.5809	Campbell County	WY(USA)
Rocky Mountain Herbarium	329531	7/25/1980	41.1248, -104.8767	Laramie County	WY(USA)
Rocky Mountain Herbarium	329531	7/25/1980	41.1248, -104.8767	Laramie County	WY(USA)
Rocky Mountain Herbarium	361100	7/29/1982	44.4935, -109.2042	Buffalo Bill Reservoir	WY(USA)
Rocky Mountain Herbarium	524271	7/18/1983	44.4128, -105.55889	Campbell County	WY(USA)
Rocky Mountain Herbarium	609437	7/7/1984	44.4618, -109.483	Park County	WY(USA)
Rocky Mountain Herbarium	389281	6/21/1987	43.1061, -108.6264	Wind River Reservation	WY(USA)
Rocky Mountain Herbarium	704685	7/8/1994	41.5886, -104.9877	Laramie County	WY(USA)
Rocky Mountain Herbarium	783428	6/7/1995	43.7922, -108.3447	Hot Springs County	WY(USA)
Rocky Mountain Herbarium	600758	8/1/1995	42.7611, -104.4461	Lusk	WY(USA)
Rocky Mountain Herbarium	600758	8/1/1995	42.7611, -104.4461	Lusk	WY(USA)
Rocky Mountain Herbarium	653597	8/13/1997	44.4645, -109.406	Park County	WY(USA)

Rocky Mountain Herbarium		8/15/1998	41.2105, -106.7877	Encampment	WY(USA)
Royal British Columbia Museum	V075731	7/29/1964	50.019722, -113.582778	Claresholm	AB(CA)
Royal British Columbia Museum	V020711	6/27/1947	49.616667, -115.633333	Windermere	BC(CA)
Royal British Columbia Museum	V034707	7/10/1958	50.466667, -115.983333	Windermere	BC(CA)
Royal British Columbia Museum	V051012	7/15/1964		Spences Bridge, Thompson-Okanagan	BC(CA)
Royal British Columbia Museum	V134152	7/18/1964	49.616667, -115.633333	East Kootenay	BC(CA)
Royal British Columbia Museum	V170481	8/20/1966	49.616667, -115.633333	Windermere	BC(CA)
Royal British Columbia Museum	V060882A	7/8/1972	50.233333, -119.216667	Coldstream	BC(CA)
Royal British Columbia Museum	V104347	7/23/1972		Kamloops, Princeton, Thompson-Okanagan	BC(CA)
Royal British Columbia Museum	V104242	7/20/1975	49.350000, -120.066667	Okanagan-Similkameen	BC(CA)
Royal British Columbia Museum	V109355	8/16/1975		Cathedral Provincial Park, Thompson-Okanagan	BC(CA)
Royal British Columbia Museum	V126087	6/13/1984	49.183333, -119.550000	Oliver	BC(CA)
Royal British Columbia Museum	V181785	9/11/1989	49.083333, -119.516667	Haynes Lease Ecological Reserve	BC(CA)
Royal British Columbia Museum	V180190	9/12/1989	50.750000, -121.000000	Thompson-Nicola	BC(CA)
Royal British Columbia Museum	V179656	7/13/1991	49.233333, -119.820000	Okanagan-Similkameen	BC(CA)
Royal British Columbia Museum	V201741	7/29/2007	48.458333, -123.497222	Victoria	BC(CA)
San Juan College Herbarium	49926	6/10/1989		Salmon Ruins, San Juan	NM(USA)
Snow College Herbarium	EPHR 000496	4/18/1977		Provo	UT(USA)
South Dakota State U Herbarium	7569	7/29/1993		Eddy	ND(USA)
U of Minnesota, Bell Museum	108748	6/23/1896	44.984523, -93.177092	Falcon Heights	MN(USA)
Univ. of Alaska, Anchorage	4108	7/29/2004		Anchorage Quad	AK(USA)
Univ. of Alberta Museums	127283	7/27/1967	53.55, -113.5	Edmonton	AB(CA)
Univ. of Alberta Museums	127096	7/1/2010	53.101817, -111.5652	Kinsella	AB(CA)
Univ. of British Columbia, Beaty Herbarium	V155368	6/30/1958		Macleoud, Champ Vague	AB(CA)
Univ. of British Columbia, Beaty Herbarium	V155368	6/30/1958		Champ Vague, Macleod	AB(CA)
Univ. of British Columbia, Beaty Herbarium	V95748	1933	50, -119	Shuswap Lake, Sorrento	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V7797	6/21/1938		Erickson	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V7796	6/27/1947		Fort Steele	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V72193	9/15/1950	49.616667, -115.616667	Fort Steele	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V72193	9/15/1950	49.616667, -115.61666	East Kootenay	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V111996	7/18/1964	49.616667, -115.61666	Fort Steele	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V140005	7/2/1972	50, -121	Ashcroft	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V190420	6/19/1986	50.750000, -121.983333	Lillooet	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V195448	8/13/1988	49.283333, -122.75	Coquitlam	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V218663	9/12/1989	50.75, -121	Thompson-Okanagan, Walhachin	BC(CA)
Univ. of British Columbia, Beaty Herbarium	V7795	7/8/1933	48.573336, -118.08704	Columbia River Valley, Northport	WA(USA)
Univ. of CA, Riverside Plant Herbarium	UCR-11266	8/2/1970			BC(CA)
Univ. of Colorado Museum of Natural History	226746	7/10/1924	38.8338819, -104.8213631	Colorado Springs	CO(USA)
Univ. of Colorado Museum of Natural History	226753	7/14/1949	40.2082377, -105.1638622	Longmont	CO(USA)
Univ. of Colorado Museum of Natural History	226613	9/12/1975	40.5139078, -107.5587807	Craig	CO(USA)
Univ. of Colorado Museum of Natural History	226738	7/18/1981		Jefferson	CO(USA)
Univ. of Colorado Museum of Natural History	226605	7/20/1987		Grand Junction, Mesa	CO(USA)
Univ. of Colorado Museum of Natural History	226621	7/26/1989	40.0583166, -106.3755897	Kremmling	CO(USA)
Univ. of Colorado Museum of Natural History	964601	7/17/2009		Boulder	CO(USA)
Univ. of Colorado Museum of Natural History	1806686	6/22/2012	40.239944, -109.010778	Dinosaur	CO(USA)
Univ. of Idaho	19791	7/7/1940		4 miles S of Rathdrum, Kootenai	ID(USA)
Univ. of Idaho	90992	8/23/1986	42.051365, -111.39631	Bear Lake County	ID(USA)

Univ. of Idaho	164982	7/31/2011		Rocker, Silver Bow	MT(USA)
Univ. of Idaho	92071	8/7/1986		Egan Range, White Pine	NV(USA)
Univ. of Idaho	106119	8/13/1991		Penedale, Sublette	WY(USA)
Univ. of Lethbridge		6/27/1958		District de Medicine Hat	AB(CA)
Univ. of Lethbridge		8/4/1968		Fort Macleod	AB(CA)
Univ. of Lethbridge		7/21/1991	49, -111.95	Coutts, Warner	AB(CA)
Univ. of Manitoba Herbarium	49813	8/8/1948	49.595000, -99.683889	Wawanesa	MB(USA)
Univ. of Manitoba Herbarium	19178	08/06/1951	50.669167, -100.811111	Rossburn	MB(USA)
Univ. of Manitoba Herbarium	19180	7/24/1953	49.595000, -99.683889	Wawanesa	MB(USA)
Univ. of Manitoba Herbarium	19181	7/26/1953	49.667000, -99.960000	Brandon	MB(USA)
Univ. of Manitoba Herbarium	19179	8/3/1953	49.268000, -100.996000	Melita	MB(USA)
Univ. of Manitoba Herbarium	58123	8/6/1955	50.702600, -96.530400	Victoria Beach	MB(USA)
Univ. of Manitoba Herbarium	25162	7/18/1971	51.653, -100.4594	Near Garland	MB(USA)
Univ. of Manitoba Herbarium	27282	8/5/1972	50.702600, -96.530400	Reader Lake, The Pas	MB(USA)
Univ. of Manitoba Herbarium	42972	8/24/1979	50.816667, -100.368056	Whitewater Lake Camp Area	MB(USA)
Univ. of Manitoba Herbarium	45440	9/10/1982	53.938333, -101.341944	Reader Lake, The Pas	MB(USA)
Univ. of Manitoba Herbarium	44054	7/23/1985	49.67, -96.65	Richer	MB(USA)
Univ. of Manitoba Herbarium	25162	7/9/1986	50.020700, -100.440000	15km W of Rivers	MB(USA)
Univ. of Manitoba Herbarium	70727	6/29/1988	49.805556, -99.641667	CFB Shilo	MB(USA)
Univ. of Manitoba Herbarium	49771	7/18/1989	49.837042, -99.594542	Shilo M.R.	MB(USA)
Univ. of Manitoba Herbarium	57567	7/28/1993	49.763333, -99.676944	CFB Shilo	MB(USA)
Univ. of Manitoba Herbarium	71893	7/15/1994	50.525000, -96.583333	S of Grand Beach	MB(USA)
Univ. of Michigan Herbarium		7/16/1913		Emmet County	MI(USA)
Univ. of Michigan Herbarium		8/15/1915		Cheboygan County	MI(USA)
Univ. of Michigan Herbarium		7/13/1917		Cheboygan County	MI(USA)
Univ. of Michigan Herbarium	MICH1314642	07/23/1918		Cheboygan County	MI(USA)
Univ. of Michigan Herbarium	MICH1314641	08/12/1927		Cheboygan County	MI(USA)
Univ. of Michigan Herbarium		7/19/1945		Grand Rapids	MI(USA)
Univ. of Michigan Herbarium		7/7/1946		Washtenaw County	MI(USA)
Univ. of Michigan Herbarium		7/20/1947		Oakland County	MI(USA)
Univ. of Michigan Herbarium		8/5/1947		Leelanau County	MI(USA)
Univ. of Michigan Herbarium	MICH1314648	9/5/1949		Houghton County	MI(USA)
Univ. of Michigan Herbarium		7/13/1950		Jackson County, Leoni TWP	MI(USA)
Univ. of Michigan Herbarium	MICH1475294	07/25/1950		Emmet County	MI(USA)
Univ. of Michigan Herbarium		7/14/1951		Macomb County	MI(USA)
Univ. of Michigan Herbarium	MICH1314639	08/16/1951		Benzie County	MI(USA)
Univ. of Michigan Herbarium	MICH1314646	07/24/1952		Emmet County	MI(USA)
Univ. of Michigan Herbarium	MICH1314640	08/11/1953		Leelanau County	MI(USA)
Univ. of Michigan Herbarium	MICH1314635	06/30/1956		Grand Traverse County	MI(USA)
Univ. of Michigan Herbarium		7/13/1956		Wayne County	MI(USA)
Univ. of Michigan Herbarium		7/26/1967		Lake	MI(USA)
Univ. of Michigan Herbarium		8/4/1969		Schoolcraft County	MI(USA)
Univ. of Michigan Herbarium		6/12/1970		Newaygo County	MI(USA)
Univ. of Michigan Herbarium		6/12/1970		Wexford County	MI(USA)
Univ. of Michigan Herbarium		07/09/1971		Emmet County	MI(USA)
Univ. of Michigan Herbarium		7/10/1972		Oakland County	MI(USA)
Univ. of Michigan Herbarium	MICH1314634	07/28/1974		Schoolcraft County	MI(USA)
Univ. of Michigan Herbarium		7/24/1978		Benzie County	MI(USA)

Univ. of Michigan Herbarium		9/13/1981		Wexford County	MI(USA)
Univ. of Michigan Herbarium	MICH1314637	07/17/1983		Emmet County	MI(USA)
Univ. of Michigan Herbarium	MICH1314649	07/06/1984		Benzie County	MI(USA)
Univ. of Michigan Herbarium	MICH1314649	7/6/1984		Benzie County	MI(USA)
Univ. of Michigan Herbarium	MICH1314638	7/23/1984		Rogers TWP, Presque Isle County	MI(USA)
Univ. of Michigan Herbarium	MICH1314647	07/24/1984		Leelanau County	MI(USA)
Univ. of Michigan Herbarium		8/26/1984		Shiawassee County, Perry TWP	MI(USA)
Univ. of Michigan Herbarium	MICH1314636	6/24/1985		Lenawee CO, Raisin TWP	MI(USA)
Univ. of Michigan Herbarium	MICH1314645	07/29/1985		Benzie County	MI(USA)
Univ. of Michigan Herbarium	MICH1314644	07/14/1991		Crawford	MI(USA)
Univ. of Michigan Herbarium	MICH1314643	07/14/1991		Antrim County	MI(USA)
Univ. of Michigan Herbarium		7/30/1997		Schoolcraft County	MI(USA)
Univ. of Michigan Herbarium		9/12/2004		Lakefield TWP, Luce	MI(USA)
Univ. of Michigan Herbarium		7/9/2008		Pellston, Emmet County	MI(USA)
Univ. of Michigan Herbarium		7/15/2012		Ludington State Park, Mason County	MI(USA)
Univ. of Minnesota, Bell Museum	353430	7/14/1934	46.862181, -94.766121	Nevis	MN(USA)
Univ. of Minnesota, Bell Museum	396865	7/30/1948	47.473563, -94.880277	Bemidji	MN(USA)
Univ. of Minnesota, Bell Museum	554063	6/2/1955	44.759815, -95.421672	Hawk Creek Township	MN(USA)
Univ. of Minnesota, Bell Museum	594503	8/4/1958	46.922181, -95.058632	Park Rapids	MN(USA)
Univ. of Minnesota, Bell Museum	568923	7/15/1960	45.428063, -93.203997	Athens Township	MN(USA)
Univ. of Minnesota, Bell Museum	584548	7/8/1963	46.922181, -95.058632	Park Rapids	MN(USA)
Univ. of Minnesota, Bell Museum	590432	9/1/1965	47.282797, -95.212519	Itasca Township	MN(USA)
Univ. of Minnesota, Bell Museum	687348	6/27/1977	47.231866, -93.522768	Grand Rapids	MN(USA)
Univ. of Minnesota, Bell Museum	690447	7/1/1977	45.695467, -94.172414	Rice	MN(USA)
Univ. of Minnesota, Bell Museum	473827	7/13/1992	46.5338898, -94.7980576	Bullard Township	MN(USA)
Univ. of Minnesota, Bell Museum	460210	7/28/1992	47.229024, -94.633282	Cass County	MN(USA)
Univ. of Minnesota, Bell Museum	838412	8/7/1993	46.32431, -92.83477	Willow River Reservoir	MN(USA)
Univ. of Minnesota, Bell Museum	479774	7/16/2001	47.058131, -95.180836	Two Inlets	MN(USA)
Univ. of Minnesota, Bell Museum	920949	8/1/2003	46.129324, -94.720884	Turtle Creek Township	MN(USA)
Univ. of Minnesota, Bell Museum	440927	7/27/2004	46.199367, -94.38887	Morrison County	MN(USA)
Univ. of Minnesota, Bell Museum	907085	6/29/2006	46.933898, -95.351552	Carsonville Township	MN(USA)
Univ. of Minnesota, Bell Museum	924746	7/23/2008	47.1975, -94.9922222	Lake George	MN(USA)
Univ. of Mississippi, Thomas M. Pullen Herbarium	MISS0022741	7/13/1968		Flathead (2.5mi N of Bigfork)	MT(USA)
Univ. of Mississippi, Thomas M. Pullen Herbarium	MISS0022740	7/24/1971		McHenry (9.5mi N of Butte)	ND(USA)
Univ. of Montana	46057	7/22/1948	47.887446, -114.117614	Flathead Lake	MT(USA)
Univ. of Montana	52322	7/15/1956	46.872702, -113.986498	Missoula	MT(USA)
Univ. of Montana	66308	7/21/1968	48.063287, -114.072613	Bigfork	MT(USA)
Univ. of Montana	66309	7/26/1968	46.592712, -112.036109	Helena	MT(USA)
Univ. of Montana	136889	8/12/1970	46.828900, -111.820900		MT(USA)
Univ. of Montana	136890	8/12/1970	46.828900, -111.820900		MT(USA)
Univ. of Montana	75494	7/15/1973	46.233333, -114.18333	Ravalli County	MT(USA)
Univ. of Nevada Herbarium	20511	7/4/1970		Washoe, Stead	NV(USA)
Univ. of Nevada Herbarium	13998	7/9/1978		Douglas	NV(USA)
Univ. of North Carolina Chapel Hill Herbarium	NCU00100830	7/26/1989		Grand	CO(USA)
Univ. of North Carolina Chapel Hill Herbarium	NCU00100825	6/2/1955		Rennville	MN(USA)
Univ. of North Carolina Chapel Hill Herbarium	NCU00100834	7/22/1948		Lake	MT(USA)
Univ. of North Carolina Chapel Hill Herbarium	NCU00100829	7/18/1970		Morton	ND(USA)
Univ. of North Carolina Chapel Hill Herbarium	NCU00100815	6/19/1947		Lawrence	SD(USA)

Univ. of North Carolina Chapel Hill Herbarium	NCU00100820	8/4/1972		Custer	SD(USA)
Univ. of Puget Sound	8729	7/17/1973		Maryhill, Klickitat	WA(USA)
Univ. of Washington Herbarium	229695	8/1/1965		Center of Missoula	MT(USA)
Univ. of Washington Herbarium	186632	11/2/1956	42.225000, -121.780600	Klamath Falls, Klamath	OR(USA)
Univ. of Washington Herbarium	18742JWT	6/25/1931	47.83556, -120.04917	Chelan	WA(USA)
Univ. of Washington Herbarium	18741	8/16/1931	48.626036, -119.46626	Okanogan	WA(USA)
Univ. of Washington Herbarium	107930	7/2/1932	48.91611, -117.78056	Northport	WA(USA)
Univ. of Washington Herbarium	107960	6/17/1944	46.73139, -117.17861	Pullman	WA(USA)
Univ. of Washington Herbarium	173712	7/2/1952	47.7675, -117.35389	Mead	WA(USA)
Univ. of Washington Herbarium	242422	9/1/1969	48.85056, -117.38972	Metaline	WA(USA)
Univ. of Washington Herbarium	368375	7/31/2006	48.680278, -120.882500	Whatcom County	WA(USA)
Univ. of Washington Herbarium	413958	5/16/2013		Chelan	WA(USA)
Univ. of Washington Herbarium		6/15/2014	47.913005, -119.045792	Grant County	WA(USA)
Univ. of Washington Herbarium	399061	6/23/2014	47.816880, -119.975560	Chelan	WA(USA)
Univ. of Washington Herbarium	365203	9/19/2002	48.105000, -119.780000	Okanogan	WA(USA)
Univ. of WI-Madison, WI State Herbarium	v0025191WIS	7/10/1959		Adams	WI(USA)
Univ. of WI-Madison, WI State Herbarium	v0025200WIS	7/28/1960		Marinette	WI(USA)
Univ. of WI-Madison, WI State Herbarium	v0025204WIS	6/24/1964		Waupaca (2mi SSW of Rural)	WI(USA)
Univ. of WI-Madison, WI State Herbarium	v0025197WIS	9/12/1972		Wisconsin Point, Douglas	WI(USA)
Univ. of WI-Madison, WI State Herbarium	v0025192WIS	7/25/1975		Ashland	WI(USA)
Univ. of WI-Madison, WI State Herbarium	v0025203WIS	7/24/1981		Oconto	WI(USA)
Washington State Univ., Marion Ownbey Herbarium	49713	1/2/1929	47.658890, -117.425000	Spokane	WA(USA)
Washington State Univ., Marion Ownbey Herbarium	76350	7/26/1931	48.098330, -119.733060	Okanogan	WA(USA)
Washington State Univ., Marion Ownbey Herbarium	241720	6/2/1956	42.230560, -121.798330	Klamath Falls, Klamath	WA(USA)
Washington State Univ., Marion Ownbey Herbarium	334226	7/10/1973	48.541878, -120.378890	Okanogan	WA(USA)
Washington State Univ., Marion Ownbey Herbarium	333550	9/15/1974	46.323890, -117.971390	Dayton	WA(USA)
Washington State Univ., Tri-Cities	WS-TC-00115	1965		Yakima River, Benton	WA(USA)
Western IL Univ, R.M. Myers Herbarium	MWI00015585	6/18/1977		Bath, Mason	IL(USA)

FIGURE LEGENDS

Figure 1. Sampling locations for assessing *Gypsophila paniculata* population structure used in this study; locations in Washington, North Dakota, and Minnesota are visualized in panel (a), locations in Michigan are visualized in panel (b). The Leelanau Peninsula is denoted by a black star.

