Grand Valley State University
ScholarWorks@GVSU

Masters Theses | Graduate Research and Creative Practice

8-2018

Genetic structure of invasive baby's breath (Gypsophila paniculata) populations in a Michigan dune system

Hailee B. Leimbach-Maus
Grand Valley State University

Follow this and additional works at: https://scholarworks.gvsu.edu/theses

 Part of the Biology Commons

ScholarWorks Citation
Leimbach-Maus, Hailee B., "Genetic structure of invasive baby's breath (Gypsophila paniculata) populations in a Michigan dune system" (2018). *Masters Theses*. 905.
https://scholarworks.gvsu.edu/theses/905

This Thesis is brought to you for free and open access by the Graduate Research and Creative Practice at ScholarWorks@GVSU. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Genetic structure of invasive baby's breath (*Gypsophila paniculata*) populations in a Michigan dune system

Hailee Brianne Leimbach – Maus

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 in Biology

Department of Biology

August 2018

Dedication

I dedicate this thesis to my family, for encouraging me to work hard and to grow; to my husband Dan, for supporting me through this adventure; and finally, to Ludovico Einaudi and Logic: if it weren't for their music, I may not have completed this thesis.

Acknowledgements

I thank Grand Valley State University and the Robert B. Annis Water Resources Institute for providing me with the resources to learn and perform valuable research. Funding was provided through the Environmental Protection Agency – Great Lakes Restoration Initiative Grant to perform this research, and support me through a graduate assistantship and tuition coverage. Additional financial support was provided by Grand Valley State University through the Scholarly and Creative Excellence Catalyst Grant and the Presidential Research Grant. I thank my graduate advisor, Dr. Charlyn Partridge, for all her guidance and support throughout the research process, and for the opportunity to be a part of her lab and to grow professionally and personally. I thank my other graduate committee members, Dr. James McNair and Dr. Timothy Evans, for their direction and encouragement whenever I needed it. Finally, I thank Emma Rice, for all her assistance in field data collection for this study and support throughout this process, to Syndell Parks, for her assistance and collaboration in laboratory data collection for this study, and to Matthew Kienitz, for his assistance in field and laboratory data collection.

Abstract

Invasive species can reduce biodiversity of a system by outcompeting native species for resources, changing the physical characteristics of a habitat, and altering natural disturbance regimes. Coastal sand dune ecosystems are dynamic with elevated levels of disturbance, and as such they are highly susceptible to plant invasions. The topography, geographic distribution of preferred habitat, and disturbance regime in an ecosystem can influence where an invasive plant becomes established, its dispersal patterns, and how densely it grows. One such invasion that is of major concern to the Great Lakes dune systems is baby's breath (*Gypsophila paniculata*). The invasion of baby's breath negatively impacts native species, including rare ones such as Pitcher's thistle (*Cirsium pitcheri*). Estimating the genetic variation and structure of invasive populations can lead to a better understanding of the invasion history, and the factors influencing invasion success. Microsatellite genetic markers can be beneficial for estimating levels of diversity present within and among populations. Our research goals were to develop microsatellite primers to analyze invasive populations, quantify the genetic diversity and estimate the genetic structure of these invasive populations of baby's breath in the Michigan dune system. We identified 16 polymorphic nuclear microsatellite loci for baby's breath out of 73 loci that successfully amplified from a primer library created using Illumina sequencing technology. We analyzed 12 populations at 14 nuclear and 2 chloroplast microsatellite loci and found moderate genetic diversity, strong genetic structure among the populations (global $F_{ST} = 0.228$), and also among two geographic regions that are separated by the Leelanau peninsula. Results from a Bayesian clustering analysis suggest two main population clusters. Isolation by distance was found over all 12 populations ($R = 0.755$, $P < 0.001$) and when only cluster 2 populations were included ($R = 0.523$, $P = 0.030$); populations within cluster 1 revealed no significant relationship ($R = 0.205$, P

= 0.494). The results suggest the possibility of at least two separate introduction events to Michigan. These results provide an understanding of the invasion history and factors contributing to invasion success. Management of invasive populations can use this to identify populations of high priority.

Table of Contents

List of Tables

Table Page

List of Figures

Figure Page

Chapter I

Introduction

Invasive species

The loss of biodiversity worldwide is attributed to two main causes: habitat loss and biological invasions (Cronk and Fuller 2001). Non-native invasive plants can reduce biodiversity of a system by outcompeting native species for resources, altering trophic webs, changing the physical characteristics of a habitat (e.g. available soil nutrients and stabilization), and altering natural disturbance regimes (Cronk and Fuller 2001, Sakai et al. 2001, Pejchar and Mooney 2009). In a global study done by Early et al. (2016), it was found that one-sixth of terrestrial habitat is vulnerable to biological invasion, with areas of high biodiversity under particular threat. In addition, abiotic factors such as global climate change have been found to often increase the success of species in novel environments (Hellmann et al. 2008, Colautti and Barrett 2013, Early et al. 2016, Moran et al. 2017). In 2005, the estimated cost associated with damages and control of non-native invasive species in the United States was $120 billion annually (Pimentel et al. 2005), and that has likely increased in the past 13 years. Due to annual increases in the spread of nonnative invasive species, the threat to biodiversity will probably continue, but the long-term response of native communities and invaded ecosystems is difficult to predict (Hellmann et a. 2008).

Historically, most non-native invasive plants were introduced to the United States for the purposes of agriculture, textiles, and ornamental decoration (Mack and Lonsdale 2001, Pimentel et al. 2005). However, the main contemporary purpose for introducing non-native invasive plant species worldwide is the latter, and they are intentionally distributed through the ornamental nursery supply chain (Mack and Lonsdale 2001, Levine and D'Antonio 2003, Pimentel et al.

2005, Hulme et al. 2018). When the intentional global distribution and introduction of ornamental plants and seeds really began in the mid 1800's, the potential harm these nonnative species could cause was not yet understood. For sake of clarity, in this study we will define non-native invasive species (hereafter invasive species) according to Executive Order number 13112 (1999) as "a species that is non-native (alien) to the ecosystem under consideration, and whose introduction causes or is likely to cause economic or environmental harm or harm to human health." Although we have countless examples today of nonnative ornamental plants that become invasive and continue to cause billions of dollars in damage to agriculture, infrastructure, and biodiversity, we continue to globally distribute them and fail to caution others of their potential destruction (Pimentel et al. 2005).

Population genetics

The field of population genetics is based on the exploration of variation in allele and genotype frequencies within and among populations over time, and how evolutionary processes can shape this observed variation (Clark 2001). In studying the genetic variation within and among invasive plant populations, we can begin to understand how the populations are different, and why these differences may be important to the success, failure, and spread of an invasion (Dlugosch et al. 2015, Lawson-Handley et al. 2011, Sakai et al. 2001). During an invasive species' introduction or range expansion, events can occur that change the demography of the invasive population, such as multiple introductions, bottlenecks, genetic admixture between populations, or inbreeding (reviewed in Dlugosch and Parker 2008). These events shape the genetic diversity that we observe in contemporary populations, and can strongly influence the success of the invader (Dlugosch and Parker 2008, Wilson et al. 2009, Crawford and Whitney 2010).

Observed genetic variation within and among populations can be used to better understand a species' invasion history, as demographic processes like founders effects, isolation, and admixture and can be inferred from contemporary populations, and can increase our knowledge of the factors contributing to successful biological invasions (Sakai et al. 2001, Piya et al. 2014, Sakata et al. 2015, Moran et al. 2017). A prevailing theory in the study of population genetics is that a colonizing population will have reduced genetic diversity due to a bottleneck event that occurred during the introduction, and this reduction in alleles confounded by genetic drift can limit the genetic variation found in that population (Dlugosch and Parker 2008, Xu et al. 2015). However, we are learning that this is not always the case (Crosby et al. 2014, Hagenblad et al. 2015, Bentley and Mauricio 2016), and multiple introduction events that add novel alleles from the native range can lead to patterns of higher genetic diversity than expected (Sakata et al. 2015). Gene flow between introduced populations can further influence the genetic variation of each, and thus the structure of the invasive populations (Dlugosch and Parker 2008, Nagy and Korpelainen 2014). This increased diversity could lessen the effect of a bottleneck event and genetic drift, and lead to possible range expansion and/or increased invasiveness (Colautti and Barrett 2013, Xu et al. 2015, Moran et al. 2017).

The use of selectively neutral markers such as microsatellites can be beneficial for estimating levels of diversity present within and among populations (Freeland et al. 2011). Microsatellites, also referred to as simple sequence repeats (SSRs), are DNA motifs that are typically 1-6 nucleotides long that are repeated several to dozens of times in tandem (Hodel et al. 2016). The variability in the length of the motif repeat is due to the high rate of mutations, which are created by slippage during DNA replication. With high levels of mutation rates, microsatellites can capture contemporary population differentiation (Sun et al. 2009, Hodel et al.

2016), and can provide a baseline for starting to understand the demography of an invasion. Microsatellites are useful for estimating genetic diversity within populations, the genetic structure among populations, and pathways of gene flow (e.g. pollination and seed dispersal) among populations (Sun et al. 2009, Freeland et al. 2011, Hodel et al. 2016). Therefore, microsatellite markers are often used to understand the partitioning of genetic variation in invasive populations to better understand the population structure and invasion history (Piya et al. 2014, Hagenblad et al. 2015, Sakata et al. 2015).

Baby's breath invasion in the Michigan dunes

Baby's breath (*Gypsophila paniculata*) is a perennial forb native to the arid steppe region of Eurasia, and is invasive across North America (Gleason and Cronquist 1963, Darwent 1975, DiTomaso and Kyser et al. 2013). It is found invading several habitat types: agricultural fields and rangelands of the west, Pacific northwest and Canada, the Great Lakes dune system, and disturbed areas such as roadsides and ditches (Darwent 1975, DiTomaso and Kyser et al. 2013, TNC 2013). For this study, we will focus on the invasion of baby's breath to the Great Lakes dune system.