Sampling location codes: Chelan, WA (CH-WA); Osborne Bay, WA (OB-WA); Knife River Historic Indian Villages, ND (KR-ND); Ottertail, MN (OT-MN); Petoskey State Park, MI (PS-MI); Sleeping Bear Dunes National Lakeshore, MI (SBD-MI); Arcadia Dunes, MI (AD-MI).

Figure 2. Results of Bayesian cluster analysis contemporary *Gypsophila paniculata* individuals genotyped at 14 microsatellite loci, performed using the program STRUCTURE (Pritchard et al. 2000). Each individual (n=145) is represented by a single column, with different colors indicating the likelihood of assignment to that cluster. Black lines delineate sampling location. Results suggest 2 population clusters (K=2). Locations are listed from west to east and north to south (MI).

Sampling location codes: Chelan, WA (CH-WA); Osborne Bay, WA (OB-WA); Knife River Historic Indian Villages, ND (KR-ND); Ottertail, MN (OT-MN); Petoskey State Park, MI (PS-MI); Sleeping Bear Dunes National Lakeshore, MI (SBD-MI); Arcadia Dunes, MI (AD-MI).

Figure 3. Principal Coordinates Analysis (PCoA) of seven contemporary sampling locations of *Gypsophila paniculata* genotyped at 14 microsatellite loci, based on a genotypic distance matrix, and performed in GenAlEx 6.502 (Peakall and Smouse, 2006, 2012). Sampling location codes: Chelan, WA (CH-WA); Osborne Bay, WA (OB-WA); Knife River Historic Indian Villages, ND (KR-ND); Ottertail, MN (OT-MN); Petoskey State Park, MI (PS-MI); Sleeping Bear Dunes National Lakeshore, MI (SBD-MI); Arcadia Dunes, MI (AD-MI).

Figure 4. Discriminant analysis of principal components (DAPC) based on contemporary *Gypsophila paniculata* individuals analyzed at 14 microsatellite loci and calculated in the '*adegenet'* package for R (Jombart et al., 2010). (a) Scatterplot showing both discriminant function axes and eigenvalues. Each point represents an individual (n=145). After cross validation, 16 of 28 PC's were retained. (b) Plot visualizing DAPC sample distribution on the primary discriminant function. (c) Individual assignment to clusters using all eigenvalues explained by the PCA.

Sampling location codes: Chelan, WA (CH-WA); Osborne Bay, WA (OB-WA); Knife River Historic Indian Villages, ND (KR-ND); Ottertail, MN (OT-MN); Petoskey State Park, MI (PS-MI); Sleeping Bear Dunes National Lakeshore, MI (SBD-MI); Arcadia Dunes, MI (AD-MI).

Figure 5. Invasion curves created using herbarium specimen data for *Gypsophila paniculata* collection in (b) North America, (c) genetic cluster 1, and (d) genetic cluster 2 (a gap

in sample collection is evidenced by the lack of points on the graph). An example invasion curve illustrating the three-stage invasion pathway typical of many invasions is visualized in panel (a).

Cluster assignment: (1) Washington, North Dakota, Minnesota, and northwest Michigan. (2) Michigan south of the Leelanau Peninsula.

Figure S1. Bayesian cluster analysis of seven sampling locations of *Gypsophila paniculata* genotyped at 14 microsatellite loci, gathered from the program STRUCTURE (Pritchard et al. 2000). (a) Mean L(K) (±SD) over 10 runs for each value of K (1-9). (b) Evanno's ΔK (Evanno et al., 2005) where the highest rate of change indicates the highest likelihood of cluster numbers. This analysis was conducted without prior sampling location information.

Figure S2. Bayesian Information Criterion for a DAPC of seven sampling locations of *Gypsophila paniculata* genotyped at 14 microsatellite loci, created using the package '*adegenet'* in R (Jombart & Collins, 2015; Jombart et al., 2010). The inflection point suggests the supported amount of genetic clusters present; both a *K* of 2 and 3 were considered in analysis.

FIGURES

Figure 1

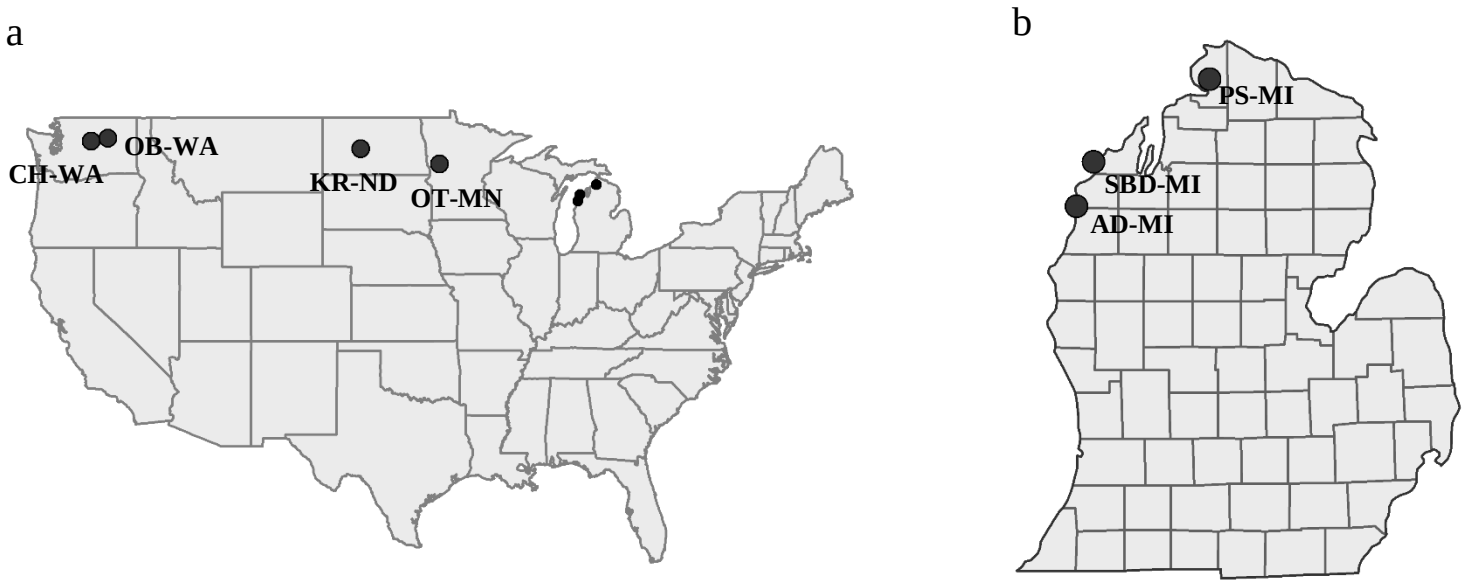


Figure 2

Figure 3

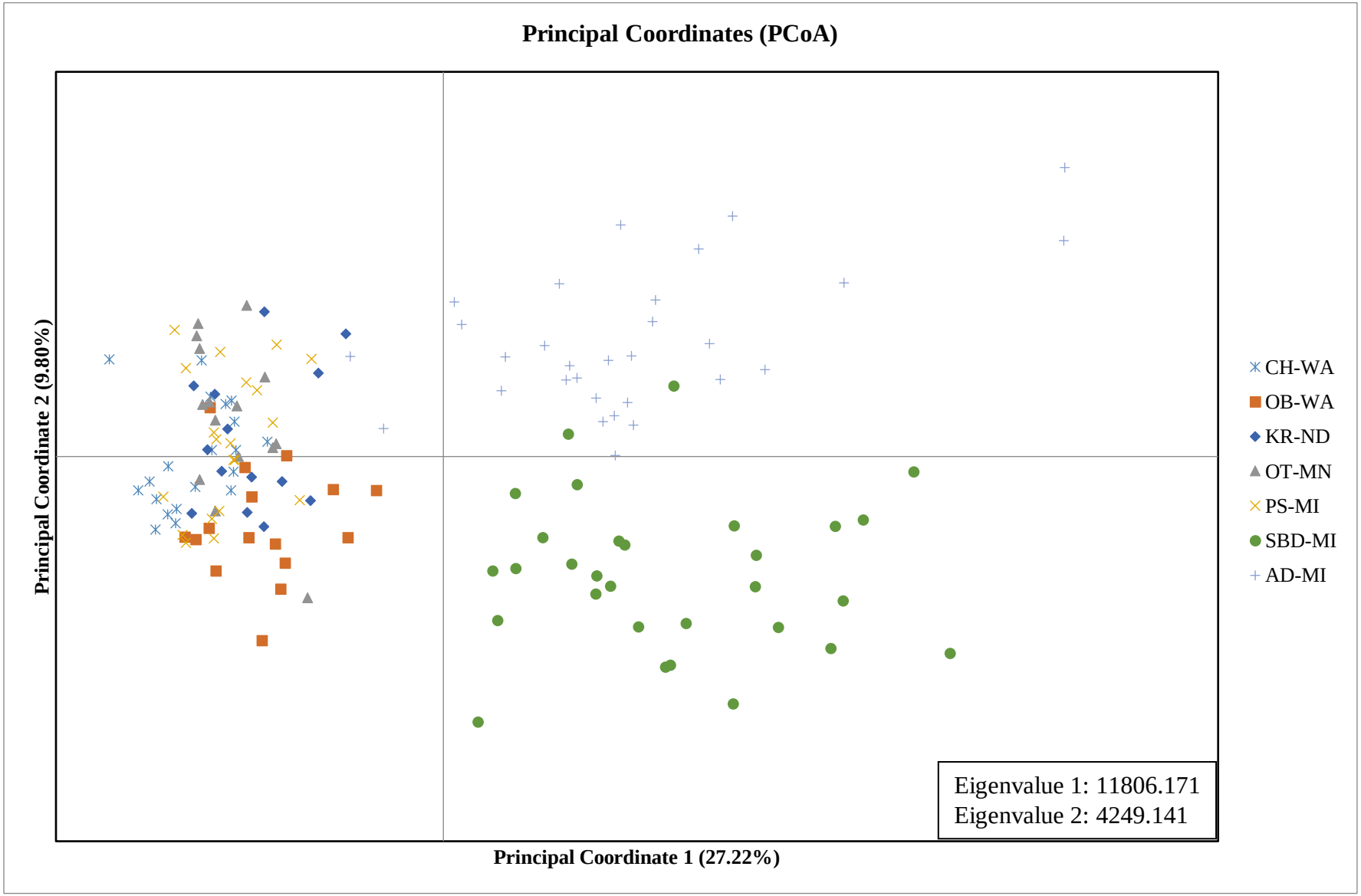
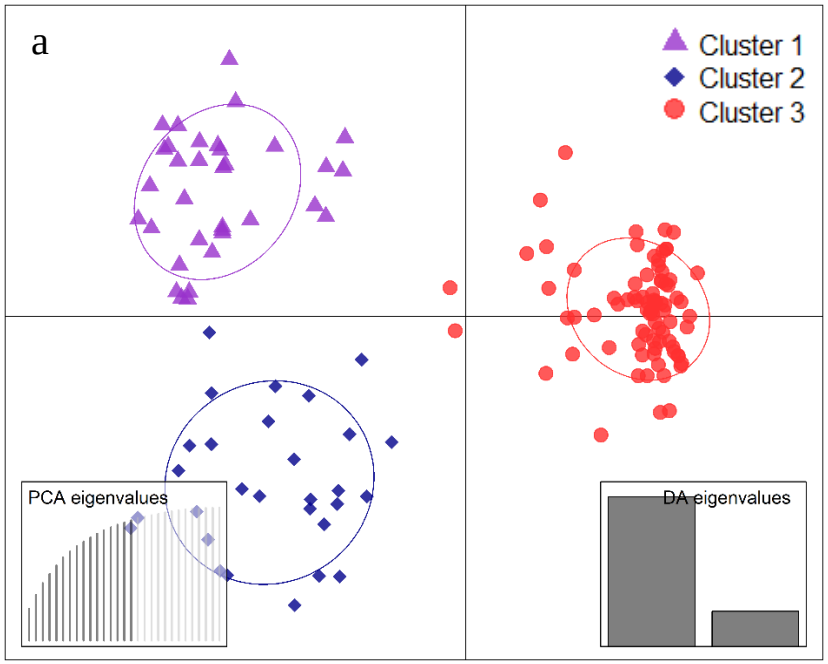
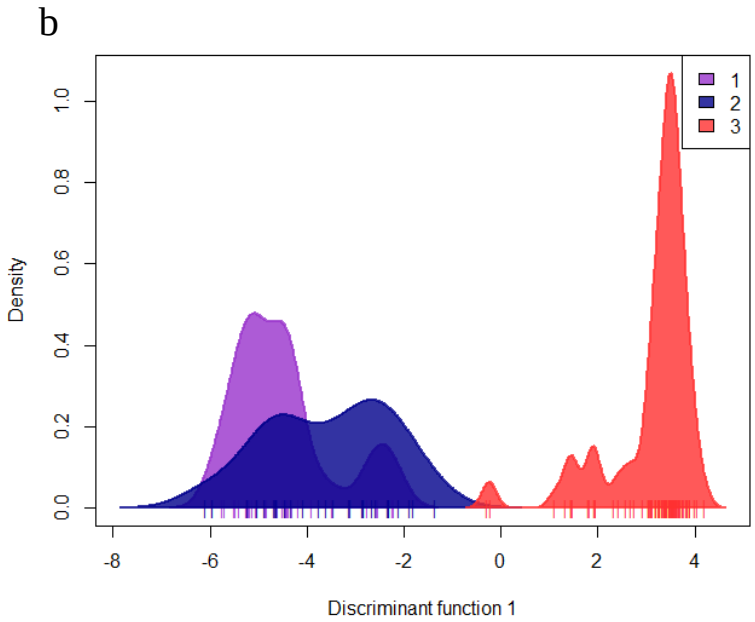

Principal Coordinates (PCoA)
Principal Coordinate 2 (9.80%)
Principal Coordinate 1 (27.22%)
CH-WA
OB-WA
KR-ND
OT-MN
PS-MI
SBD-MI
AD-MI
Eigenvalue 1: 11806.171
Eigenvalue 2: 4249.141

Figure 4





c

Population	Cluster 1 ▲	Cluster 2 ◆	Cluster 3 ●
CH-WA	0	0	20
OB-WA	5	0	11
KR-ND	0	0	14
OT-MN	0	0	15
PS-MI	0	0	20
SBD-MI	28	2	0
AD-MI	1	27	2

Figure 5

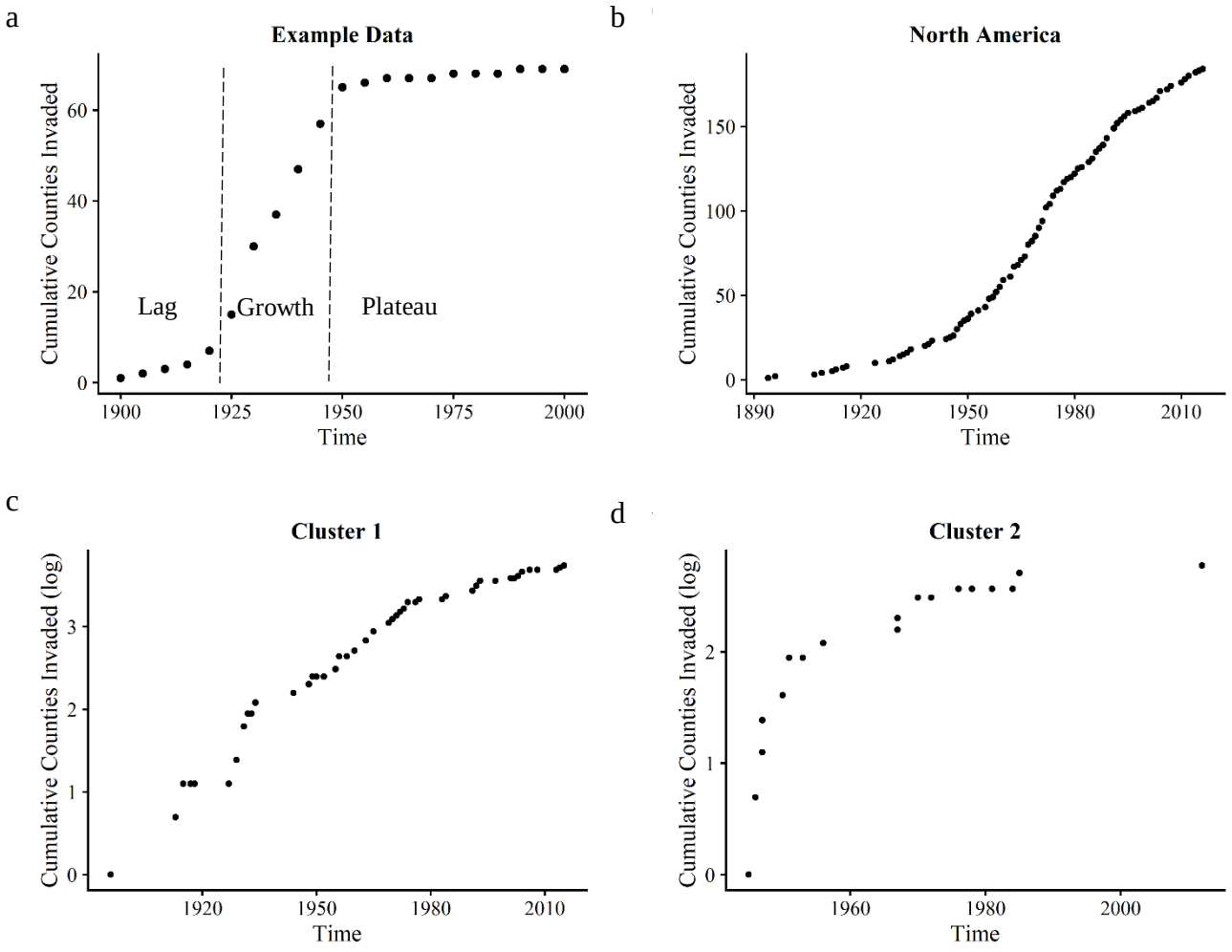
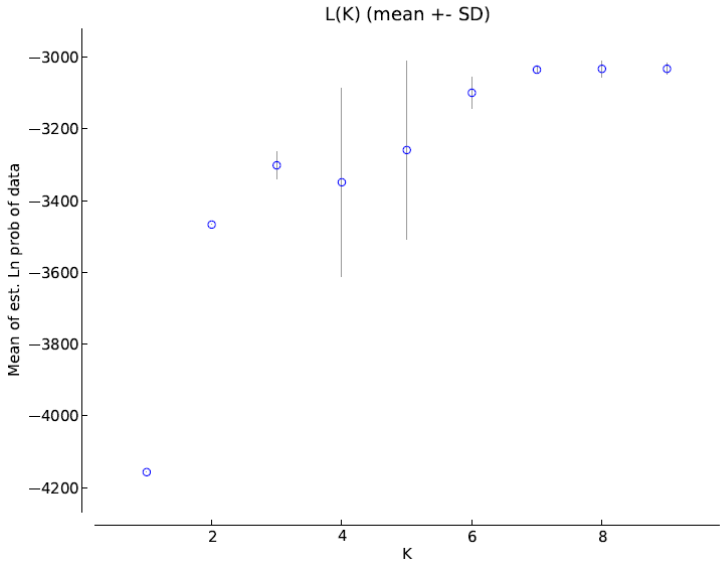

a
Example Data
Cumulative Counties Invaded
Lag
Growth
Plateau
Time
b
North America
Cumulative Counties Invaded
Time
c
Cluster 1
Cumulative Counties Invaded (log)
Time
d
Cluster 2
Cumulative Counties Invaded (log)
Time