The Great Lakes dune system comprises the largest freshwater dune complex in the world, and as such is economically and environmentally important. Specifically, the Lake Michigan dune system is over 1000 km^2, receiving millions of visitors annually to recreate and enjoy the unique landscape (Stynes 2011). This ecosystem harbors endemic species including the federally threatened pitcher's thistle (*Cirsium pitcheri*) and endangered piping plover (*Charadrius melodus*) (Albert 2000). Because abiotic disturbances are so high in the Michigan dune system, biological invasions continuously pose a threat to the dune community (Albert

2000, Jolls et al. 2015). Alongside invaders such as bladder campion (*Silene vulgaris*) and spotted knapweed (*Centaurea stoebe* subsp. *micranthos*), the *G. paniculata* invasion has been found to negatively affect the native plant communities (TNC 2013, NPS 2015). It does this by crowding out sensitive species such as *C. pitcheri* through direct competition for limited resources, forming monotypic stands in the open dune habitat, and limiting pollinator visits to native species (Jolls et al. 2015, Emery and Doran 2013, Baskett et al. 2011). The invasion of *G. paniculata* has the potential to disrupt the dynamism of the dune landscape (Emery et al. 2013) and alter trophic levels of the native community (Baskett et al. 2011, Emery and Doran 2013) in the northwest Michigan dune system, and this threat has led to increased concern over its pervasiveness regionally and nationally.

Purpose

There has been little research into the population structure and invasion history of baby's breath. Due to the potential negative impacts of baby's breath to the Great Lakes dune system, the goal of this research is to better understand the invasive populations in Michigan and how the population distribution may be shaped by the dune landscape. To do this, we used a population genetics approach, and developed microsatellite molecular markers specific to perennial baby's breath (*Gypsophila paniculata*), which then allowed us to analyze the invasion in the Michigan dunes on a population level, and estimate the variation found within and between these populations.

Scope

The scope of this research focused on estimating the genetic structure of the baby's breath invasion in both the upper and lower peninsulas of Michigan, which to the best of our knowledge, also make up the largest contemporary infestations within the Great Lakes region. However, the molecular markers and methods developed in this study can be applied to both invasive populations of baby's breath across North America, and populations in its native range.

Assumptions

In performing this research, there are several main assumptions we made about our data. The sampling design we implemented assumes that we adequately captured the majority of principal alleles present in each location. Our conclusions of the data assume that baby's breath seeds are naturally wind-dispersed. The statistical analyses we performed also assume that the molecular markers we developed are neutral in regards to natural selection, are independently inherited, and that each population is in Hardy-Weinberg Equilibrium (HWE).

Hypotheses

Through this research, we sought to complete three objectives: (1) develop molecular markers specific to perennial baby's breath (chapter 2), (2) quantify the genetic diversity of invasive baby's breath populations in the Michigan dune system (chapter 3), and (3) estimate the genetic structure of these invasive populations (chapter 3). Based on suggestions by land managers that the initial site of the baby's breath invasion was at the Zetterberg Preserve site (TNC 2013), we hypothesized that we would see an isolation by distance effect in populations at increasing distances from this site. Given that invasive species populations often experience a bottleneck

due to founder's effect upon introduction in a novel environment, we also hypothesized that we would find limited genetic variation in these populations (Dlugosch and Parker 2008).

Significance

This research on the invasion of baby's breath in Michigan is significant for several reasons. Identifying the genetic structure of invasive baby's breath in Michigan will increase our understanding of invasion biology by characterizing how genetic variation is maintained across a highly-structured geographic region, and the influence of human-mediation in perpetuating the issue of invasive species. The development of species-specific molecular markers provides scientists and resource managers with a set of genetic tools to continue research on populations across its invasive and native range.

Furthermore, the invasion of baby's breath is seen as detrimental to the integrity of the dynamic dune landscape in Michigan, and The Nature Conservancy and the National Park Service are already implementing more focused management practices informed by this research. These agencies have revised their strategy to prioritize sites that have high visitor traffic and where population maintenance is likely buffered by hitchhiking seeds, sites that are on the northern end of Sleeping Bear Dunes National Lakeshore (hereafter Sleeping Bear Dunes or SBDNL) in order to prevent the invasion from moving past the north end of the park, and sites where baby's breath populations can be kept at preferred management levels. The adjustments made according to this research can improve the protection of the unique biodiversity in the Michigan dune system.

References

Albert, D. A. 2000. *Borne of the Wind, Michigan Sand Dunes*.

Baskett, C.A., Emery, S. M., and J.A. Rudgers. 2011. Pollinator Visits to Threatened Species Are Restored Following Invasive Plant Removal. *International Journal of Plant Sciences* 172(3): 411–422.

Bentley, K.E., and R. Mauricio. 2016. High degree of clonal reproduction and lack of large-scale geographic patterning mark the introduced range of the invasive vine, kudzu (*Pueraria montana* var. *lobata*), in North America. *American Journal of Botany* 103(8): 1499–1507.

Clark A.G. 2001. Population genetics. *Encyclopedia of Genetics*. Cambridge: Academic Press. Science Direct. Web. 10 June 2018.

Colautti, R.I., and S.C.H. Barrett. 2013. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* 342(6156): 364–366.

Crawford, K. M., and K.D. Whitney. 2010. Population genetic diversity influences colonization success. *Molecular Ecology* 19: 1253–1263.

Cronk, Q.C.B., and J.L. Fuller. 2001. Plant invaders: the threat to natural ecosystems, 241. Retrieved from https://books.google.com/books?id=t8dQAwAAQBAJ

Crosby, K., Stokes, T.O., and R.G. Latta. 2014. Evolving California genotypes of *Avena barbata* are derived from multiple introductions but still maintain substantial population structure. *PeerJ* 2: e633.

Darwent, A.L. 1975. The Biology of Canadian Weeds. 14. Gypsophila paniculata L. *Canadian Journal of Plant Science* 55(4): 1049–1058.

DiTomaso, J., and G.B. Kyser. 2013. *Weed Control in Natural Areas in the Western United States*. UC Davis Weed Research and Information Center. Retrieved from wric.ucdavis.edu%5Cn(wsweedscience.org%5Cncal-ipc.org

Dlugosch, K.M., and I.M. Parker. 2008. Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17: 431-449.

Dlugosch, K.M., Anderson, S.R., Braasch, J., Cang, F.A., and H.D. Gillette. 2015. The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology* 24(9): 2095–2111.

Early, R., Bradley, B.A., Dukes, J.S., Lawler, J.J., Olden, J.D., Blumenthal, D.M., … and A.J. Tatem. 2005. Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications* 7: 12485.

Emery, S.M., and P.J. Doran. 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, S.M., Doran, P.J., Legge, J.T., Kleitch, M., and S. Howard. 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.

Exec. Order No. 13112, 3 C.F.R. 64 (25). 1999.

Freeland, J.R., S.D. Petersen, and H. Kirk. 2011. *Molecular Ecology*, 2nd ed. Wiley-Blackwell, Hoboken, New Jersey, USA.

Gleason, H.A., and A. Cronquist. 1963. *Manual of vascular plants of Northeastern United States and adjacent Canada*. Boston, Mass: W. Grant.

Hagenblad, J., Hülskötter, J., Acharya, K.P., Brunet, J., Chabrerie, O., Cousins, S.A.O., Dar, P.A., Diekman, M., De Frenne, P., Hermy, M., Jamoneau, A., Kolb, A., Lemke, I., Plue, J., Reshi, Z.A., and B.J. Graae. 2015. Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. *BMC Genetics* 16:103.

Hellmann J.J., Byers J.E., Bierwagen B.G., and J.S. Dukes. 2008. Five Potential Consequences of Climate Change for Invasive Species. *Conservation Biology* 22: 534-543.

Hodel, R.G.J., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A., Douglas, N.A., Germain – Aubrey, C.C., Chen, S., Soltis, D.E., and P.S. Soltis. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st Century. *Applications in Plant Sciences* 4(6): 1600025.

Hulme, P.E., Brundu, G., Carboni, M., Dehnen-Schmutz, K., Dullinger, S., Early, R., Essl, F., González – Moreno, P., Groom Q.J., Kueffer, C., Kühn, I., Maurel, N., Novoa, A., Pergl, J., Pyšek, P., Seebens, H., Tanner, R., Touza, J.M., van Kleunen, M., and L.N.H. Verbrugge. 2018. Integrating invasive species policies across ornamental horticulture supply chains to prevent plant invasions. *Journal of Applied Ecology* 55(1): 92–98.

Jolls C.L., Marik, J.E., Hamzé, S.I., and K. Havens. 2015. Population viability analysis and the effects of light availability and litter on populations of *Cirsium pitcheri*, a rare, monocarpic perennial of Great Lakes shorelines. *Biological Conservation* 187: 82–90.

Lawson – Handley, L.J., Estoup, A., Evans, D.M., Thomas, C.E., Lombaert, E., Facon, B., …and H.E. Roy. 2011. Ecological genetics of invasive alien species. *BioControl* 56(4): 409–428.

Levine, J.M. and C.M. D'Antonio. 2003. Forecasting Biological Invasions with Increasing International Trade. *Conservation Biology* 17: 322 – 326.

Mack, R.N., and W.M. Lonsdale. 2001. Humans as global plant dispersers: getting more than we bargained for. *BioScience* 51(2): 9 –102.

Moran, E.V., Reid, A., and J.M. Levine. 2017. Population genetics and adaptation to climate along elevation gradients in invasive *Solidago canadensis*. *PLoS ONE* 12(9): e0185539.

Nagy, A.M., and H. Korpelainen. 2014. Population genetics of Himalayan balsam (*Impatiens glandulifera*): comparison of native and introduced populations. *Plant Ecology & Diversity* 8:3: 317 – 321.

National Park Service (NPS). 2015. "Nonnative Species." Website https://www.nps.gov/slbe/learn/nature/nonnativespecies.htm [accessed 10 November 2017].

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

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

Piya, S., Nepal, M.P., Butler, J.L., Larson, G.E., and A. Neupane. 2014. Genetic diversity and population structure of sickleweed (*Falcaria vulgaris*; Apiaceae) in the upper Midwest USA. *Biological Invasions* 16(10): 2115–2125.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., … and S.G. Weller. 2001. The population biology of invasive species. *Annual Review Ecology and Systematics* 32: 305–332.

Sakata, Y., Itami, J., Isagi, Y., and T. Ohgushi. 2015. Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range. *Journal of Plant Research* 128(6): 909–921.

Sun, J.X., Mullikin, J.C., Patterson, N., and D.E. Reich. 2009. Microsatellites are molecular clocks that support accurate inferences about history. *Molecular Biology and Evolution* 26(5): 1017–1027.

Stynes, D. 2011. *Economic Benefits to Local Communities from National Park Visitation and Payroll, 2010*. NPS Natural Resource Report. East Lansing, MI.

The Nature Conservancy, and partners. 2013. Lake Michigan Coastal Dunes Restoration Report. The Nature Conservancy. Website https://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/michigan/lake-michigan-dune-report-2013.pdf [accessed 01 November 2017].