Figure S1

a


L(K) (mean +- SD)
Mean of est. Ln prob of data
−3000
−3200
−3400
−3600
−3800
−4000
−4200
2
4
6
8
K

b

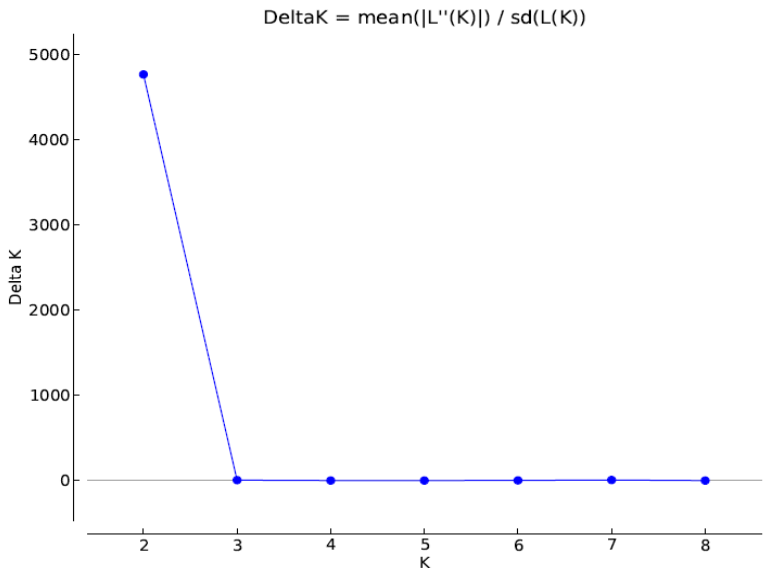
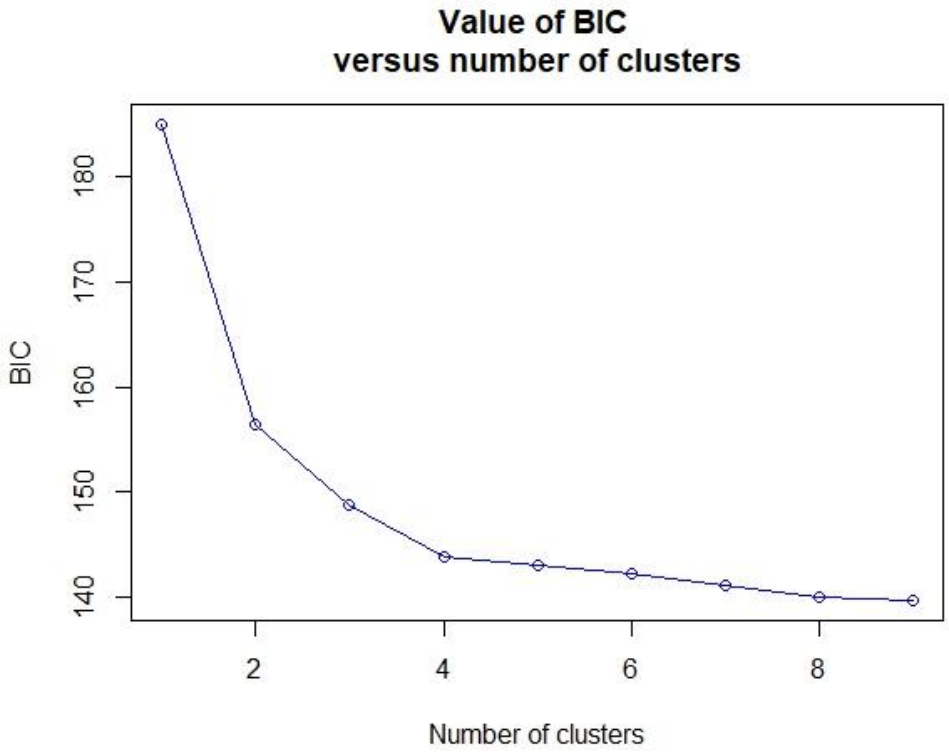

DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
5000
4000
3000
2000
1000
0
2
3
4
5
6
7
8
K

Figure S2


Value of BIC
versus number of clusters
BIC
140
150
160
170
180
2
4
6
8
Number of clusters

CHAPTER III

Manuscript

Using RNA-seq analysis and common garden growth trials to investigate potential adaptation of the invasive weed *Gypsophila paniculata* growing in distinct ecoregions

ABSTRACT

The ability of invasive species to succeed in environments in which they did not evolve has long been a topic of interest in the ecological community. Additionally, as global trade and transport increases, so do the cost and prevalence of invasive species. Despite the growing interest in invasive success, the mechanisms behind this phenomenon are not fully understood. This study investigated the possible mechanisms of success of the invasive forb *Gypsophila paniculata* (common baby's breath) growing in two distinct ecoregions in North America. In the spring of 2018, tissue samples were collected from *G. paniculata* populations growing in a sagebrush steppe in Chelan, Washington, USA (CHWA) and from a primary successional dune shore in northern Petoskey, Michigan, USA (PSMI). RNA-seq was used to sequence the transcriptomes for these populations. In addition, seeds were collected from these same populations and germination trials conducted for 12 days. We found a total of 1,149 genes were differentially expressed across all tissue types (root, stem, and leaf). Root tissue had the highest number of differentially expressed genes (8,135). Notably, biological processes overrepresented in PSMI were related to nutrient level homeostasis, particularly phosphate, potassium, and nitrogen starvation. Biological processes overrepresented in CHWA were related to calcium ion homeostasis, responses to heat and light intensity, response to water deprivation, and circadian rhythm. In germination trials seeds collected from CHWA were found to germinate significantly faster and demonstrate greater emergence than seeds collected from PSMI. No differences in above vs belowground tissue allocation were observed between populations growing in a common garden, though family effects were present. The combination of this transcriptomic and phenotypic data reveal responses suggesting that *G. paniculata* may be coping with the novel environmental stressors present in these divergent environments through alternate life history tactics related to early life history growth strategies and differential gene expression.

INTRODUCTION

The ability of invasive species to invade, adapt, and thrive in novel ecosystems has long been a focus of ecological research. Coined the "paradox of invasions", examining how invasive populations respond to novel environmental stressors after an assumed reduction in population size during introduction has become an entire field of scientific inquiry (Dlugosch et al. 2015; Sork 2018). More recently, this paradox has been called into question as research shows that while many invasive species may undergo a reduction in demographic size relative to their source population after an invasion event, this is not always linked with a subsequent reduction in genetic diversity (Frankham 2005; Dlugosch et al. 2015). Additionally, differences between the total genetic diversity of a population and the adaptive variation of a population can be large (McKay and Latta 2002; Leinonen et al. 2008). Due to the possible issues associated with using total genetic diversity as a measure of invasive potential, researchers are now assessing how local adaptation and phenotypic plasticity influences the fitness of invasive populations (Kawecki and Ebert 2004; Lande 2015; Sork 2018).

While local adaptation and phenotypic plasticity are independently viable strategies for coping with novel environmental stressors, they are not mutually exclusive (Kawecki and Ebert 2004; Lande 2015). Phenotypic plasticity can be adaptive, maladaptive, or neutral to an individual's fitness, and can occur independently of or in conjunction with changes in mean trait values (Ghalambor et al. 2007). Phenotypic plasticity can allow populations to persist through the sudden application of strong directional selection that often accompanies an introduction event. When the plastic response of the phenotype that is acted upon moves the trait value towards the new environment's optimum, the plasticity becomes adaptive (Conover and Schultz 1995; Van Tienderen 1997; Ghalambor et al. 2007). Ultimately, local adaptation results in a

phenotype that is more fit in its current range than that phenotype would be in other environments, and this shift in fitness can be the difference between persistence and perishing in introduced populations (Kawecki and Ebert 2004; Richards et al. 2006).

In the study of invasive species, the ability to parse out plastic and adaptive responses is often limited by the relative lack of background genetic data available for the species. However, with the development of technologies like RNA-seq, which allows for the assembly of transcriptomes *de novo*, gene expression data have become more widely available for use in non-model systems (Wang et al. 2009; Sork 2018). RNA-seq derived gene expression data can be used to answer questions related to genotype-by-environment (GxE) interactions, which are often used as proxy for phenotypic plasticity in populations located in discrete environments (Via and Lande 2006; Lande 2015; Sork 2018). These gene expression profiles allow researchers to estimate the prevalence of phenotypic plasticity in response to novel environmental stressors and to answer questions about how invasive species may be adapting in their introduced environments (Lande 2015; Sork 2018). A more classic approach to quantifying the genetic basis of phenotypic differences is the common garden growth experiment (Langlet 1971). Particularly useful for sessile organisms like plants, this approach allows researchers to parse out which phenotypic differences are a function of the individual's environment and which may have a genetic architecture behind them (Langlet 1971; Blumenthal and Hufbauer 2007; Colautti et al. 2009).

Gypsophila paniculata (common baby's breath) is a perennial forb native to Eurasia, originally introduced into North America in the late 1800's for its use in the floral industry (Darwent and Coupland 1966; Darwent 1975). After introduction, *G. paniculata* quickly spread and can now be found growing in diverse ecosystems across North America, often outcompeting

and crowding out the native species that live there (Baskett et al. 2011; Rice 2018). *Gypsophila paniculata* does this with the aid of a characteristic taproot which can grow to be over 3m deep (Barkoudah 1962; Darwent and Coupland 1966); the plant also produces as many as 13,000 seeds/year, allowing it to quickly form dense stands (Stevens 1957). *Gypsophila paniculata* is considered a priority invasive by The Nature Conservancy in Michigan (Emery and Doran 2013; Swearingen and Bargeron 2016) and a widespread and noxious weed in Washington and California (USDA 2019).

In chapter II of this thesis, we genotyped *G. paniculata* growing in seven locations across its North American introduced range at 14 microsatellite (nSSR) loci. These sampling locations consisted of varied habitats, such as sagebrush steppes in central Washington, prairie in North Dakota, and quartz-sand dune shores in the Great Lakes Basin. While many of the sampling locations were both climatically distinct and geographically separated, two genetic clusters were formed among all seven locations. To examine how *G. paniculata* populations respond to distinct environments we chose two populations that occur in distinct ecosystems (Michigan sand dunes and Washington sagebrush steppe), yet clustered together based upon genetic analysis. We conducted habitat characterization, soil analyses, RNA-seq analysis, and common garden germination and growth trials to explore differences between these populations in relation to their distinct habitats.

METHODS

Study Site Characterization

Petoskey, Michigan (PSMI) is a state park located along Lake Michigan's primary successional quartz-sand dune system. Vegetation is sparse and is chiefly comprised of *Ammophila*

breviligulata (dune grass), *Silene vulgaris* (bladder campion), *Juniper horizontalis* (creeping juniper) and *J. communis* (common juniper), and *Cirsium pitcher* (Pitcher's thistle). Herbarium records indicate that *G. paniculata* has been present in PSMI since at least 1913. Chelan, Washington (CHWA) is a disturbed habitat situated on slopes surrounding Lake Chelan and dominated by sagebrush (*Artemis* spp.). Herbarium records for CHWA suggest that *G. paniculata* has been present in the area since 1931. Average climate data for these two locations were collected from stations operated by the National Oceanic and Atmospheric Organization (NOAA) in Petoskey, MI and Entiat, WA (near Chelan, WA) and is summarized in Table 1.

Soil Analysis

In the spring of 2018, we collected soil samples from PSMI and CHWA (Table 2). Sampling locations differed in collection depths due to soil characteristics in CHWA that made deeper collection impossible (large boulders, hard soil). At both locations, we collected two sets of soil samples from all depths. In PSMI, we collected soil from 10cm, 50cm, and 1m, while in CHWA, we collected soil from 10cm, 25cm, and 50cm depths. We stored samples in airtight plastic bags and maintained them at 4˚C until analysis.

Using these soil samples, we conducted particle size analysis (PSA) via the sieve method, in which we dried, weighed, and washed soil through multiple size filters (1700, 1000, 500, 250, 125, and 63μm) (American Society for Testing and Materials Committee D18 on Soil and Rock 2004). We then oven-dried the sorted samples and weighed the final amount of soil caught in each filter size. For each location, we used soil samples taken from 10cm and 50cm depths for PSA.

We sent soil samples from all depths at both locations to A&L Great Lakes Laboratories (Fort Wayne, IN) for nutrient analysis. Samples were tested for: organic matter (%), phosphorus (P), potassium (K), magnesium (Mg), calcium (Ca), soil pH, total nitrogen (N), cation exchange capacity (CEC), and percent cation saturation of K, Mg, and Ca. At the laboratories, samples were dried overnight at 40˚C before being crushed and filtered through a 2 mm sieve. The following methods were then used for each analysis: organic matter content (loss on ignition at 360˚C). pH (pH meter), phosphorus, potassium, magnesium, and calcium content (Mehlich III Extraction and inductively coupled plasma mass spectrometry). Total nitrogen was determined using the Dumas method (thermal conductance). Results of nutrient analysis were explored using a principal component analysis (PCA) in the statistical program R v6.0 (R Development Core Team 2017).

RNA Extraction

We collected seedlings from CHWA and PSMI concurrently with soil samples (Table 2). We located *G. paniculata* seedlings separated by at least 2 meters to reduce the risk of sampling close relatives whenever possible. We then dissected seedlings into three tissue types (root, stem, and leaf), placed tissue in RNAlater™ (Thermo Fisher Scientific, Waltham, MA), and flash froze them in an ethanol and dry ice bath. Samples were kept on dry ice for transport and maintained at -80˚C until extraction in the lab.

We extracted total RNA from frozen tissue using a standard TRIzol® (Thermo Fisher Scientific) extraction protocol (Rio et al. 2010). We resuspended the extracted RNA pellet in DNase/RNase free water, before treating it with a DNA-Free Kit (Invitrogen, Carlsbad, CA). We assessed RNA quality with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and NanoDrop™ 2000 (Thermo Fisher Scientific). RIN (RNA Integrity Number) values for

individuals used in this study ranged from 6.1-8.3. However, because both chloroplast and mitochondrial rRNA's can artificially deflate RIN values in plant leaf tissue, we deemed these values to be sufficient for further analysis (see Babu C. V. and Gassmann, 2016). Finally, we submitted total RNA to the Van Andel Research Institute for cDNA library construction and sequencing.

cDNA Library Construction and Sequencing

Prior to sequencing, all samples were treated with a Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA). cDNA libraries were constructed using the Collibri Stranded Library Prep Kit (Thermo Fisher Scientific) before being sequenced on a NovaSeq 6000 (Illumina) using S1 and S2 flow cells. Sequencing was performed using paired end 2 x 100 bp format and produced approximately 60 million reads per sample, with 94% of reads having a Q-score >30.

Transcriptome Assembly

Prior to transcriptome assembly, read quality was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed and filtered using SortMeRNA (https://bioinfo.lifl.fr/RNA/sortmerna/), keeping only non-rRNA reads for downstream processing. A reference transcriptome was assembled *de novo* using Trinity v2.8.2 (Grabherr et al. 2011; Haas et al. 2013) with a normalized max read coverage of 100, a minimum k-mer coverage of 10, and k-mer size set to 32. The assembled transcriptome was annotated using Trinotate v3.1.1. and consisted of 223,810 genes and 474,313 transcripts from 59 samples. Data were filtered to exclude transcripts that were expressed less than 10 times or in fewer than 10 samples. Following filtering, 111,042 genes (49.61%) and 188,108 transcripts (39.66%) remained. Considering tissue type, 127,591 transcripts remained in the data from 20 root samples

(26.90%), 125,261 transcripts remained in the 19 samples from stem tissue (26.41%), and 112,499 transcripts remained in the 20 leaf tissue samples (23.72%).

Differential Expression

Differential expression was analyzed using the *edgeR* framework in R; to be considered significant, genes needed to have a p-value below 0.05 after false discovery rate correction and a log2 fold change greater than 2. For transcripts that were differentially expressed at the level of both tissue type and population, we identified Gene Ontology (GO) biological processes that were either over-represented or under-represented using the PANTHER classification system v14.1, where transcripts were assessed against the *Arabidopsis thaliana* genome (http://pantherdb.org/webservices/go/overrep.jsp).

Germination Trial

On August 11, 2018 we returned to our sample sites in CHWA and PSMI and collected seeds from 20 plants per location; Rice (2018) previously determined this collection date to yield over 90% seed germination for *G. paniculata* collected from Empire, MI. To collect seeds, we manually broke seed pods off and placed them inside paper envelopes in bags half-filled with silica beads. We stored bags in the dark at 20 to 23˚C until the germination trial began.

We counted one hundred seeds from twenty plants per population and placed them in a petri dish lined with filter paper (n = 2,000 seeds per population). We established a control dish using 100 seeds from the 'Early Snowball' commercial cultivar *(G. paniculata*) sold by W. Atlee Burpee & Co in 2018, known to have germination percentages in excess of 90%. Incubators had a 12:12h dark:light photoperiod and growth chamber conditions were set at 20˚C with 114 μmol m^{-2} s^{-1} photosynthetically active radiation from fluorescent light bulbs. Each day we randomized petri dish locations within the incubator to avoid bias in temperature or light regimes. We

conducted this study for fourteen days, at which point there had been no germination in any dish for two days. The same individual checked all seeds (n=4,100) daily within the same three-hour time window to minimize bias for germination, functionally defined as radicle emergence (Baskin and Baskin 2001). Once a seed had germinated, we removed it from the dish (method adapted from Rice, 2018).

Using the statistical program R v6.0, we fitted data to a nonparametric Kaplan-Meier time-to-event curve (McNair et al. 2012; R Development Core Team 2017). We then compared germination patterns between collection localities using a pairwise log-rank test (McNair et al. 2012). To test homogeneity within localities, we again conducted a log-rank test. Finally, to investigate the presence of family effects, we ran a series of pairwise log-rank tests with a Holm correction for multiple comparisons (McNair et al. 2012). For all analyses in this study, we set the alpha level to 0.05.

Common Garden Growth Trial

Finally, we conducted a common garden growth trial for seven weeks. Greenhouse conditions were set at 7:17 h dark:light photoperiod. Relative humidity and temperature settings during the day were 55% and 21˚C while nighttime conditions were 60% and 15.5˚C. We planted seeds on the same day to a standardized depth of 6 mm in a sand/potting soil mixture. Each day we watered plants until soil appeared fully wet and randomized plant position to prevent bias in temperature, light, or water regime. We used 6 seeds from 20 individuals per population (n = 120 per population, n = 240 total) for this trial to investigate both population level differences and potential family effects. At the end of the trial period, plants were carefully removed from the soil and the length of tissue above and below the caudex was measured with a caliper.

To compare emergence values between populations, we ran a two sided proportion test in the R statistical program (R Development Core Team 2017). We analyzed any difference in the ratio of above/belowground tissue between sampling locations and the presence or absence of family effects using a completely randomized design with subsampling ANOVA in SAS v9.4 (SAS Institute Inc. 2013). Analyses were run both with and without plants that did not emerge.

RESULTS

Habitat Characterization

Climate data collected from NOAA monitoring stations revealed differences in mean temperature and precipitation between our two sampling locations. CHWA had a 3˚C and 5˚C higher mean temperature in 2017 and 2018 than PSMI. Additionally, PSMI had greater rainfall in both 2017 (109.8 cm vs 38 cm) and 2018 (88.6 cm vs 27.8 cm).

PSA revealed that soils collected from PSMI primarily consisted of particles size 250-500 μm (avg. 83.95%), while soils collected from SPWA had a more even particle size distribution spanning all ranges, largely in part to the many rocks and pebbles present in the soil. Once rocks had been removed, roughly 20% of each sample consisted of particle sizes <125 μm. A PCA conducted on nutrient data collected from the soil showed that 87.6% of variation present could be explained by the first PC, with the second PC explaining only 7.6% of variation present (Figure 1) (Table S1). Soils collected from CHWA were characterized by higher levels of total nitrogen, phosphorus, magnesium, and potassium. In contrast, soils from PSMI had a higher pH and more available calcium.

Differential Gene Expression by Tissue Type

Root Tissue

There were 8,135 genes differentially expressed between CHWA and PSMI root tissue (Figure 2). Of those, 3,004 showed higher expression in PSMI, while 5,131 were more highly expressed in CHWA. The five most significant biological processes for each tissue type are detailed in Table 3, however, many more genes were significantly differentially expressed between sampling locations. Notably, genes that showed higher expression in PSMI were overrepresented in GO biological processes related to nutrient level homeostasis, particularly phosphate, potassium, and nitrogen starvation. PSMI also had an overexpression of genes related to positive regulation of flower development (Table 3). GO terms overrepresented by genes with increased expression in CHWA were related to calcium ion homeostasis, responses to heat and light intensity, response to water deprivation, and circadian rhythm (Table 3).

Stem Tissue

There were 5,374 differentially expressed genes in stem tissue collected from CHWA and PSMI (Figure 2). Of those, 2,393 genes were overrepresented in PSMI while 2,421 were overrepresented in CHWA. GO biological processes overrepresented in PSMI were related to biological processes like phosphate starvation, positive regulation of reproductive processes, and salt stress, while those overrepresented in CHWA were related to features such as high light intensity response and heat acclimation (Table 3).

Leaf Tissue

A total of 5,666 genes were differentially expressed between leaf tissues from PSMI and CHWA (Figure 2). Of those, 2,380 genes displayed higher expression in PSMI and enriched GO biological processes related to these genes included responses to high light intensity, responses to red and far red light, long-day photoperiodism, and flowering time regulation (Table 3). CHWA

had 3,286 genes with increased expression relative to PSMI. The enriched GO biological processes for CHWA included calcium ion homeostasis genes and high light intensity responses (Table 3).

All Tissue Types

There were 1,149 genes differentially expressed among all tissue types (Figure 2). Enriched GO biological processes related to these genes were associated with high light intensity response, cellular response to heat, and circadian rhythm.

Germination Trial

Results of a log-rank test comparing time-to-germination curves for each locality indicated strong statistical differences between seeds collected from PSMI and CHWA, with seeds from CHWA germinating more quickly ($p < 2.0 \times 10^{-16}$) (Figure 3). While there was a difference in germination curves, both localities reached 90% germination by the end of the germination trial. Log-rank tests looking at homogeneity within groups found strong statistical support for variation among time-to-germination curves for seeds from different plants for both populations (both $p < 2.0 \times 10^{-16}$). To investigate this, pairwise log-rank tests with a Holm correction for multiple comparisons showed that 38% of pairwise comparisons for CHWA seeds and 52% of pairwise comparisons for PSMI seeds showed a statistically significant difference in germination pattern among seeds collected from different parent plants.

Common Garden Growth Trial

A two-sided proportion test indicated a significant difference between total emergence in seeds collected from CHWA and PSMI, with CHWA seeds emerging more often than seeds from PSMI ($p<0.0002$) (Figure 4a). When excluding plants that did not emerge, ANOVA results

indicated no robust difference in the ratio of above/belowground tissue allocation between populations (p=0.605) (Figure 4b). However, significant family effects remained (p=0.0301) (Figure S1).

DISCUSSION

The ability of invasive species to adapt to novel environments over relatively short periods of evolutionary time is a process not yet fully understood. In this study, we analyzed two populations of the invasive weed *Gypsophila paniculata* for differences in gene expression and phenotype that may confer advantages in novel environments, particularly a quartz-sand dune shore in the Great Lakes Basin and a sagebrush steppe in central Washington's high desert. This was accomplished through initial habitat characterization, RNA-seq transcriptome analysis, germination trials, and common garden growth trials.

RNA-seq transcriptomes from populations of *G. paniculata* growing in CHWA and PSMI had 1,149 genes differentially expressed among all tissue types (root, stem, and leaf). Root tissue had the highest number of differentially expressed genes (8,135), which may indicate increased response to the environment for that tissue type relative to others. Soil differences between the populations were substantial, particularly in nutrient availability, and genes related to nutrient homeostasis can be found differentially expressed in the root tissue. CHWA had higher levels of K, PO_4, Mg, and N than PSMI, while PSMI had more available calcium than CHWA. Genes overrepresented in CHWA root tissue were related to calcium ion homeostasis, and PSMI root tissue had an overrepresentation of genes related to nutrient level homeostasis, particularly PO_4, K, and N starvation. Additionally, CHWA received less precipitation annually than PSMI, and CHWA root tissue had an overrepresentation of genes related to water

deprivation relative to PSMI. Stem tissue between the two populations revealed an overrepresentation of genes related to PO_4 starvation in PSMI, while CHWA stem tissue had an overrepresentation of genes related to high light intensity and heat responses. CHWA's mean temperature was 3˚C and 5˚C higher than PSMI's in 2017 and 2018. Finally, genes overrepresented in CHWA leaf tissue were related to high light intensity responses, a potential coping mechanism to the high elevation and subsequently increased UV present in CHWA relative to PSMI (elevation of 313m vs 182.6m). These differences in gene expression of *G. paniculata* found growing in distinct environments illustrate how despite relatively similar levels of background genetic diversity at neutral loci, plants are able to respond to different selection pressures.

Populations of *G. paniculata* growing in CHWA and PSMI displayed differences in their germination rate, with seeds collected from CHWA germinating significantly more quickly; however, there was no difference in overall germination success. This increased germination rate could be a function of maternal investment in that population during the collection year or indicate increased selection pressure relative to PSMI. First, there is lower precipitation in CHWA relative to PSMI, which could lead to seeds collected there being predisposed to germinate at the first instance of heavy watering. Secondly, this could be due to vegetative characteristics in the region; CHWA has much less open ground than the primary-successional dune shore of PSMI. *Gypsophila paniculata* has to compete with the many *Artemis* spp. that grow there, and survival is not dependent on merely whether the plant can grow, but whether or not it can compete, and early germination can confer advantages in nutrient and water limited environments like a sagebrush steppe. Additionally, family effects were observed, which could be a function of genetic differences or maternal investment. The collection date could have

biased results of our germination experiment, particularly since the number of growing degree days differs among these regions. While this date was previously determined by Rice (2018) to be a date after which populations of *G. paniculata* growing in Michigan have a greater than 90% germination success rate, no data existed for optimal seed collection time for *G. paniculata* growing in CHWA. It is possible that the collection date could influence results, but we would expect to see it favor PSMI seeds, as Michigan populations of *G. paniculata* were the basis for its selection. Whether the results seen in our germination experiment are a function of maternal investment or genetic architecture is impossible to say without a reciprocal transplant experiment or multigenerational analysis. Regardless of the underlying cause, the data indicate differences in life history traits that may be specific to the divergent environmental pressures present in these two populations.

In our common garden growth trial, CHWA seedlings had significantly higher emergence than seedlings sprouted from PSMI seeds, but there was no difference in the ratio of above:below ground tissue allocation once seedlings had emerged. However, family effects were again observed, and could be a function of either maternal investment or genetic differences. The differences in emergence for these populations could indicate that the CHWA population of *G. paniculata* has a competitive edge to PSMI and could again be due to the harsher conditions present in the sagebrush steppe. Specifically, *G. paniculata g*rowing in CHWA must compete against woody perennials that are already established aboveground at the start of every growing season, while *G. paniculata* growing along the dune shore competes with grass species that sprout new leaves every year.

This study sheds light on the biological processes that may be helping populations of the invasive weed *G. paniculata* adapt to novel environments. Differences in gene expression reveal

responses to the many novel environmental stressors these populations face while growing in harsh, divergent ecosystems. Differences in early life history strategies were observed that suggest that the population growing in CHWA may be facing stronger selective pressures and a harsher environment, leading to rapid germination and increased emergence relative to PSMI. Family effects were observed in both germination trials and common garden growth trials, but whether those effects are a function of maternal investment or are based in genetic differences is currently unclear. This study adds to the growing body of work investigating the success of invasive plant species in novel environments and sheds light on the plastic responses of plants in different ecosystems.

ACKNOWLEDGEMENTS

The authors would like to thank Thermo Fisher Scientific, the Michigan Botanical Foundation, Grand Valley State University, and the Environmental Protection Agency – Great Lakes Restoration Initiative (C.G.P., Grant #00E01934) for financial support of this project. Additional thanks to Marie Adams and Ian Beddows at the Van Andel Research Institute for bioinformatics assistance and to Hailee Leimbach-Maus for help with field collection.

REFERENCES

American Society for Testing and Materials Committee D18 on Soil and Rock. 2004. Standard Test Methods for Particle-Size Distribution (Gradation) of Soils Using Sieve. In: Annual book of ASTM standards. ATSM International.

Babu C. V. S, Gassmann M. 2016. Assessing integrity of plant RNA with the Agilent 2100 Bioanalyzer System. Agilent Application Note. 5990-8850E.

Barkoudah YI. 1962. A revision of *Gypsophila, Bolanthus, Ankyropetalum* and *Phryna*. Wentia. 9: 1–203.

Baskett CA, Emery SM, Rudgers JA. 2011. Pollinator visits to threatened species are restored following invasive plant removal. International Journal of Plant Sciences. 172(3): 411–422.

Baskin JM, Baskin CC. 2001. Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination. 2nd ed. New York: Academic Press.

Blumenthal DM, Hufbauer RA. 2007. Increased plant size in exotic populations: A common-garden test with 14 invasive species. Ecology. 88(11): 2758–2765.

Colautti RI, Maron JL, Barrett SCH. 2009. Common garden comparisons of native and introduced plant populations: Latitudinal clines can obscure evolutionary inferences. Evolutionary Applications. 2(2): 187–199.

Conover DO, Schultz ET. 1995. Phenotypic similarity significance of countergradient variation. Trends in Ecology & Evolution. 10(6): 248–252.

Darwent AL. 1975. The biology of Canadian weeds: 14. *Gypsophila paniculat*a L. The Biology of Canadian Weeds. 55(14): 1049–1058.

Darwent AL, Coupland RT. 1966. Life History of *Gypsophila paniculata*. Weeds. 14(4): 313.

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD. 2015. The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. Molecular Ecology. 24(9): 2095–2111.

Emery SM, Doran PJ. 2013. Presence and management of the invasive plant *Gypsophila paniculata* (baby's breath) on sand dunes alters arthropod abundance and community structure. Biological Conservation. 161: 174–181.

Frankham R. 2005. Resolving the genetic paradox in invasive species. Heredity. 94(4): 385–385.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology. 21(3): 394–407.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology. 29(7): 644–52.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols. 8(8): 1494–512.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecology Letters. 7(12): 1225–1241.

Lande R. 2015. Evolution of phenotypic plasticity in colonizing species. Molecular Ecology. 24(9): 2038–2045.

Langlet O. 1971. Two hundred years genecology. Taxon.(20): 653–721.

Leinonen T, O'Hara RB, Cano JM, Merilä J. 2008. Comparative studies of quantitative trait and neutral marker divergence: A meta-analysis. Journal of Evolutionary Biology. 21(1): 1–17.

McKay JK, Latta RG. 2002. Adaptive population divergence: Markers, QTL and traits. Trends in Ecology and Evolution. 17(6): 285–291.

McNair JN, Sunkara A, Frobish D. 2012. How to analyse seed germination data using statistical time-to-event analysis: non-parametric and semi-parametric methods. Seed Science Research. 22(2): 77–95.

R Development Core Team. 2017. R: A language and environment for statistical computing.

Rice E. 2018. Assessment of invasive *Gypsophila paniculata* control methods in the northwest Michigan dunes. Masters Theses. 888.

Richards CL, Bossdorf O, Muth NZ, Gurevitch J, Pigliucci M. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecology Letters. 9(8): 981–993.

Rio DC, Ares M, Hannon GJ, Nilsen TW. 2010. Purification of RNA using TRIzol (TRI Reagent). Cold Spring Harbor Protocols. 5(6).

SAS Institute Inc. 2013. SAS/ACCESS® 9.4.

Sork VL. 2018. Genomic studies of local adaptation in natural plant populations. Journal of Heredity. 109(1): 3–15.

Stevens OA. 1957. Weights of Seeds and Numbers Per Plant. Weeds. 5(1): 46–55.

Swearingen J, Bargeron C. 2016. Invasive plant atlas of the United States. University of Georgia Center for Invasive Species and Ecosystem Health. invasiveplantatlas.org.

Van Tienderen PH. 1997. Generalists, specialists, and the evolution of phenotypic plasticity in sympatric populations of distinct species. Evolution. 51(5): 1372–1380.

USDA. 2019. Plants Profile for *Gypsophila paniculata* (baby's breath).

Via S, Lande R. 2006. Genotype-Environment interaction and the evolution of phenotypic plasticity. Evolution. 39(3): 505.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 10(1): 57–63.

TABLES

Table 1. Climate data for sampling locations, taken from National Oceanic and Atmospheric Organization (NOAA) weather stations in Petoskey, MI and Entiat, WA (near Chelan, WA).

	Station ID	GPS Coordinates	Elevation (m)	2017 Mean Temp. (˚C)	2018 Mean Temp. (˚C)	2017 Precipitation (cm)	2018 Precipitation (cm)
Entiat Fish Hatchery (WA)	USC00452563	47.6983°, -120.3228°	313	10.33	12.22	37.95	27.81
Petoskey (MI)	USC00206507	45.3725°, -84.9766°	182.6	7.66	7.17	109.75	88.62

Table 2. Sample location information and details of tissue collection for transcriptome analysis for populations of *Gypsophila paniculata* used in this study.

	Sample Code	Collection Date	GPS Coordinates	Leaf Tissue (*n*)	Stem Tissue (*n*)	Root Tissue (*n*)	Total Individuals Used for Sampling
Chelan, WA	CHWA	June 7-8, 2018	47.7421˚N 120.2177˚W	10	9	10	16
Petoskey, MI	PSMI	June 1, 2018	45.4037˚N 84.9121˚W	10	10	10	14

Table 3. Five most significant gene ontology (GO) terms for *Gypsophila paniculata* growing in Chelan, Washington (CHWA) and Petoskey, Michigan (PSMI) by tissue type.

Comparisons	GO Term	GO Name	P-value	FDR
Root Tissue				
Higher in CHWA	GO:0048205	COPI coating of Golgi vesicle	1.42E-03	3.30E-02
	GO:0045041	protein import into mitochondrial intermembrane space	2.40E-03	4.89E-02
	GO:0051560	mitochondrial calcium ion homeostasis	5.68E-04	1.56E-02
	GO:0009643	photosynthetic acclimation	1.74E-03	3.85E-02
	GO:0010337	regulation of salicylic acid metabolic process	1.07E-04	3.61E-03
Higher in PSMI	GO:0015857	uracil transport	9.16E-04	3.15E-02
	GO:0042276	error-prone translesion synthesis	9.16E-04	3.11E-02
	GO:0032107	regulation of response to nutrient levels	9.16E-04	3.10E-02
	GO:0015800	acidic amino acid transport	1.98E-04	8.59E-03
	GO:0055062	phosphate ion homeostasis	1.68E-05	1.13E-03
Stem Tissue				
Higher in CHWA	GO:0016560	protein import into peroxisome matrix, docking	3.74E-04	1.56E-02
	GO:0006516	glycoprotein catabolic process	5.89E-04	2.34E-02
	GO:0043247	telomere maintenance in response to DNA damage	8.69E-04	3.19E-02
	GO:0006517	protein deglycosylation	1.10E-04	6.33E-03
	GO:0042542	response to hydrogen peroxide	4.86E-12	1.60E-09
Higher in PSMI	GO:0015714	phosphoenolpyruvate transport	8.66E-04	3.51E-02
	GO:0015857	uracil transport	1.35E-03	4.66E-02
	GO:0006145	purine nucleobase catabolic process	1.35E-03	4.60E-02
	GO:0031507	heterochromatin assembly	1.35E-03	4.58E-02
	GO:0009558	embryo sac cellularization	1.35E-03	4.55E-02
Leaf Tissue				
Higher in CHWA	GO:0051560	mitochondrial calcium ion homeostasis	1.53E-04	5.21E-03

	GO:0046345	abscisic acid catabolic process	1.39E-03	3.38E-02
	GO:0071486	cellular response to high light intensity	2.05E-03	4.56E-02
	GO:0080151	positive regulation of salicylic acid mediated signaling pathway	2.05E-03	4.51E-02
	GO:0016558	protein import into peroxisome matrix	6.63E-04	1.86E-02
Higher in PSMI	GO:0071492	cellular response to UV-A	1.43E-05	1.88E-03
	GO:0071486	cellular response to high light intensity	4.14E-05	4.70E-03
	GO:0009395	phospholipid catabolic process	4.15E-04	2.53E-02
	GO:0048574	long-day photoperiodism, flowering	6.53E-04	3.65E-02
	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	1.55E-04	1.25E-02

Table S1. Results of soil nutrient analysis for samples collected from locations in Chelan, WA (CHWA) and Petoskey, MI (PSMI).