Wilson, J.R.U., Dormontt, E.E., Prentis, P.J., Lowe, A.J., and D.M. Richardson. 2009. Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution* 24(3): 136 – 144.

Xu, C.-Y., Tang, S., Fatemi, M., Gross, C.L., Julien, M.H., Curtis, C., and R.D. van Klinken. 2015. Population structure and genetic diversity of invasive *Phyla canescens*: implications for the evolutionary potential. Ecosphere 6(9): 162.

Chapter II

Leimbach – Maus et al. – *Gypsophila paniculata* microsatellites

Microsatellite primer development for invasive perennial herb, *Gypsophila paniculata* (Caryophyllaceae)[1]

Hailee B. Leimbach-Maus[2,3], Syndell R. Parks[2], and Charlyn G. Partridge[2]

[2]Annis Water Resources Institute, Grand Valley State University (AWRI-GVSU), 740 W. Shoreline Dr., Muskegon, MI 49441 USA

Email addresses: HBLM: hailee.leimbachmaus@gmail.com

SRP: syndell.parks@gmail.com

CGP: partridc@gvsu.edu

Number of words: 1219.

[1]Manuscript received _____; revision accepted_____.

[3]Author for correspondence: hailee.leimbachmaus@gmail.com

The authors thank the Grand Valley State University Center for Scholarly and Creative Excellence Catalyst Grant (C.G.P) and the Environmental Protection Agency – Great Lakes Restoration Initiative Grant (C.G.P., Grant #00E01934) for financial support. The authors also thank E.K. Rice for help in field sampling, and M. Kienitz for help in the laboratory.

ABSTRACT

- *Premise of the study: Gypsophila paniculata* L. (baby's breath) is an herbaceous perennial that has invaded much of northern and western United States and Canada, outcompeting and crowding out native and endemic species. Microsatellite primers were developed to analyze the genetic structure of invasive populations.
- *Methods and Results:* We have identified 16 polymorphic nuclear microsatellite loci for *G. paniculata* out of 73 loci that successfully amplified from a primer library created using Illumina sequencing technology. The developed primers amplified microsatellite loci in 3 invasive populations in Michigan. Primers amplified di-, tri-, and tetra-nucleotide repeats. Five of these developed primers also amplified in *G. elegans*.
- *Conclusions:* These markers will be useful in characterizing the genetic structure of invasive populations throughout North America to aid targeted management efforts, and in native Eurasian populations to better understand invasion history.

Key words: Genetic diversity; *Gypsophila paniculata*; invasive species; microsatellites.

INTRODUCTION

The herbaceous perennial forb, *Gypsophila paniculata* L., was introduced to North America in the late 1800's (Darwent and Coupland, 1966). Invasive populations have since been documented throughout the northern and western United States and Canada, specifically in agriculture fields, rangeland, roadsides, and sandy coastlines along the Great Lakes (Darwent, 1975; Emery and Doran, 2013). Despite its wide invasive range, little information exists on how populations throughout North America are related or how they are spreading. Due to its aggressive invasion, negative impacts on native biota (Emery and Doran, 2013), and a lack of data regarding its spread, it is important to develop molecular markers that can characterize the genetic structure of invasive populations of *G. paniculata*. These markers will be directly used to investigate invasions within the Lake Michigan coastal dune system where an 1,800-acre infestation occurs (TNC, 2013). However, these markers and optimized protocols can be used to characterize populations of *G. paniculata* throughout its invasive and native range to further assess its invasion history and spread.

Calistri et al. (2014) examined the genetic relationship of 5 *Gypsophila spp*. (including *G. paniculata*) within their native range and 13 commercial hybrid strains using a combination of amplified fragment length polymorphisms (AFLPs), inter simple sequence repeats (ISSRs), target region amplification polymorphism (TRAP), and universal chloroplast simple sequence repeats (cpSSRs). However, the majority of these markers are dominant and thus do not fully distinguish between homozygotes and heterozygotes, a characteristic that would allow for fine-scale population genetic analyses (Freeland et al., 2011). Thus, the development of microsatellite markers for *G. paniculata* is necessary to adequately characterize invasive populations throughout North America.

METHODS AND RESULTS

Microsatellite Library Development, Assembly and Identification—

Adventitious buds growing from the caudex of five *G. paniculata* plants were collected from Sleeping Bear Dunes National Lakeshore (hereafter Sleeping Bear Dunes or SBDNL) in 2015 to develop the microsatellite library. Tissue was stored in indicator silica until DNA extraction. Genomic DNA was extracted using DNeasy plant mini kits (QIAGEN, Hilden, Germany), with modifications including extra wash steps with AW2 buffer. Extracted DNA was run through Zymo OneStep PCR Inhibitor Removal Columns twice (Irvine, California, USA), and checked using a Thermo Fisher Scientific Nanodrop 2000 (Waltham, Massachusetts, USA). For microsatellite library development, each sample was diluted to 50 ng/µL and submitted to Cornell University, Department of Ecology and Evolutionary Biology (CU-EEB). Libraries were then submitted to the Sequencing and Genotyping Facility at the Cornell Life Science Core Laboratory Center (CLC) for sequencing using a 2x250 paired-end format on an Illumina MiSeq (Appendix S1). Raw sequence files for the microsatellite library have been deposited to NCBI's Short Read Archive (Bioproject No: PRJNA431197). A total of 58,907 contigs containing microsatellite loci were obtained. For primer design, Msatcommander (v 1.0.3) (Faircloth, 2008) identified 3,892 potentially unique primers that yielded products of 150-450 bp, had a GC content between 30-70%, and that had a T_m between 58-62ºC, with an optimum of 60ºC (Appendix S2).