Sample ID	Depth (cm)	Organic Matter (%)	Phosphorus (ppm-P)	Potassium (ppm)	Magnesium (ppm)	Calcium (ppm)	Soil pH	CEC (meq/100g)	Total Nitrogen (%)
PSMI-1	10	0.1	1	1	45	2300	7.8	11.9	0.007
PSMI-2	50	0.1	1	1	35	1850	8.3	9.5	0.004
PSMI-3	100	0.1	1	1	40	1800	8.1	9.3	0.008
PSMI-4	10	0.2	1	6	45	3450	8.2	17.6	0.011
PSMI-5	50	0.1	1	1	40	2400	8.2	12.3	0.007
PSMI-6	100	0.1	1	1	40	2550	8.3	13.1	0.010
CHWA-1	10	1.4	19	144	105	850	7.2	5.5	0.044
CHWA-2	25	1.0	10	100	135	700	7.5	4.9	0.017
CHWA-3	50	1.1	20	129	165	1100	6.8	7.4	0.029
CHWA-4	10	1.3	27	194	110	800	7.0	5.4	0.046
CHWA-5	25	1.3	23	182	135	1000	7.1	6.6	0.037
CHWA-6	50	1.4	29	201	140	1050	7.4	6.9	0.037
PSMI Avg.	53.3	0.12	1.0	1.8	40.8	2391.7	8.2	12.3	0.008
CHWA Avg.	28.3	1.25	21.3	158.3	131.7	916.7	7.2	6.1	0.035

FIGURE LEGENDS

Figure 1. Principal component analysis (PCA) of soil nutrient data for sampling locations in Chelan, WA (CHWA) and Petoskey, Michigan (PSMI). Ca = calcium, CEC = cation exchange capacity, K = potassium, Mg = magnesium, N = nitrogen, OM = organic matter, and PO_4= phosphorus

Figure 2. Intersection plot visualizing the number of differentially expressed genes in *Gypsophila paniculata* growing in Chelan, Washington (CHWA) and Petoskey, Michigan (PSMI) broken down by tissue type (root, stem, and leaf tissue).

Figure 3. Germination curves for *Gypsophila paniculata* seeds collected Chelan, Washington (CHWA, n = 2,000) and Petoskey, Michigan (PSMI, n = 2,000) on August 11, 2018 and incubated for 12 days. Burpee commercial cultivar seeds (n = 100) known to have germination success in excess of 90% were used for an experimental control

Figure 4. Results of a common garden growth trial of *Gypsophila paniculata* plants conducted for seven weeks (n=120 per population). a) Emergence per sampling location b) Ratio of above:below ground tissue allocation per sampling location.
CHWA = Chelan, Washington; PSMI = Petoskey, Michigan

Figure S1. Bar plot showing the ratio of above:below ground tissue allocation in *Gypsophila paniculata* plants grown in a common garden for seven weeks. Bars represent standard error. Six seeds were planted from each parent plant, with 20 parent plants per population (n = 120 plants per population).
CHWA = Chelan, Washington; PSMI = Petoskey, Michigan

FIGURES

Figure 1

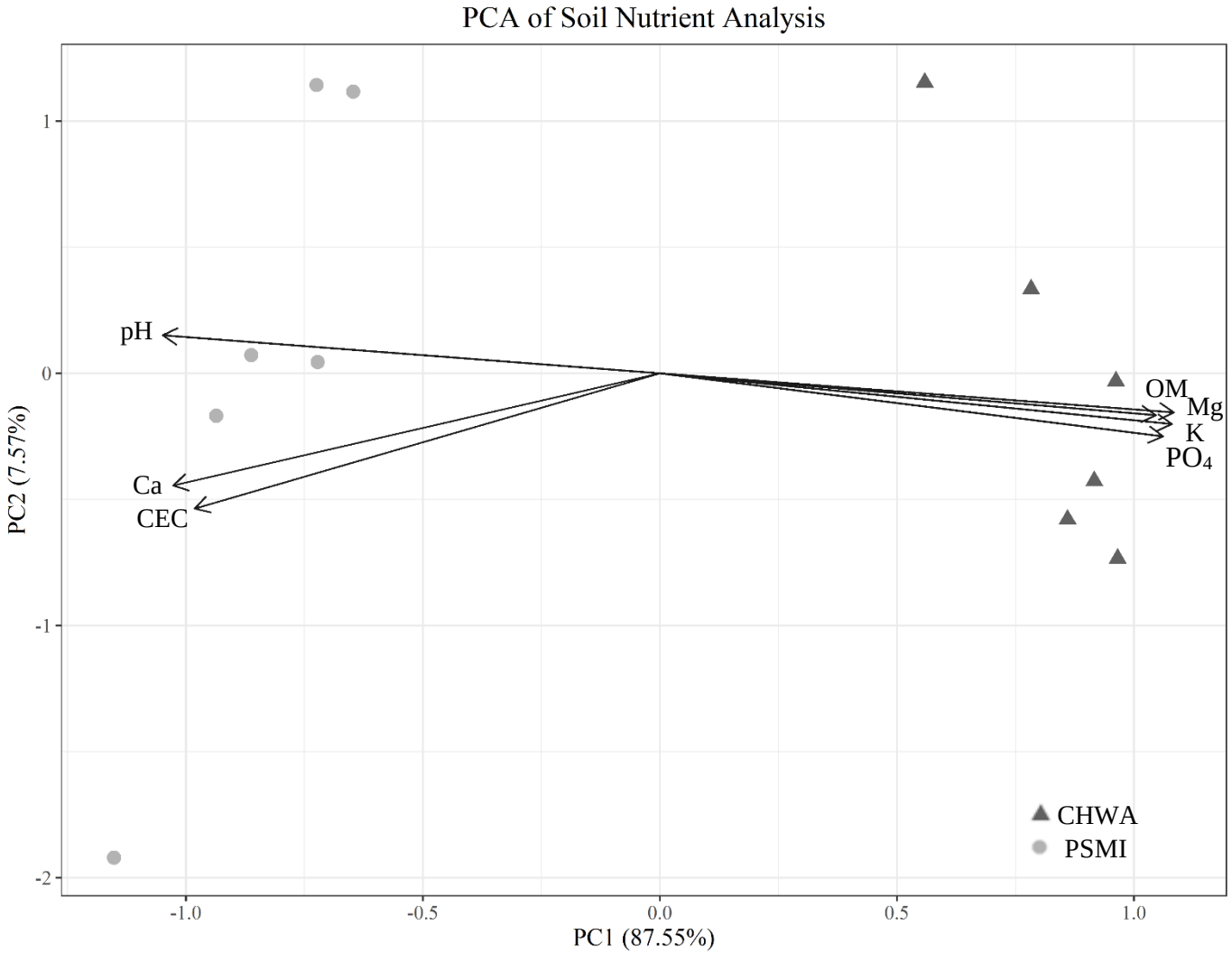


Figure 2

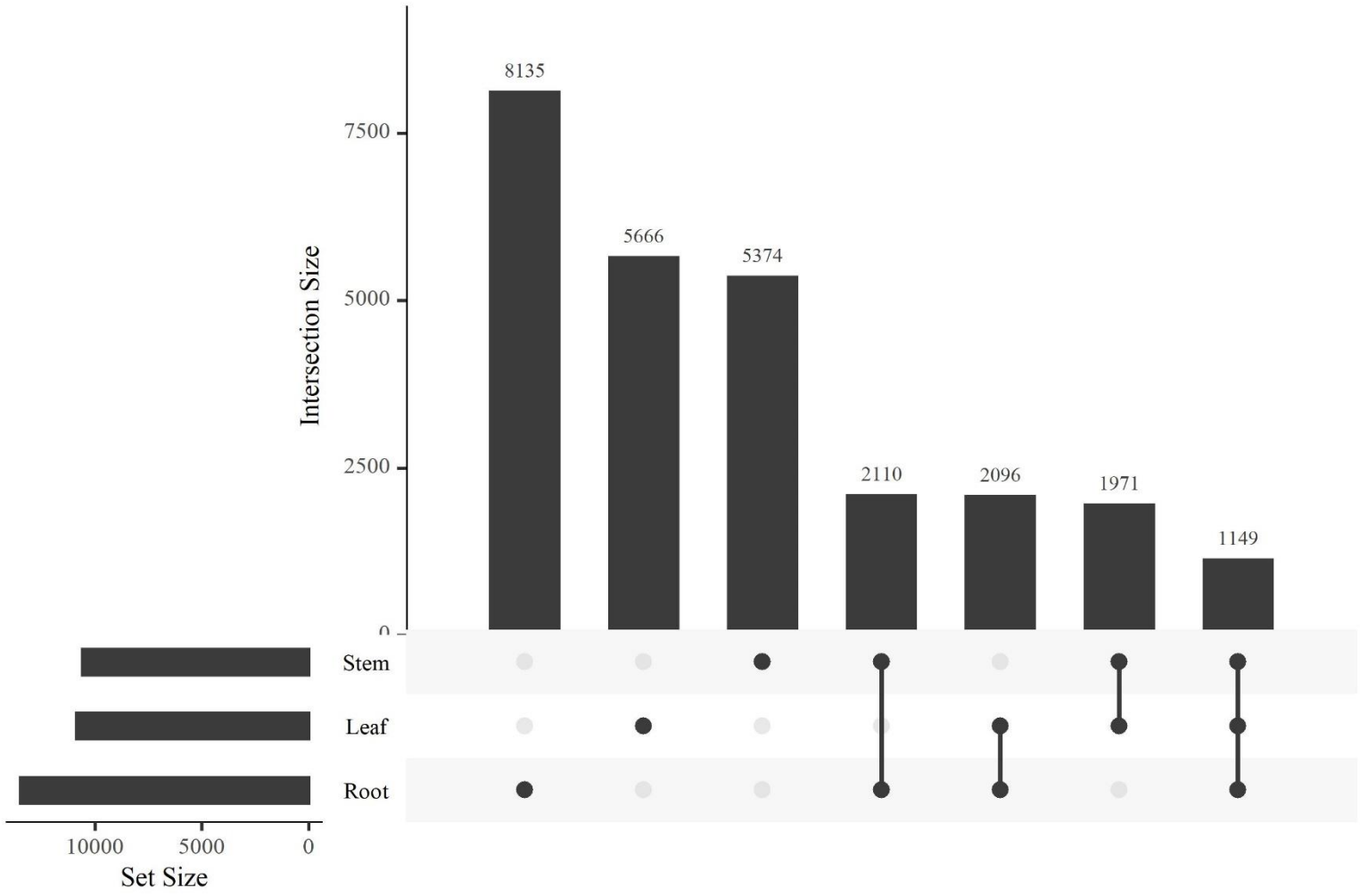

8135
5666
5374
2110
2096
1971
1149
7500
5000
2500
0
Intersection Size
Stem
Leaf
Root
10000
5000
0
Set Size

Figure 3

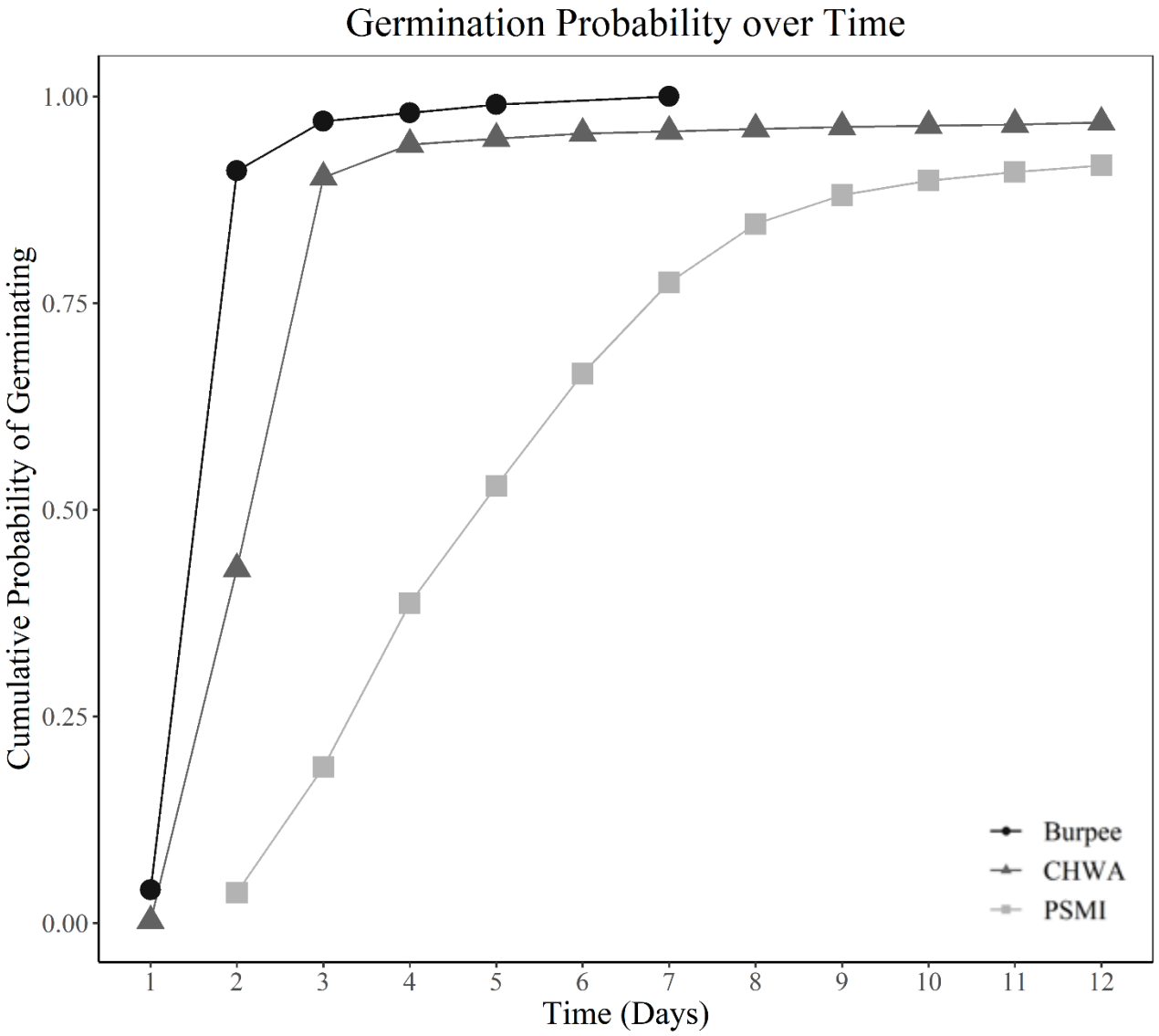

Germination Probability over Time
Cumulative Probability of Germinating
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
Time (Days)
Burpee
CHWA
PSMI

Figure 4

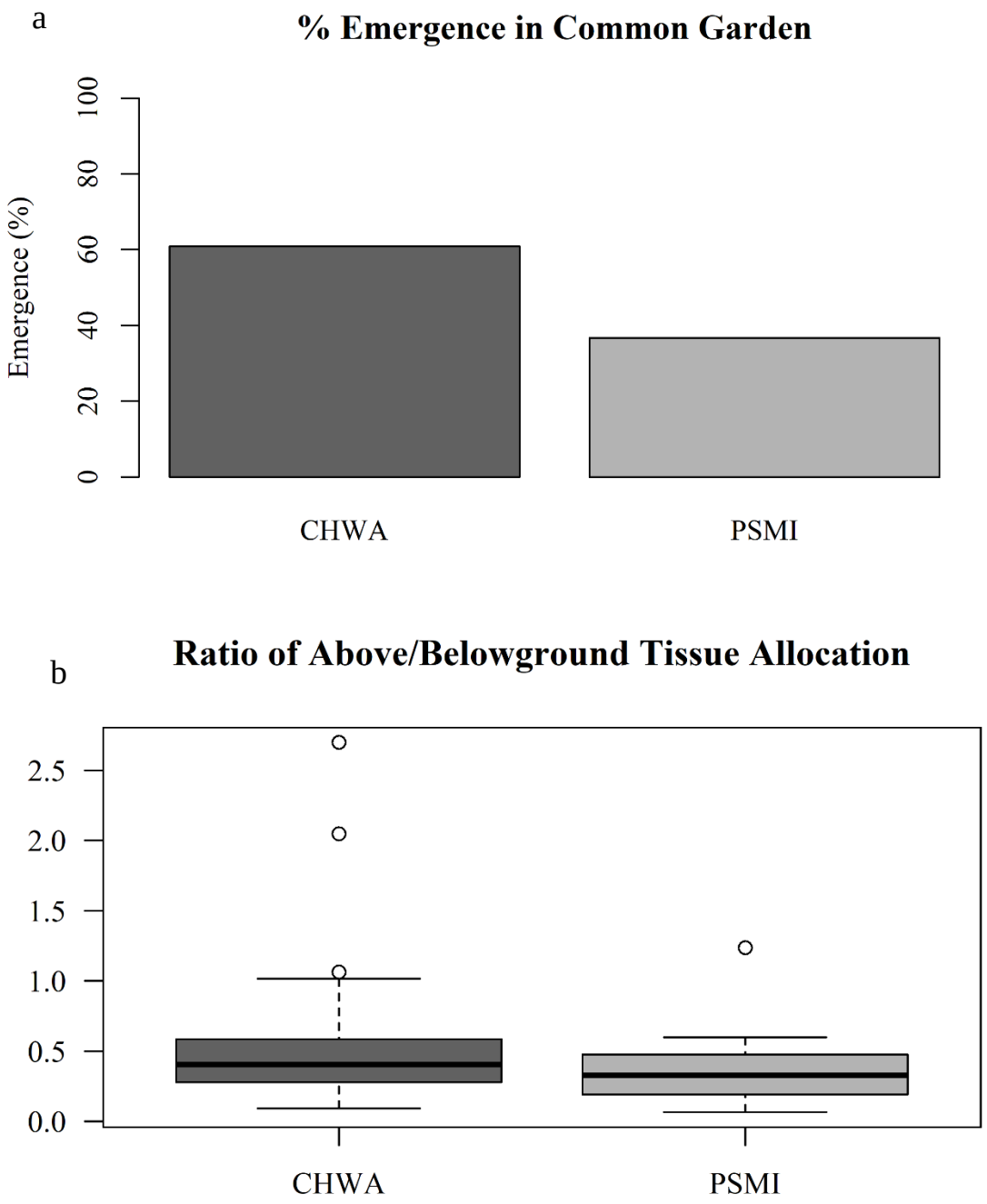
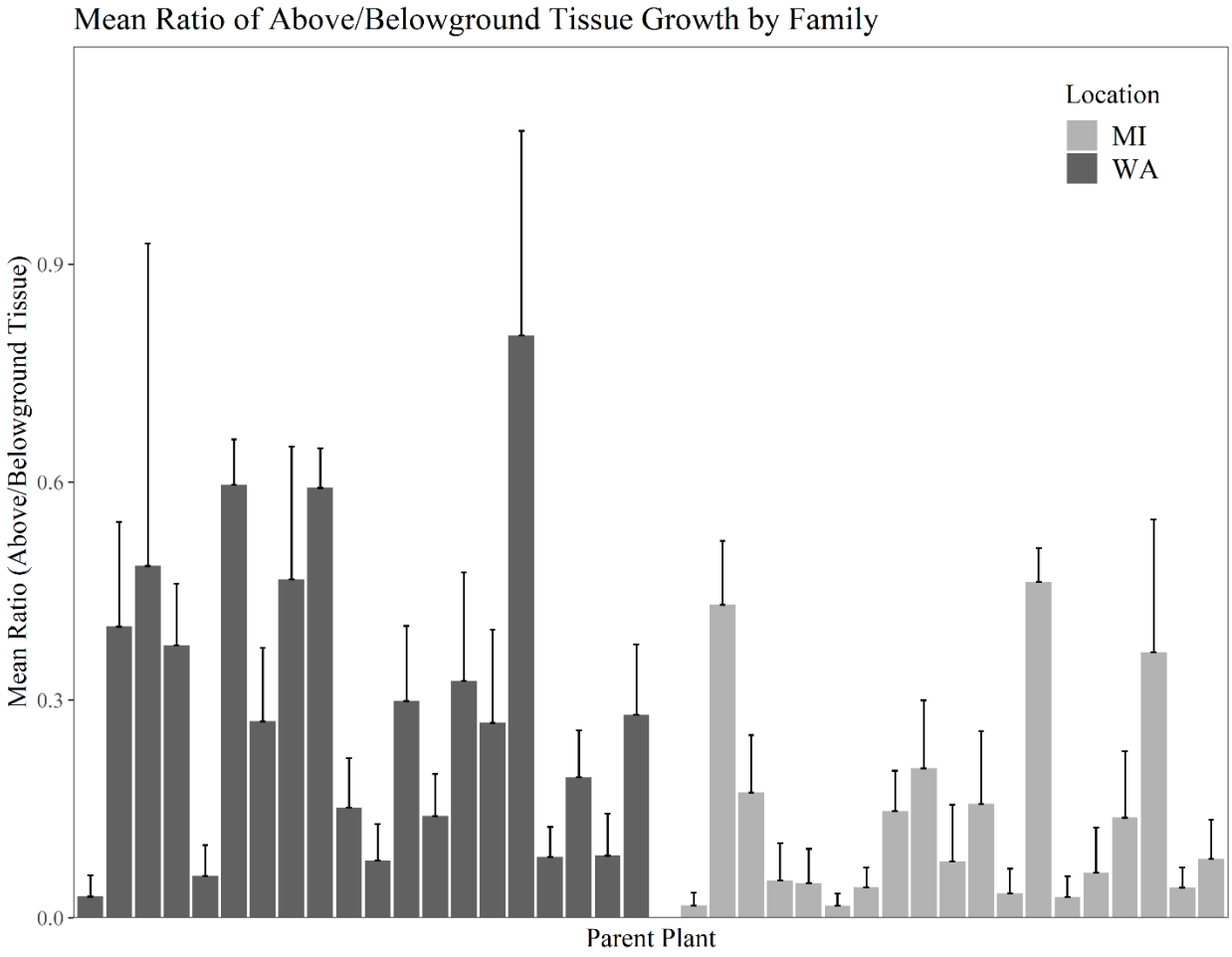

a
% Emergence in Common Garden
Emergence (%)
0
20
40
60
80
100
CHWA
PSMI
b
Ratio of Above/Belowground Tissue Allocation
0.0
0.5
1.0
1.5
2.0
2.5
CHWA
PSMI

Figure S1


Mean Ratio of Above/Belowground Tissue Growth by Family
Location
MI
WA
Mean Ratio (Above/Belowground Tissue)
0.0
0.3
0.6
0.9
Parent Plant

CHAPTER IV

EXTENDED REVIEW OF LITERATURE

Invasive Species

Introduction and Establishment

Invasive species are not an invention of recent history; since the Neolithic era humans have been trading and transporting (both incidentally and intentionally) biota across the globe. However, the number of species being introduced to novel environments has increased dramatically in the 200 years since the Industrial Revolution (Mack et al. 2000; Hulme 2009). More recently, new technical and logistic advancements have further increased the ease with which global trade and transport is conducted, leading to an even greater amplification in the number of invasive species being moved worldwide in the past few decades (Mack et al. 2000; Hulme 2009).

Broadly, invasive species are introduced to new environments through one of three mechanisms: commodity import, transport vector arrival ("hitchhiking"), or range expansion (Hulme et al. 2008). These three broad mechanisms can be further broken down into six major pathways: release, escape, contaminant, stowaway, corridor, or unaided expansion (Hulme et al. 2008). Each pathway tends to be associated with a particular mechanism. For example, release, escape, and contaminant are all associated with the import of commodities (Hulme et al. 2008). This mechanism is responsible for the majority of well-known invaders, with vertebrates typically being classified as deliberate releases, invertebrates as contaminants of other commodities, and plants as escaped commodities (Hulme et al. 2008). The contamination pathway is also significant in the introduction of fungi and microscopic pathogens, though they can be associated with vector travel as well (Hulme et al. 2008).

Upon introduction, an alien species must overcome multiple potential barriers to establishment in order to begin the invasion of a new environment, such as surviving transport and colonization and beginning to reproduce successfully (Larkin 2012). These initial obstacles that occur shortly after introduction often result in a "lag phase", or a period of slow growth and expansion while an introduced species naturalizes to its new environment (Larkin 2012). This period of relative dormancy after introduction may be the result of a species needing time to overcome the Allee effect or for environmental and biotic conditions to change in such a way as to favor the spread of the introduced species (Mack et al. 2000; Crooks 2007). Lag phases can last for long periods of time and make it difficult for researchers to parse out which species will eventually overcome this period of relative dormancy and which species will ultimately fail to establish (Mack et al. 2000; Crooks 2007; Larkin 2012). Once a species has successfully overcome the lag phase of an invasion, there is typically a period of rapid range expansion or population growth, which will ultimately plateau as the invader saturates its new range (Larkin 2012).

Adaptation and Invasion Success

By definition, invasive species are succeeding in an environment in which they did not evolve. This success can often have negative effects on the native species that live there (Pyšek 1995; Richardson and Pyšek 2006). These effects can be both indirect and direct, and include things such as niche displacements, disruption of mutualisms, competitive exclusion, and even native extinctions (Mooney and Cleland 2001). The ability of invasive species to enter a novel ecosystem in relatively low frequencies and outcompete locally-adapted native species is a phenomenon coined the "paradox of invasions" (Sax and Brown 2000). The paradox of invasive success has been studied since the advent of invasion science, focusing mainly on the questions

of what makes a “good” invader, and what makes a “susceptible” ecosystem (Sax and Brown 2000). Many life history characteristics have been tentatively identified as predisposing species to invasive success, but results across studies are varied and contradictions are rife throughout the literature (see Kolar & Lodge, 2001; Mooney & Cleland, 2001; Williamson & Fitter, 1996). The same statement holds true for invaded ecosystems, though there is consensus that island ecosystems may be particularly vulnerable to invasion, perhaps due to their long period of relative isolation and naïve fauna (Reaser et al. 2007; Bellard et al. 2017; Russell et al. 2017).

More recently, the genetics behind the paradox of invasions have begun to receive attention. This research is concerned with the genetic mechanisms that may be allowing invasive species to quickly adapt to the novel environments they find themselves in after introduction (Dlugosch et al. 2015; Sork 2018). When a species undergoes an introduction event, they are often assumed to also undergo a significant bottleneck in both population size and genetic diversity relative to their source population (Sakai et al. 2001; Frankham 2005; Dlugosch et al. 2015). Extensive research on this bottleneck event suggests that while invaders typically do undergo a reduction in population size during an invasion event, this is not always linked with a reduction in genetic diversity relative to their source population (Frankham 2005; Dlugosch et al. 2015). For example, a founding population comprised of a single breeding pair of an outbreeding species will only result in a 25% reduction in heterozygosity, as long as the population grows steadily and does not undergo extended periods of small population size (Frankham 2005; Lande 2015). Therefore, if a founding population is large enough to overcome the many demographic barriers to introduction, genetic diversity will likely be retained at a sufficient enough level to overcome the genetic barriers to establishment (Frankham 2005; Dlugosch et al. 2015). Species that do experience a dramatic reduction in genetic diversity after an introduction event may

experience inbreeding depression and a limited ability to evolve to novel environmental pressures (Sakai et al. 2001). However, a reduction in total genetic diversity is not a guarantee of invasive failure; research shows that more important than *how much* genetic diversity is retained is *what* genetic diversity is retained (Dlugosch et al. 2015). While total genetic diversity is often used as a proxy for adaptive potential, the difference between total and adaptive variation can often be large (McKay and Latta 2002; Leinonen et al. 2008).

Some degree of pre-adaption to their introduced environments is necessary for the success of invasive populations, but there will likely be aspects of the new ecosystem that are still novel (Mack et al. 2000; Sakai et al. 2001). The study of adaptive potential in invasive species focuses on the many sources of potential adaptive variation that are able to withstand population bottlenecks, such as: loci of large effect, genetic variation that is cryptic in the native range but becomes apparent in the introduced range, and the importance of serial invasion events, particularly in relation to the admixture of previously isolated alleles (see Dlugosch et al., 2015). More recently, gene expression has become apparent as an important factor in local adaptation of invasive species that are succeeding in novel environments (Sork 2018).

With the advent of new technologies like RNA-seq that allow for the assembly of transcriptomes *de novo*, gene expression data has become more attainable for use in invasive studies (Wang et al. 2009; Sork 2018). Gene expression profiles obtained through RNA-seq methods can be used to investigate genotype-by-environment (GxE) interactions which function as a measure of phenotypic plasticity when considering populations found in discrete environments (Via and Lande 2006; Sork 2018). These methods of investigating gene expression allow researchers to find evidence for the prevalence of phenotypic plasticity in response to novel environmental stressors (Lande 2015; Sork 2018). One study looking at multiple

populations of valley oak (*Quercus lobata*) taken from climactically distinct regions exposed acorns to differing levels of water stress (Gugger et al. 2016). The authors found that 52% of approximately 68,000 contigs were differentially expressed before and after the water stress treatment, and that 56 contigs showed a population by treatment interaction (Gugger et al. 2016). However, these results are study specific, and there is not a clear trend that suggests an increase in plasticity in invasive species as a whole (Lande 2015). When species are expressing a plastic response to the novel environmental stressors they encounter, the length and timing of that response will be determined by factors such as the optimal and mean phenotype, variation and predictability of the environmental stressors found in the introduced range relative to the species' native range, the cost of plasticity, and whether the population is experiencing one-shot or labile plasticity (Lande 2015).

The ability of invasive species to adapt to their introduced environments is essential for the persistence of introduced populations (Sakai et al. 2001; Sork 2018). Whether these adaptations occur through genetic changes associated with a population's adaptive potential or initially through plastic means which may become adaptive, like differential gene expression, these molecular mechanisms are often what makes the difference between a successful invader and one that will not persist past the lag phase of invasion.

Importance

The far-reaching impacts of invasive species are felt across almost every ecosystem. In general, the proportion of non-native species occupying environments increases as you move towards the equator. This trend continues until you reach the life-rich tropics, which tend to withstand invasions relatively well (Vitousek et al. 1997). Islands with extensive travel and trade feel the burden of introduction and subsequent invasions more heavily than both continental habitats and

more isolated islands, and the sum of these worldwide introductions are now considered a prominent factor in global change (Vitousek et al. 1997).

There have been an estimated 50,000 non-native species introduced to the U.S. since its founding (Pimentel et al. 2005). Many of these introductions have been intentional for use in various industries (e.g. agriculture, textile, horticulture) (Vitousek et al. 1997; Pimentel et al. 2001; Pimentel et al. 2005). For example, non-native species now comprise 98% of the U.S. food system and provide at least $800 billion in profit every year (Pimentel et al. 2005). However, many of these introduction events have resulted in invasive species that are detrimental to their new environments. Invasive species are estimated to cost the U.S. $120 billion to mitigate on an annual basis (Pimentel et al. 2005). Just under half of all plant and animal species listed as either threatened or endangered on the Endangered Species Act cite competition with non-native species as a primary threat (Wilcove et al. 1998). Additionally, half of all extinctions for which the cause is identified can be attributed to either over-exploitation or invasive species (Bellard et al. 2016). Most of these extinctions are due to the introduction of vertebrate predators to islands (Bellard et al. 2017). Because of the impacts they can have on both ecosystems and economies, invasive species are an important and growing area of ecological research.

Baby's Breath (*Gypsophila paniculata*)

Physical Characteristics and Life History

Gypsophila paniculata (common baby's breath) is a perennial forb belonging to the Caryophyleaceae family. Previously, *G. paniculata* has been recorded at heights up to 1 m tall (Barkoudah 1962), though during the course of this study plants were regularly found reaching over 1 m in height. *Gypsophila paniculata* is a heavily-branched shrub with opposite, lance-shaped leaves that are covered in glandular hairs (Barkoudah 1962; Darwent and Coupland 1966;

Darwent 1975). Stems change color throughout the plant's life cycle, beginning a dark green shade and shifting to more purple-gray as the plant nears senescence, although different color morphs have recently been identified (Darwent and Coupland 1966; Yang et al. 2019) . *Gypsophila paniculata* has an extensive root system that has been recorded reaching depths of over 3 m, characterized by a thick primary taproot up to 7 cm in diameter (Barkoudah 1962; Darwent and Coupland 1966). This root system stores an abundance of food reserves that allows the plant to persist through the winter (Darwent 1975). Flowers are not produced until the second or third year of growing, and may not be produced every year in mature plants (Darwent and Coupland 1966). Flowers are arranged in a panicle-like manner (Darwent 1975), the characteristic from which the species gets its name. Flowers are small and range from white to shades of pale pastels and have a strong odor (Barkoudah 1962; Darwent and Coupland 1966; Darwent 1975). Pollination by insects has been observed, but self-pollination is also suspected to be a viable option for this species (Darwent and Coupland 1966; Baskett et al. 2011). Seeds are dark, globular spirals with limited to no dormancy period that are primarily wind-dispersed (Darwent and Coupland 1966). A single *G. paniculata* plant can produce almost 14,000 seeds per year (Stevens 1957); the mean reported weight of 100 seeds ranges from 67 mg (Darwent and Coupland 1966) to 86 mg (Stevens 1957). During the latter part of the growing season, *G. paniculata*'s stems become brown and brittle and are easily broken off by strong winds (Darwent and Coupland 1966). This can lead to entire mature plants breaking off above the caudex and tumbling across the landscape, spreading seeds up to 0.8 km (Darwent and Coupland 1966).

Gypsophila paniculata is a hardy forb able to withstand a wide range of temperature, moisture, and light regimes (Darwent 1975). The plant species was first thoroughly described as an invasive species in Canada, where the mean number of annual degree days above 5.5˚C

ranged from 832 to 2,220 in areas invaded. The mean annual precipitation in these areas ranged from 25 to 112 cm, with the plant being most aggressive in areas of lower rainfall (Darwent 1975). *Gypsophila paniculata* has been shown to be drought-tolerant and able to withstand harsh wind conditions across its range, though severe drought can lead to periods of seedling mortality (Darwent 1975). Additionally, *G. paniculata* increases above-ground biomass in the presence of snow melts (Blumenthal et al. 2008). In fact, it is one of the few perennial forbs suggested for growth in areas of permafrost (Harris 1970). *Gypsophila paniculata* succeeds on a variety of substrata, the most common being sandy soils (Barkoudah 1962; Darwent and Coupland 1966; Darwent et al. 1967). However, soils that are too fine may prevent the growth of *G. paniculata*'s taproot (Darwent 1975).

Geographic Distribution

The *Gypsophila* genus originates from the region surrounding the Black Sea, the Caucasus Mountains, and northern Iraq and Iran (Barkoudah 1962). *Gypsophila paniculata* is distributed throughout Central and Eastern Europe (spanning from Austria to European Russia), Asiatic Russia, Mongolia, and western China (Barkoudah 1962). The plant is now found growing as an adventive garden escape in Western Europe and North America (Barkoudah 1962; Darwent 1975).

In North America, *G. paniculata* persists in at least 7 Canadian provinces and 30 U.S. states (EDDMapS 2019; USDA 2019). In the U.S. states of Washington and California, *G. paniculata* is listed as a widespread and noxious weed (USDA 2019). It is considered a priority invasive by The Nature Conservancy in Michigan (Emery and Doran 2013; Swearingen and Bargeron 2016). In some areas of Sleeping Bear Dunes National Lakeshore, a national park located in Michigan, *G. paniculata* occupies as much as 75% of the groundcover present

(Karamanski 2000; Rice 2018). *Gypsophila paniculata* is also listed in *Weeds of the West*, a book listing common weeds found throughout the Western U.S. and used by the U.S. Department of Agriculture to identify species of concern (Whitson et al. 1991; USDA 2019).

Use and Economic Importance

Gypsophila paniculata is primarily cultivated for its popularity as a garden ornamental and by the floral industry as a backdrop for showier blooms in bouquets (Darwent and Coupland 1966; Darwent 1975). For example, the eastern side of the Cascade Mountain Range that cuts through the Pacific Northwest is responsible for producing all *G. paniculata* in the region; in the 1990's this crop harvest led to over $50 million dollars in revenue for the area (Schlosser et al. 1991; Schlosser and Blatner 1997). Livestock have been observed to graze on the plant during its younger life stages, when crude protein content is highest (16.1%) and crude fiber content is lowest (10.9%). Grazing slowed as the plants matured and stems became brittle; as plants matured, crude protein content also dropped to 5.2% while crude fiber rose to 40.5% (Darwent et al. 1967). While this increase in fiber content could prove beneficial to ruminant species, plant palatability appears to decrease with maturity (Darwent et al. 1967).

Additionally, the genus has been cultivated for use of the high saponin contents of its roots in Europe and Asia (Barkoudah 1962). At least nine unique bioactive saponins have been isolated from the roots of *G. paniculata* for use in pharmacy since 2010 (Yao et al. 2010; Shun et al. 2011). The saponins isolated from *G. paniculata* are sometimes consumed for use as a medical purgative, while compounds from the roots of other members of the genus (e.g. *G. arrostii*) are used to treat skin maladies and as a diuretic (Usher 1974). Additionally, saponins from the roots of *G. paniculata* have been found to effectively control two species of nematodes, *Xiphinema index* and *X. diversicaudatum*. These nematode species transmit the two viruses

responsible for most grapevine damage and grape crop loss worldwide, grapevine fanleaf virus and Arabais mosaic virus. Saponins isolated from the roots of *G. paniculata* proved to be an effective control against these nematode species, while not harming the delicate communities of mycorrhizae living in the soil (Pensec et al. 2013).