Primer Optimization—

Prior to PCR optimization, contigs containing potential primers were aligned using ClustalOmega to ensure they were targeting unique microsatellite regions (Sievers et al., 2011).

We focused on 107 primer pairs that consisted of either tetrameric, trimeric, or dimeric motifs, and yielded products between 150-300 bp. Of the 107 primer pairs that were tested, 73 successfully amplified, and 16 were determined to be polymorphic and easily scorable (Appendix S3). DNA from leaf tissue collected from three populations (Zetterberg Preserve, SBDNL, Petoskey State Park) along eastern Lake Michigan in 2016 was used for primer optimization (population geographic details included as footnote on Table 1.2). A minimum of 30 tissue samples were collected from each population. Tissue storage and DNA extraction methods are the same as previously stated.

PCR reactions consisted of 1x KCl buffer (Thermo Fisher, Waltham, Massachusetts, USA), 2.0-2.5 mM $MgCl_2$ depending on the locus (Thermo Fisher, Waltham, Massachusetts, USA) (Table 1.1), 300 µM dNTP (New England BioLabs, Ipswich, Massachusetts, USA), 0.08 mg/mL BSA (Thermo Fisher, Waltham, Massachusetts, USA), 0.4 µM forward primer fluorescently labeled with either FAM, VIC, NED, or PET (Applied Biosystems, Foster City, California, USA), 0.4 µM reverse primer (Integrated DNA Technologies, Coralville, Iowa, USA), 0.25 units of *Taq* polymerase (Thermo Fisher, Waltham, Massachusetts, USA), and a minimum of 50 ng DNA template, all in a 10.0 µL reaction volume. The thermal cycle profile consisted of 94˚C for 5 minutes followed by 35 cycles of 94˚C for 1 minute, primer-specific annealing temperature (Table 1.1) for 1 min, 72˚C for 1 min, and a final elongation step of 72˚C for 10 minutes. Successful amplification was determined by visualizing the amplicons on a 2% agarose gel stained with ethidium bromide. Fragment analysis of the amplicons was performed on an ABI3130xl Genetic Analyzer (Applied Biosystems, Foster City, California, USA).

Microsatellite marker data analysis—

Alleles were scored using Genemapper v5 software (Applied Biosystems, Foster City, California, USA), and the software Micro-Checker v2.2.3 was used to identify null alleles and potential scoring errors from stuttering or large allele dropout (Van Oosterhout et al., 2004, Van Oosterhout et al., 2006). There was no significant evidence of null alleles ($p > 0.05$) in the Zetterberg Preserve and SBDNL populations. However, null alleles were suggested for loci BB_3968, BB_5021, and BB_8681 in the Petoskey State Park population (Table 1.2). Homozygote excess for Petoskey State Park is not surprising, given this population's reduced number of alleles at each locus and small comparative population size. We characterized genetic diversity by examining the number of alleles, and expected and observed heterozygosity for each locus averaged over each population (Table 1.2) using the package STRATAG in the R statistical program (Archer et al., 2016). The number of observed alleles ranged from 1 – 10. Some loci were monomorphic for one population, but polymorphic when analysis included all populations (e.g., BB_4258).

The Zetterberg Preserve population displayed slightly higher heterozygosity values than Sleeping Bear Dunes, but the Petoskey State Park population had much lower heterozygosity values in comparison. A probability test for Hardy-Weinberg Equilibrium (HWE), calculation of the fixation index (F_{IS}), and linkage disequilibrium were performed in GENEPOP 4.2 (Raymond and Rousset, 1995; Rousset, 2008). The default parameters of a Markov Chain Monte Carlo (MCMC) iteration were used to calculate HWE. All loci were in HWE except locus BB_3968 for SBDNL, and locus BB_21680 for Petoskey State Park (Table 1.2). The F_{IS} estimates were calculated using the probability model following Robertson and Hill (1984). Statistical tests for genetic linkage disequilibrium were performed using the log likelihood ratio statistic (G-test) and

MCMC algorithm by Raymond and Rousset (1995). Two pairs of loci were significantly out of linkage disequilibrium ($p < 0.05$) for both Zetterberg Preserve and Sleeping Bear Dunes: BB_5021 and BB_2888, and BB_3913 and BB_1355. Out of 16 loci, five successfully amplified in a related species *G. elegans* (BB_4443, BB_4258, BB_7213, BB_5151, BB_1355) (Table 1.3).

CONCLUSIONS

The 16 microsatellite primers developed for *G. paniculata* provide a tool for estimating genetic diversity and structure of invasive populations, which will aid in understanding its invasion history, identifying source populations, and examining dispersal patterns. Though we developed these markers to study the Lake Michigan dune system invasion, it is invasive throughout North America. With these markers, we can begin to understand the invasion of *G. paniculata* in North America in order to improve management efforts and prevent the further spread of this species.

LITERATURE CITED

Calistri, E., M. Buiatti, and P. Bogani. 2014. Characterization of Gypsophila species and commercial hybrids with nuclear whole-genome and cytoplasmic molecular markers. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*. 150: 1–11.

Darwent, A.L. 1975. The biology of Canadian weeds. 14. *Gypsophila paniculata* L. *Canadian Journal of Plant Science.* 55: 1049–1058.

Darwent, A.L., and R.T. Coupland. 1966. Life history of *Gypsophila paniculata*. *Weeds.* 14: 313–318.

Emery, S.M., and P.J. Doran. 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.

Freeland, J.R., S.D. Petersen, and H. Kirk. 2011. Molecular Ecology, 2nd ed. Wiley-Blackwell, Hoboken, New Jersey, USA.

Sievers, F., A. Wilm, D.G. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology.* 7: 539.

The Nature Conservancy, and partners. 2013. Lake Michigan Coastal Dunes Restoration Report. The Nature Conservancy. Website https://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/michigan/lake-michigan-dune-report-2013.pdf [accessed 01 November 2017].

Faircloth, B.C. 2008. Msatcommander: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*. 8: 92–94.

Archer, F.I., P.E. Adams, and B.B. Schneiders. 2016. STRATAG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources.* 17: 5–11.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 4.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*. 86: 248–249.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources.* 8: 103–106.

Robertson, A., and W.G. Hill. 1984. Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics.* 107: 703–718.

Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources.* 4: 535–538.

Van Oosterhout, C., D. Weetman, and W.F. Hutchinson. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Resources.* 6: 255–256.

DATA ACCESSIBILITY

A summary of the microsatellite library development and sequence analysis protocols (unpublished data) provided to us by CU-EEB are in Appendix S1. Fasta sequences for the 16 microsatellite primers developed here are in Appendix S5. The fasta file listing all identified contigs containing microsatellite regions are in Appendix S4. Potential primer pairs for the identified microsatellite-containing contigs are in Appendix S2. The 107 *G. paniculata* – specific primer pairs tested during primer optimization are in Appendix S3. Raw sequence files for the microsatellite library have been deposited to NCBI's Short Read Archive (Bioproject No: PRJNA431197) and microsatellite sequences have been deposited to GenBank.

TABLES

Table 1.1. Characteristics of 16 nuclear microsatellite loci developed for *Gypsophila paniculata.*

Locus[a]	Primer sequences (5' - 3')	Repeat motif	Allele size range (bp)	Annealing temperature (°C)	Fluorescent label
BB_3335[b]	F: TCCACCAAACTCTTAAACTGCC R: CACAGACACAAAGGATCCAACC	$(AGG)_5$	215 - 244	62	NED
BB_3913	F: GGCTGTCGGGTAATAAACACAG R: TCCCAACTCAAGTCATAGCCTAG	$(ACAG)_5$	159 - 171	62	PET
BB_5567	F: GGCTAGGGAAAGTAGGAAGACC R: CGTGTCCTGTTTCTCCATGATC	$(AAT)_5$	198 - 222	62	VIC
BB_4443	F: TAGGGTGGGTGCTTGTACTAAC R: AAAGTGGTGCTGCAGAAGAATC	$(AAG)_{16}$	171 - 211	62	NED
BB_21680	F: ACTACACACAGACTCGATCCTC R: CTTTGATTGTTTGGTGTAAGTTGC	$(AAAG)_5$	199 - 218	62	PET
BB_3968	F: CATGGAGGACAATGAGAAGACG R: ACGGTGGTAATGAAGTTTGGTG	$(AGG)_6$	207 - 219	62	FAM
BB_1355	F: GCTGATCTTTGTCGTCAGGAAG R: ACTCTAGGTGTTAGGAAGGCAC	$(AAAC)_5$	220 - 224	62	NED
BB_5151	F: TCCACCTTATAACTCACCACCC R: TGAGGAAGGATAACAGCTCTCG	$(ACC)_5$	205 - 210	62	PET
BB_14751[b]	F: CCTCAAACCCTAACAATGCTCC R: TCAGCCGATCCTCTAACACG	$(AAG)_{12}$	195 - 248	62	FAM
BB_4258[b]	F: TCACAAGAGGCCCAATTTCTTC R: ACTTGAACCCGAACCTATACCC	$(AAT)_5$	178 - 195	62	VIC
BB_6627	F: CAAACTCAACCAACCAGACACC R: CACCTCAGCAACAACAGAGTG	$(AAAC)_5$	151 - 155	62	FAM
BB_5021	F: ATTGTCGGTGGTCATTGGTTTC R: CTTAGTCCGCAGTGTAAACAAAG	$(AC)_8$	162 - 207	62	VIC
BB_7213[b]	F: TTGCATTCCCACCATTTCATCC R: AGCCAACCTCGTATTAATTGCC	$(AC)_7$	161 - 248	62	PET
BB_2888[b]	F: CTTCATTCATGTACAAGAGCGC R: AGAACTGGCTATGGATCGAAATG	$(AC)_{16}$	219 - 232	63	FAM
BB_8681[b]	F: ATCTCCAGTTTCCGTGATTTGC R: TACGTCACAAGAGCTTTCAACC	$(ACC)_8$	204 - 222	62	NED
BB_31555	F: TGTATAACTGAGATAACCCAGACG R: TTGTTACCTTGTTCCGGCAAAG	$(AC)_7$	150 - 156	62	VIC

[a]Optimal annealing temperature was 62˚C for all loci except for BB_2888, where it was 63 ˚C. [b]Used 2.5 mM $MgCl_2$ per sample to amplify locus.