Ecological Impacts and Control

In the areas *G. paniculata* invades, it can form dense monocultures that drastically alter the existing plant community (Baskett et al. 2011; Rice 2018). Most research done on the ecological impacts of *G. paniculata* has occurred in the primary successional sand dunes surrounding Lake Michigan in the Great Lakes Basin of the U.S. In this system *G. paniculata* has been found to alter pollinator abundance and visitation to native species (Baskett et al. 2011), plant cover (Reid and Emery 2018), nematode richness and abundance (Reid and Emery 2018), and arthropod community structure (Emery and Doran 2013). While the presence of *G. paniculata* increased pollinator abundance at a landscape level, it also decreased the number of pollinator visits to native and threatened species when present in the same plot (Baskett et al. 2011). This increase in pollinator abundance may be due to the increased plant cover that *G. paniculata* provides the sparse community present in a primary successional dune system, as well as its abundant floral displays (Baskett et al. 2011). The presence of invasive *G. paniculata* did not alter plant alpha diversity, suggesting that it is not replacing native species in these environments, but rather colonizing the large swaths of bare ground present in the dune communities (Reid and Emery 2018). However, the presence of *G. paniculata* did alter plant community composition at a plot-scale, resulting in reduced heterogeneity as monocultures of *G. paniculata* took over study plots and reduced abundance of the threatened and endemic plant species *Cirsium pitcheri* (Emery et al. 2013; Reid and Emery 2018). As plant cover increased, so did nematode abundance, likely

due to this increase in plant biomass (Reid and Emery 2018). The impacts of *G. paniculata* invasion on arthropods in this system were dependent on functional feeding group (Emery and Doran 2013). While these impacts may be specific to the dune system these studies were conducted in, the potential of *G. paniculata* to alter the intricate dynamics of the ecosystems it invades is clear.

While much research has been done on the best methods for cultivating *G. paniculata* for horticultural uses (e.g. Fudano, 2007; Shibuya, Murakawa, Nishidate, Nishiyama, & Kanayama, 2017), less research has been conducted on the best methods of control for this problematic invasive. Removal typically consists of manual cutting with a spade below the caudex of the plant (Emery et al. 2013; Rice 2018). However, treatment with herbicide is also effective and yields a minimal amount of resprouting (Rice 2018). Both methods are most effective when treatment is conducted for multiple years and in late June through early July (Rice 2018). When treated with the broad-spectrum herbicide glyphosate, the germinability of seeds was reduced. This reduction was more marked when glyphosate was applied early in the growing season (Rice 2018).

EXTENDED METHODOLOGY

Chapter II

Study Sites and Contemporary Sample Collection

To investigate contemporary population structure of *G. paniculata*, tissue samples from five locations across the United States were collected in the summer of 2018: Petoskey, MI; Knife River Indian Villages National Historic Site, ND; Ottertail, MN; Chelan, WA; and Osborne Bay, WA (Figure 1, Table 1). Samples from two additional locations in Sleeping Bear Dunes National Lakeshore, MI and Arcadia Dunes, MI were collected in the summer of 2016 (Table 1) (Leimbach-Maus, Parks, & Partridge, 2018a). Leaf tissue was collected from 15-30 individuals per location (5-10 leaves per plant). Tissue samples were placed inside coin envelopes and stored in silica until DNA extraction. Individuals were collected for sampling by identifying a plant of any size separated from other sampled individuals by at least 2 meters, in efforts to minimize the likelihood of sampling closely related plants.

Microsatellite Analysis of Contemporary Samples

For each contemporary sample (n=145), DNA was extracted from 0.25 g of dried leaf tissue using a Qiagen DNeasy plant mini kit (QIAGEN, Hilde, Germany), modified to include an extra wash with AW2 buffer. Extracted DNA was cleaned twice using a Zymo OneStep PCR Inhibitor Removal Column (Zymo, Irvine, CA). Samples were amplified at 14 nuclear microsatellite loci identified as polymorphic and specific to *G. paniculata* (Leimbach-Maus, Parks, & Partridge, 2018b). PCR was conducted using a 5' fluorescently-labelled primer (6-FAM, PET, NED, or VIC) (Applied Biosystems, Foster City, CA) and an unlabeled reverse primer. Reaction mixtures consisted of 1x KCl buffer, 2.0-2.5 mM $MgCl_2$, 300 µM dNTP, 0.08 mg/mL BSA, 0.4 µM forward primer, 0.4 µM reverse primer, 0.25 units Taq polymerase, and 50 ng DNA template.

The thermal cycling profile consisted of 5 minutes of denaturation at 94˚C, followed by 35 cycles of 94˚C for 1 minute, 1 minute of annealing at 62˚ (with the exception of locus BB_2888, see Leimbach-Maus et al. 2018b), 1 minute of extension at 72˚C, and a final elongation step of 10 minutes at 72˚C. PCR products were visualized on a 2% agarose gel using GelRed™ (Biotium, Freemont, CA) before multiplexing with consideration to dye color and allele size. Genescan 500 LIZ size standard (Thermo Fisher Scientific, Waltham, MA) was added to multiplexed product with Hi-Di™ Formamide (Thermo Fisher Scientific, Waltham, MA) to aid in denaturing. Fragment analysis was conducted on an ABI3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). Individuals were genotyped using the automatic binning procedure on GENEMAPPER v5 (Applied Biosystems, Foster City, CA) before being visually verified to reduce error. A subsample of 20 individuals were genotyped twice to ensure consistent allele scoring.

The presence of null alleles was investigated using MICRO-CHECKER v2.2.3; using this method, none were found (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004). Data were screened using the 'STRATAG' package in the R statistical program v3.4.3 (Archer, Adams, & Schneiders, 2016; R Development Core Team, 2017) for any individual that was missing greater than 20% of loci and any locus that was missing greater than 10% of individuals; on this basis, no data were removed.

Measures of Genetic Diversity and Structure in Contemporary Populations

Linkage disequilibrium and a test for Hardy-Weinberg equilibrium were calculated using GENEPOP v4.6 with 1,000 batches of 1,000 Markov chain Monte Carlo iterations (Raymond & Rousset, 1995; Rousset, 2008). There was no significant deviation from linkage equilibrium across populations and no data were removed on this basis. Expected versus observed

heterozygosity, number of private alleles, and Weir and Cockerham's population pairwise F_{ST} values were conducted using GENALEX v6.502 in Microsoft Excel (Peakall & Smouse, 2006, 2012; Weir & Cockerham, 1983). Inbreeding coefficient (F_{IS}) values were calculated in GENEPOP.

A principal coordinate analysis (PCoA) was conducted using a genetic distance matrix in GENALEX (Peakall & Smouse, 2006, 2012). Population clustering was analyzed in STRUCTURE v2.3.2 (Pritchard, Stephens, & Donnelly, 2000) using an admixture model, both with and without *a priori* location information, and a burn-in length of 100,000 with 1,000,000 MCMC replicates after burn-in. Ten iterations were run for each *K* value (1-9). The number of genetic clusters was determined using the Evanno ΔK method (Evanno, Regnaut, & Goudet, 2005). Because ΔK is based on a rate of change, it does not evaluate *K*=1 and can be biased towards *K*=2 (Dupuis et al., 2017). Considering this, we also used discriminant analysis of principal components (DAPC) to support our STRUCTURE findings (Jombart, Devillard, & Balloux, 2010). DAPC separates variance into within-group and between-group categories and works to maximize cluster discrimination; this analysis was conducted using the package 'adegenet' v2.1.1 in R (Jombart et al., 2010). Because retaining too many principal components (PC's) can lead to instability in cluster membership properties, a cross-validation was performed to inform the analysis of the optimal number of PC's. After cross-validation, 16 of 28 PC's and all eigenvalues were retained. An analysis of molecular variance (AMOVA) was run using 9,999 permutations in GENALEX to test how much variance could be explained by between-population and within-population variation.

Invasion Curves

To create invasion curves for *G. paniculata* population clusters, public herbarium databases were searched for specimen records of this species; species identification was visually confirmed when possible. Records that did not include location data (either GPS, county (U.S.) or regional municipality (Canada)) and year were discarded, resulting in 307 records from 65 North American institutions (Table S1). All locality information was standardized to the county scale to reduce the risk of redundant specimen collection while maintaining adequate resolution (Antunes & Schamp, 2017). Earliest samples were found in the late 1890's-early 1900s in California, Michigan, Minnesota, and New York and this is consistent with the earliest times in which *G. paniculata* seeds were first being sold in the United States (1886), based on a search of the Henry G. Gilbert Nursery and Seed Trade Catalog Collection from the Biodiversity Heritage Library (https://www.biodiversitylibrary.org/).

To examine the invasion status of populations belonging to genetic clusters identified from our population genetics analysis, herbarium records were grouped according to desired spatial scales (cumulative North America, current location of genetic cluster 1, and current location of genetic cluster 2 in contemporary samples). Only specimen records for the first collection of *G. paniculata* in each county or regional municipality were kept. Cumulative records for all of North America had 184 unique municipalities represented, while records from the geographic area of both genetic clusters had fewer unique localities (cluster 1 = 42, cluster 2 =16) and required log transformation for better visualization. Data were plotted as the cumulative number of localities invaded over time using the statistical program R v6.0.

EXTENDED METHODOLOGY

Chapter III

Soil Analysis

In the spring of 2018, we collected soil and tissue samples from two locations of *G. paniculata* infestation (Table 2). Sampling locations differed in collection depths due to soil characteristics in CHWA that made deeper collection impossible (large boulders, hard soil). At both locations, we collected two sets of soil samples from all depths. In PSMI, we collected soil from 10cm, 50cm, and 1m, while in CHWA, we collected soil from 10cm, 25cm, and 50cm depths. We stored samples in airtight plastic bags and maintained them at 4˚C until analysis.

Using these soil samples, we conducted particle size analysis (PSA) via the sieve method, in which we dried, weighed, and washed soil through multiple size filters (1700, 1000, 500, 250, 125, and 63µm) (American Society for Testing and Materials Committee D18 on Soil and Rock 2004). We then oven-dried the sorted samples and weighed the final amount of soil caught in each filter size. For each location, we used soil samples taken from 10cm and 25cm depths for PSA.

We sent soil samples from all depths at both locations out for nutrient analysis to A&L Great Lakes Laboratories (Fort Wayne, IN), where samples were tested for: organic matter (% OM), phosphorus (P), potassium (K), magnesium (Mg), calcium (Ca), soil pH, total nitrogen (N), cation exchange capacity, and percent cation saturation of K, Mg, and Ca. Upon receipt, samples were dried overnight at 40˚C, crushed with a grinder, and sieved through a 2mm sieve.

To analyze OM, the soil was scooped into dried, weighed crucibles. The crucibles were then placed into an oven at 105˚C to dry any remaining moisture in the samples. After re-

weighing soil samples, they were transferred into a Blue M Oven (Thermal Product Solutions, New Columbia, PA). Soils were heated to 360˚C for two hours, then decreased to 105˚C. Finally, soil samples were removed and reweighed. OM content is determined by calculating the sample weight lost at 360˚C divided by the original dried soil weight. This is converted to a percentage and multiplied by a conversion factor (0.98) to regress the loss on ignition.

To determine pH of the soil, soil is mixed into a 1:1 soil:water slurry. The slurry is left to acclimate for roughly 20 minutes before being measured with a calibrated electrode. Available phosphorus and exchangeable calcium, magnesium, and potassium were all measured using soil that had been extracted with a Mehlich III extraction (Mehlich 1984), where soil is mixed with Mehlich III extracting solution (0.2 N acetic acid, 0.25 N ammonium nitrate, 0.015 N ammonium fluoride, 0.013 N nitric acid, and 0.001 M EDTA at pH 0.25 ± 0.05) and shaken on an oscillating extractor at 180 opm for 5 minutes. The soil is then poured through a paper filter to remove soil particles from the liquid solution. This solution was then analyzed via inductively coupled plasma spectroscopy for available phosphorus and exchangeable calcium, magnesium, and potassium content.

Total nitrogen was analyzed via thermal conductance, or the Dumas method. First, soil was weighed and loaded into a TruMac Nitrogen Analyzer (LECO, St. Joseph, MI). Analysis consists of a burn phase and an analyze phase. During the burn phase, the sample was placed in a combustion chamber heated to 1350˚C and filled with O_2. Combustion produced gasses (CO_2, H_2O, O_2, NO_x, and N_2) that were cooled to remove water and collected in the ballast volume. All gasses collected mixed freely and became homogenous. In the analysis phase, a sample of this gas mixture was collected. This sample was then mixed with pure argon and passed through a hot copper tube to remove O_2 and convert NO_X to N_2. The gas mixture was then run through

sodium hydroxide to remove CO_2 and magnesium perchlorate to remove any remaining water. Finally, the only remaining gas product of combustion (N_2) was measured via gas thermal conductivity. Results of nutrient analysis were analyzed using a principal component analysis in the R statistical program v6.0 (R Development Core Team 2017).

RNA Extraction

We collected seedlings from CHWA and PSMI concurrently with soil samples. First, we located *G. paniculata* seedlings separated by at least 2 meters to reduce the risk of redundant sampling whenever possible. We then dissected seedlings into three tissue types (root, stem, and leaf), placed tissue in RNAlater™ (Thermo Fisher Scientific, Waltham, MA), and flash froze them in an ethanol and dry ice bath. We kept sample on dry ice during transport and maintained them at -80˚C until extraction in the lab.

We extracted total RNA from frozen tissue using a standard TRIzol® (Thermo Fisher Scientific) extraction protocol (Rio et al. 2010). Briefly, we ground frozen leaf tissue in liquid nitrogen with a chilled pestle and mortar. We then dissolved the powdered sample in TRIzol® and added chloroform before mixing thoroughly. The mixture was allowed to acclimate at room temperature before being centrifuged at 10,000g for 10 minutes. After centrifugation, we transferred the clear upper layer into a new tube; for every mL of clear phase, we added 0.5mL of isopropanol. We then mixed samples vigorously and incubated them at -20ºC for 10 min., before centrifuging them at 10,000g for 15 min. at 4°C to pellet precipitated RNA. We carefully decanted the supernatant away and washed the pellet twice with 75% ethanol. Finally, we resuspended the extracted RNA pellet in DNase/RNase free water, before treating it with a DNA-Free Kit (Invitrogen, Carlsbad, CA). We assessed RNA quality with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and NanoDrop™ 2000 (Thermo Fisher Scientific). RIN

(RNA Integrity Number) values for individuals used in this study ranged from 6.1-8.3. However, because both chloroplast and mitochondrial rRNA's can artificially deflate RIN values in plant leaf tissue, we deemed these values to be sufficient for further analysis (see Babu C. V. and Gassmann, 2016). Finally, we submitted total RNA to the Van Andel Research Institute for cDNA library construction and sequencing.

cDNA Library Construction and Sequencing

Prior to sequencing, all samples were treated with a Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA). cDNA libraries were constructed using the Collibri Stranded Library Prep Kit (Thermo Fisher Scientific) before being sequenced on a NovaSeq 6000 (Illumina) using S1 and S2 flow cells. Sequencing was performed on a paired end 2 x 100 bp format and produced approximately 60 million reads per sample, with 94% of reads having a Q-score >30.

Transcriptome Assembly

Prior to transcriptome assembly, read quality was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed and filtered using SortMeRNA (https://bioinfo.lifl.fr/RNA/sortmerna/), keeping only non-rRNA reads for downstream processing. A reference transcriptome was assembled *de novo* using Trinity v2.8.2 (Grabherr et al. 2011; Haas et al. 2013) with a normalized max read coverage of 100, a minimum k-mer coverage of 10, and k-mer size set to 32. The assembled transcriptome was annotated using Trinotate v3.1.1. and consisted of 223,810 genes and 474,313 transcripts from 59 samples. Data were filtered to exclude transcripts that were expressed less than 10 times or in fewer than 10 samples. Following filtering, 111,042 genes (49.61%) and 188,108 transcripts (39.66%) remained. Considering tissue type, 127,591 transcripts remained in the data from 20 root samples

(26.90%), 125,261 transcripts remained in the 19 samples from stem tissue (26.41%), and 112,499 transcripts remained in the 20 leaf tissue samples (23.72%).

Differential Expression

Differential expression was analyzed using the *edgeR* framework in R; to be considered significant, genes needed to have a p-value below 0.05 after false discovery rate correction and a log2 fold change greater than 2. For transcripts that were differentially expressed at the level of both tissue type and population, Gene Ontology (GO) analysis was conducted using the PANTHER classification system v14.1, where transcripts were assessed against the *Arabidopsis thaliana* genome (http://pantherdb.org/webservices/go/overrep.jsp).

Germination Trial

On August 11, 2018 we returned to our sample sites in CHWA and PSMI and collected seeds from 20 plants per location; Rice (2018) previously determined this collection date to yield over 90% seed germination for *G. paniculata*. To collect seeds, we manually broke seed pods off and placed them inside paper envelopes in bags half-filled with silica beads. We stored bags in the dark at 20 to 23˚C until the germination trial began.

We counted one hundred seeds from twenty plants per population and placed them in a petri dish lined with filter paper. We established a control dish using 100 seeds from the 'Early Snowball' commercial cultivar *(G. paniculata*) sold by W. Atlee Burpee & Co in 2018, known to have germination percentages in excess of 90%. Incubators had a 12:12h dark:light photoperiod and growth chamber conditions were set at 20˚C with 114 μmol m^{-2} s^{-1} photosynthetically active radiation from fluorescent light bulbs. Each day we randomized petri dish locations within the incubator to avoid bias in temperature or light regimes. We conducted this study for fourteen

days, at which point there had been no germination in any dish for two days. The same individual checked all seeds (n=4,100) daily within the same three-hour time window to minimize bias for germination, functionally defined as radicle emergence (Baskin and Baskin 2001). Once a seed had germinated, we removed it from the dish (method adapted from Rice, 2018).

Using the statistical program R v6.0, we fitted data to a nonparametric Kaplan-Meier time-to-event curve (McNair et al. 2012; R Development Core Team 2017). We then compared germination patterns between collection localities using a pairwise log-rank test (McNair et al. 2012). To test homogeneity within localities, we again conducted a log-rank test. Finally, to investigate the presence of family effects, we ran a series of pairwise log-rank tests with a Holm correction for multiple comparisons (McNair et al. 2012). For all analyses in this study, we set the alpha level to 0.05.

Common Garden Growth Trial

Using seeds collected previously, we conducted a common garden growth trial for seven weeks. Greenhouse conditions were set at 7:17 h dark:light photoperiod. Relative humidity and temperature settings during the day were 55% and 21˚C while nighttime conditions were 60% and 15.5˚C. We planted seeds on the same day to a standardized depth of 6 mm in a sand/potting soil mixture. Each day we watered plants until soil appeared fully wet and randomized plant position to prevent bias in temperature, light, or water regime. We used 6 seeds from 20 individuals per population for this trial to investigate both population level differences and potential family effects. At the end of the trial period, plants were carefully removed from the soil and the length of tissue above and below the caudex was measured with a caliper.

To compare emergence values between populations, we ran a two sided proportion test in the R statistical program v6.0 (R Development Core Team 2017). We analyzed any difference in the ratio of above/belowground tissue between sampling locations and the presence or absence of family effects using a completely randomized design ANOVA in SAS v9.4 (SAS Institute Inc. 2013). Analyses were run both with and without plants that did not emerge.

CONCLUDING STATEMENTS

This thesis aimed to add to our knowledge of the North American invasion of *Gypsophila paniculata* and the greater field of invasion science. In chapter II, I recreated the invasion history of *G. paniculata* using a combination of herbarium records and microsatellite analyses. Results found the presence of two genetic clusters among our seven sampling locations, suggesting at least two invasion events. When herbarium records were grouped according to contemporary population genetic clusters, two distinct expansion phases become apparent. In chapter III, gene expression profiles and phenotypic differences for populations of this species that clustered together according to microsatellite analysis but were growing in distinct environments were explored. Gene expression profiles showed differences in genes related to stress and nutrient starvation. Germination trials show that seeds collected from Washington germinated significantly quicker than those collected from Petoskey, MI and had higher levels of emergence when grown in a common garden. Family effects were found in both germination and growth trials. Future work is needed to parse out possible maternal effects from differences due to genetic architecture, however, differences in phenotypic traits that may confer increased fitness in these two distinct ecosystems are visible. The combination of these data provides greater context for this invasion and can be applied more broadly to our understanding of how invasive plant species invade, adapt, and thrive in novel environments.

BIBLIOGRAPHY

Abdelkrim J, Pascal M, Calmet C, Samadi S. 2005. Importance of assessing population genetic structure before eradication on invasive species: Examples from insular norway rat populations. Conservation Biology. 19(5): 1509–1518.

Allendorf FW, Lundquist LL. 2003. Introduction: Population Biology, Evolution, and Control of Invasive Species. Conservation Biology. 17(1): 24–30.

American Society for Testing and Materials Committee D18 on Soil and Rock. 2004. Standard Test Methods for Particle-Size Distribution (Gradation) of Soils Using Sieve. In: Annual book of ASTM standards. ATSM International.

Antunes PM, Schamp B. 2017. Constructing standard invasion curves from herbarium data—toward increased predictability of plant invasions. Invasive Plant Science and Management. 10(04): 293–303.

Archer FI, Adams PE, Schneiders BB. 2016. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources. 17(1): 5–11.

Babu C. V. S, Gassmann M. 2016. Assessing integrity of plant RNA with the Agilent 2100 Bioanalyzer System. Agilent Application Note. 5990-8850E.

Balch JK, Bradley BA, D'Antonio CM, Gómez-Dans J. 2013. Introduced annual grass increases regional fire activity across the arid western USA (1980-2009). Global Change Biology. 19(1): 173–183.

Barkoudah YI. 1962. A revision of *Gypsophila, Bolanthus, Ankyropetalum* and *Phryna*. Wentia. 9: 1–203.

Baskett CA, Emery SM, Rudgers JA. 2011. Pollinator visits to threatened species are restored following invasive plant removal. International Journal of Plant Sciences. 172(3): 411–422.

Baskin JM, Baskin CC. 2001. Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination. 2nd ed. New York: Academic Press.

Bellard C, Cassey P, Blackburn TM. 2016. Alien species as a driver of recent extinctions. Biology letters. 12(2): 20150623.

Bellard C, Rysman JF, Leroy B, Claud C, Mace GM. 2017. A global picture of biological invasion threat on islands. Nature Ecology and Evolution. 1(12): 1862-1869

Blumenthal D, Chimner RA, Welker JM, Morgan JA. 2008. Increased snow facilitates plant invasion in mixedgrass prairie. New Phytologist. 179(2): 440–448.

Blumenthal DM, Hufbauer RA. 2007. Increased plant size in exotic populations: A common-garden test with 14 invasive species. Ecology. 88(11): 2758–2765.

Boag AE, Eckert CG. 2013. The effect of host abundance on the distribution and impact of biocontrol agents on purple loosestrife (*Lythrum salicaria*, Lythraceae). Écoscience. 20(1): 90–99.

Colautti RI, Maron JL, Barrett SCH. 2009. Common garden comparisons of native and introduced plant populations: Latitudinal clines can obscure evolutionary inferences. Evolutionary Applications. 2(2): 187–199.

Conover DO, Schultz ET. 1995. Phenotypic similarity significance of countergradient variation. Trends in Ecology & Evolution. 10(6): 248–252.

Crooks JA. 2007. Lag times and exotic species: The ecology and management of biological invasions in slow-motion. Écoscience. 12(3): 316–329.

Darwent AL. 1975. The biology of Canadian weeds: 14. *Gypsophila paniculata* L. The Biology of Canadian Weeds. 55(14): 1049–1058.

Darwent AL, Coupland RT. 1966. Life History of *Gypsophila paniculata*. Weeds. 14(4): 313.

Darwent AL, Coupland RT, Skoglund NA. 1967. The potential of *Gypsophila paniculata* as a weed in Saskatchewan. Canadian Journal of Plant Science. 47(2): 125–134.

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD. 2015. The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. Molecular Ecology. 24(9): 2095–2111.

Dupuis JR, Janes JK, Malenfant RM, Gorrell JC, Andrew RL, Miller JM, Cullingham CI. 2017. The K = 2 conundrum. Molecular Ecology. 26(14): 3594–3602.

EDDMapS. 2019. Early Detection & Distribution Mapping System. The University of Georgia - Center for Invasive Species and Ecosystem Health. Available online athttp://www.eddmaps.org/.

Ellstrand NC, Elam DR. 2003. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annual Review of Ecology and Systematics. 24(1): 217–242.

Emery SM, Doran PJ. 2013. Presence and management of the invasive plant *Gypsophila paniculata* (baby's breath) on sand dunes alters arthropod abundance and community structure. Biological Conservation. 161: 174–181.

Emery SM, Doran PJ, Legge JT, Kleitch M, Howard S. 2013. Aboveground and belowground impacts following removal of the invasive species baby's breath (*Gypsophila paniculata*) on Lake Michigan sand dunes. Restoration Ecology. 21(4): 506–514.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology. 14(8): 2611–2620.

Frankham R. 2005. Resolving the genetic paradox in invasive species. Heredity. 94(4): 385–385.

Fudano T. 2007. Effect of plant density and variety on allometry of inflorescence architecture in *Gypsophila paniculata*. Journal of the Japanese Society for Horticultural Science. 76(4): 327–332.

Fuentes N, Ugarte E, Kühn I, Klotz S. 2008. Alien plants in Chile: Inferring invasion periods from herbarium records. Biological Invasions. 10(5): 649–657.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology. 21(3): 394–407.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology. 29(7): 644–52.

Greene DF, Johnson EA. 1994. Estimating the mean annual seed production of trees. Ecology. 75(3): 642–647.

Grime JP, Thompson K, Hunt R, Hodgson JG, Cornelissen JHC, Rorison IH, Hendry GAF, Ashenden TW, Askew AP, Band SR, et al. 1997. Integrated screening validates primary axes of specialisation in plants. Oikos. 79: 259–281.

Gugger PF, Fitz-Gibbon S, Pellegrini M, Sork VL. 2016. Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. Molecular Ecology. 25(8): 1665–1680.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols. 8(8): 1494–512.

Harper JL, Lovell PH, Moore KG. 1970. The shapes and sizes of seeds. Annual Review of Ecology and Systematics. 1(1): 327–356.

Harris RE. 1970. Gardening on permafrost. Publications Department of Agriculture, Canada. 1408.

Haynes GD, Gilligan DM, Grewe P, Nicholas FW. 2009. Population genetics and management units of invasive common carp *Cyprinus carpio* in the Murray-Darling Basin, Australia. Journal of Fish Biology. 75(2): 295–320.

Hulme PE. 2009. Trade, transport and trouble: Managing invasive species pathways in an era of globalization. Journal of Applied Ecology. 46(1): 10–18.

Hulme PE, Bacher S, Kenis M, Klotz S, Kühn I, Minchin D, Nentwig W, Olenin S, Panov V, Pergl J, et al. 2008. Grasping at the routes of biological invasions: a framework for integrating pathways into policy. Journal of Applied Ecology. 45(2): 403–414.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics. 11(94).

Karamanski TJ. 2000. A nationalized lakeshore: the creation and administration of Sleeping Bear Dunes National Lakeshore.

Kawano S. 1981. Trade-off relationships between some reproductive characteristics in plants with special reference to life history strategy. The Botanical Magazine Tokyo. 94(3): 285–294.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecology Letters. 7(12): 1225–1241.

Kolar CS, Lodge DM. 2001. Progress in invasion biology: Predicting invaders. Trends in Ecology and Evolution. 16(4): 199–204.

Lande R. 2015. Evolution of phenotypic plasticity in colonizing species. Molecular Ecology. 24(9): 2038–2045.

Langlet O. 1971. Two hundred years genecology. Taxon.(20): 653–721.

Larkin DJ. 2012. Lengths and correlates of lag phases in upper-Midwest plant invasions. Biological Invasions. 14(4): 827–838.

Lavoie C, Jodoin Y, De Merlis AG. 2007. How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Quebec? A historical analysis using herbarium records. Journal of Biogeography. 34(10): 1751–1761.

Leimbach-Maus Hailee B, Parks SR, Partridge CG. 2018. Genetic structure of invasive baby's breath (*Gypsophila paniculata*) populations in a freshwater Michigan dune system. Pre-print. 8042(440): 83–87.

Leimbach-Maus Hailee B., Parks SR, Partridge CG. 2018. Microsatellite primer development for the invasive perennial herb *Gypsophila paniculata* (Caryophyllaceae). Applications in Plant Sciences. 6(12): e1203.

Leinonen T, O'Hara RB, Cano JM, Merilä J. 2008. Comparative studies of quantitative trait and neutral marker divergence: A meta-analysis. Journal of Evolutionary Biology. 21(1): 1–17.

Lockwood JL, Cassey P, Blackburn T. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution. 20(5): 223–228.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications. 10(3): 689-710.

Maranon T, Grubb PJ. 1993. Physiological basis and ecological significance of the seed size and relative growth rate relationship in Mediterranean annuals. Functional Ecology. 7(5): 591–599.

McKay JK, Latta RG. 2002. Adaptive population divergence: Markers, QTL and traits. Trends in Ecology and Evolution. 17(6): 285–291.

McNair JN, Sunkara A, Frobish D. 2012. How to analyse seed germination data using statistical time-to-event analysis: non-parametric and semi-parametric methods. Seed Science Research. 22(2): 77–95.

Mehlich A. 1984. Mehlich 3 Soil Test Extractant: A modification of Mehlich 2 Extractant. Communications in Soil Science and Plant Analysis. 15(12): 1409–1416.

Meimberg H, Milan NF, Karatassiou M, Espeland EK, McKay JK, Rice KJ. 2010. Patterns of introduction and adaptation during the invasion of *Aegilops triuncialis* (Poaceae) into Californian serpentine soils. Molecular Ecology. 19(23): 5308–5319.

Michigan Department of Natural Resources. 2015. 2015 Michigan Invasive Species Grant Program Handbook.

Mooney HA, Cleland EE. 2001. The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences. 98(10): 5446–5451.

Moritz C. 1994. Defining 'Evolutionarily Significant Units'. Trends in Ecology and Evolution. 9(10): 373–375.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 4(3): 535–538.

Palsbøll PJ, Bérubé M, Allendorf FW. 2007. Identification of management units using population genetic data. Trends in Ecology and Evolution. 22(1): 11–16.

Peakall R, Smouse PE. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 6(1): 288–295.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 28(19): 2537–2539.

Pejchar L, Mooney HA. 2009. Invasive species, ecosystem services and human well-being. Trends in Ecology and Evolution. 24(9): 497–504.

Pensec F, Marmonier A, Marchal A, Gersch S, Nassr N, Chong J, Henry M, Demangeat G, Bertsch C. 2013. *Gypsophila paniculata* root saponins as an environmentally safe treatment against two nematodes, natural vectors of grapevine fanleaf degeneration. Australian Journal of Grape and Wine Research. 19(3): 439–445.

Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T, et al. 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. Agriculture, Ecosystems and Environment. 84(1): 1–20.

Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics. 52(3): 273–288.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus data. Genetics. 155: 945–959.

Pyšek P. 1995. On the terminology used in plant invasion studies. In: Plant lnvasions - General Aspects and Special Problems. Amsterdam, The Netherlands: SPB Academic Publishing. p. 71–81.

Pyšek P, Jarošík V, Müllerová J, Pergl J, Wild J. 2008. Comparing the rate of invasion by *Heracleum mantegazzianum* at continental, regional, and local scales. Diversity and Distributions. 14(2): 355–363.

R Development Core Team. 2017. R: A language and environment for statistical computing.

Raymond M, Rousset F. 1995. An exact test for population differentiation. Evolution. 49(6): 1280–1283.

Reaser JK, Meyerson LA, Cronk Q, De Poorter M, Eldredge LG, Green E, Kairo M, Latasi P, Mack RN, Mauremootoo J, et al. 2007. Ecological and socioeconomic impacts of invasive alien species in island ecosystems. Environmental Conservation. 34(2): 98–111.

Reid ML, Emery SM. 2018. Scale-dependent effects of *Gypsophila paniculata* invasion and management on plant and soil nematode community diversity and heterogeneity. Biological Conservation. 224: 153–161.

Reinhold CG. 1856. The farmer's promotion book. Pittsburgh, PA: W.S. Haven.

Rejmánek M, Richardson DM. 1996. What attributes make some plant species more invasive? Ecology. 77(6): 1655–1661.

Rejmánek M, Richardson DM, Higgins SI, Pitcairn MJ, Grotkopp E. 2005. Ecology of invasive plants: State of the art. In: Invasive Alien Species: A New Synthesis. p. 104–161.

Rice E. 2018. Assessment of invasive *Gypsophila paniculata* control methods in the northwest Michigan dunes. Masters Theses. 888.

Richards CL, Bossdorf O, Muth NZ, Gurevitch J, Pigliucci M. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecology Letters. 9(8): 981–993.

Richardson DM, Pyšek P. 2006. Plant invasions: Merging the concepts of species invasiveness and community invasibility. Progress in Physical Geography. 30(3): 409–431.

Rio DC, Ares M, Hannon GJ, Nilsen TW. 2010. Purification of RNA using TRIzol (TRI Reagent). Cold Spring Harbor Protocols. 5(6).

Roman J, Darling JA. 2007. Paradox lost: genetic diversity and the success of aquatic invasions. Trends in Ecology and Evolution. 22(9): 454–464.

Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources. 8(1): 103–106.

Russell JC, Meyer J-Y, Holmes ND, Pagad S. 2017. Invasive alien species on islands: impacts, distribution, interactions and management. Environmental Conservation. 44(4): 359–370.

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, et al. 2001. The Population Biology of Invasive Species. Annual Review of Ecology and Systematics. 32(1): 305–332.

SAS Institute Inc. 2013. SAS/ACCESS® 9.4.

Sax DF, Brown JH. 2000. The paradox of invasion. Global Ecology and Biogeography. 9: 363–371.

Schlosser W, Blatner K. 1997. Special forest products: An east-side perspective. Gen. Tech. Rep. GTR-PNW-380. Portland, Oregon.

Schlosser WE, Blatner KA, Chapman RC. 1991. Economic and Marketing Implications of Special Forest Products Harvest in the Coastal Pacific Northwest. Western Journal of Applied Forestry. 6: 67–72.

Shibuya T, Murakawa Y, Nishidate K, Nishiyama M, Kanayama Y. 2017. Characterization of flowering-related genes and flowering response in relation to blue light in *Gypsophila paniculata*. The Horticulture Journal. 86(1): 94–104.

Shih JG, Finkelstein SA. 2008. Range dynamics and invasive tendencies in *Typha latifolia* and *Typha angustifolia* in eastern North. Wetlands. 28(1): 1–16.

Shun YAO, Jian-guang LUO, Li MA, Ling-yi K. 2011. Two New Triterpenoid Saponins from the Roots of *Gypsophila paniculata* with Potent α-Glucosidase Inhibition Activity. Chinese Journal of Natural Medicines. 9(6): 2–6.

Sork VL. 2018. Genomic studies of local adaptation in natural plant populations. Journal of Heredity. 109(1): 3–15.

Stevens OA. 1957. Weights of Seeds and Numbers Per Plant. Weeds. 5(1): 46–55.

Swearingen J, Bargeron C. 2016. Invasive plant atlas of the United States. University of Georgia Center for Invasive Species and Ecosystem Health. invasiveplantatlas.org.

Van Tienderen PH. 1997. Generalists, specialists, and the evolution of phenotypic plasticity in sympatric populations of distinct species. Evolution. 51(5): 1372–1380.

USDA. 2019. Plants Profile for *Gypsophila paniculata* (baby's breath).

Usher G. 1974. A Dictionary of Plants Used by Man. London: Constable and Company Limited.

Via S, Lande R. 2006. Genotype-Environment interaction and the evolution of phenotypic plasticity. Evolution. 39(3): 505.

Vilà M, Espinar JL, Hejda M, Hulme PE, Jarošík V, Maron JL, Pergl J, Schaffner U, Sun Y, Pyšek P. 2011. Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. Ecology Letters. 14(7): 702–708.

Vitousek PM, D'Antonio CM, Loope LL, Rejmanek M, Westbrooks R. 1997. Introduced species: a significant component of human-caused global change. New Zealand Journal of Ecology. 21(1): 1–16.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 10(1): 57–63.

Weidenhamer JD, Callaway RM. 2010. Direct and indirect effects of invasive plants on soil chemistry and ecosystem function. Journal of Chemical Ecology. 36(1): 59–69.

Weir BS, Cockerham CC. 1983. Estimating F-statistics for the analysis of population structure. Evolution. 38(6): 1358.

Whitson TD, Burrill LC, Dewey SA, Cudney DW, Nelson BE, Lee RD, Parker R. 1991. Weeds of the West.

Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E. 1998. Quantifying threats to imperiled species in the United States. BioScience. 48(8): 607–615.

Williamson M, Fitter A. 1996. The varying success of invaders. Ecology. 77(6): 1661–1666.

Yang ML, Rice E, Leimbach-Maus H, Partridge CG. 2019. Identification and characterization of *Gypsophila paniculata* color morphs in Sleeping Bear Dunes National Lakeshore, MI, USA. PeerJ. 7: e7100.

Yao S, Ma L, Luo JG, Wang JS, Kong LY. 2010. New triterpenoid saponins from the roots of *Gypsophila paniculata* L. Helvetica Chimica Acta. 93(2): 361–374.

Zalewski A, Michalska-Parda A, Bartoszewicz M, Kozakiewicz M, Brzeziński M. 2010. Multiple introductions determine the genetic structure of an invasive species population: American mink *Neovison vison* in Poland. Biological Conservation. 143(6): 1355–1363.

Zedler JB, Kercher S. 2004. Causes and consequences of invasive plants in wetlands: Opportunities, opportunists, and outcomes. Critical Reviews in Plant Sciences. 23(5): 431–452.