Table 1.2. Genetic properties of the 16 developed nuclear microsatellite loci for 3 populations of *Gypsophila paniculata.*

	Zetterberg Preserve (n = 30)					Sleeping Bear Dunes (n = 33)					Petoskey State Park (n = 30)				
Locus	*A*	H_O[a]	H_E	F_{IS}	*r*	*A*	H_O[a]	H_E	F_{IS}	*r*	*A*	H_O[a]	H_E	F_{IS}	*r*
BB_3335	8	0.70	0.82	0.09	0.07	7	0.80	0.80	-0.01	0.00	3	0.50	0.56	0.02	0.05
BB_3913	4	0.74	0.68	-0.04	-0.06	4	0.59	0.61	0.05	0.00	2	0.13	0.13	-0.06	-0.06
BB_5567	4	0.63	0.60	-0.04	-0.04	4	0.63	0.63	0.01	-0.01	3	0.53	0.62	0.14	0.05
BB_4443	10	0.60	0.64	0.03	0.03	8	0.76	0.77	-0.02	0.00	5	0.80	0.71	-0.06	-0.09
BB_21680	4	0.53	0.56	0.01	0.02	3	0.52	0.56	0.02	0.05	4	0.67*	0.58	0.23	-0.10
BB_3968	3	0.50	0.42	-0.14	-0.11	4	0.35*	0.49	0.15	0.13	2	0.04	0.04	–	0.28
BB_5021	5	0.77	0.73	0.00	-0.05	5	0.54	0.60	0.15	0.02	1	–	–	–	0.30
BB_7213	5	0.63	0.61	-0.05	-0.02	3	0.59	0.58	-0.13	0.04	2	0.37	0.38	0.04	0.01
BB_2888	6	0.90	0.80	-0.11	-0.07	6	0.71	0.82	0.20	0.06	2	0.57	0.50	-0.13	-0.08
BB_8681	4	0.47	0.51	0.23	0.01	4	0.38	0.36	-0.07	-0.05	3	0.41	0.53	0.28	0.17
BB_1355	2	0.17	0.26	0.37	0.13	2	0.14	0.13	-0.07	-0.08	2	0.37	0.51	0.29	0.13
BB_5151	2	0.20	0.18	-0.10	-0.11	2	0.49	0.51	0.11	0.00	2	0.03	0.03	–	-0.02
BB_14751	6	0.79	0.71	-0.06	-0.07	6	0.76	0.76	-0.02	-0.01	3	0.20	0.29	0.21	0.12
BB_4258	3	0.40	0.33	-0.11	-0.22	1	0.00	0.00	–	0.00	1	–	–	–	0.00
BB_6627	2	0.30	0.26	-0.16	-0.16	2	0.46	0.49	-0.03	0.04	1	–	–	–	0.00
BB_31555	4	0.60	0.67	0.07	0.05	4	0.46	0.59	0.11	0.12	2	0.23	0.26	0.10	0.04

Note: n = number of individuals sampled; A = number of alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = fixation index; r = null allele frequency; – = data not available because only 1 allele present at locus or because $A > 1$ is due to rare allele.

Data based on 3 populations with the following geographic coordinates: Zetterberg Preserve = 44.68231, -86.25322; Sleeping Bear Dunes = 44.87372, -86.06170; Petoskey State Park = 44.40859, -84.91238. All 3 populations are located in the northwest region of Michigan's lower peninsula.

[a]Significant deviation from Hardy–Weinberg equilibrium: *P < 0.05.

Table 1.3. Results of cross - amplification of microsatellite loci isolated from *Gypsophila paniculata* and tested in 12 *G. elegans* individuals.[a]

		Successful amplification (X) in each *G. elegans* individuals (n = 12).											
		GelegansA	GelegansB	GYEL01	GYEL02	GYEL07	GYEL08	GYEL09	GYEL12	GYEL13	GYEL15	GYEL18	GYEL19
Locus	Allele size range (bp)												
BB_3335													
BB_3913													
BB_5567													
BB_4443	152	X	X	X	X	X	X	X	X	X	X	X	X
BB_21680													
BB_3968													
BB_1355	209 - 222	X	X	X	X	X	X	X	X	X	X	X	X
BB_5151	202 - 205	X	X	X	X	X	X	X	X	X	X	X	X
BB_14751													
BB_4258	160 - 169	X	X	X	X	X	X	X	X	X	X	X	X
BB_6627													
BB_5021													
BB_7213	161 - 171	X	X	X	X	X	X	X	X	X	X	X	X
BB_2888													
BB_8681													
BB_31555													

[a] *G. elegans* tissue sourced from individuals grown in greenhouse at AWRI-GVSU facilities.

Chapter III

Genetic structure of invasive baby's breath (*Gypsophila paniculata*) populations in a freshwater Michigan dune system

Hailee B. Leimbach – Maus[1,2], Syndell R. Parks[1], and Charlyn G. Partridge[1]

[1]Annis Water Resources Institute, Grand Valley State University (AWRI-GVSU), 740 W. Shoreline Dr., Muskegon, MI 49441 USA

[2]Corresponding author: Hailee B. Leimbach – Maus

Annis Water Resources Institute, Grand Valley State University (AWRI-GVSU)

740 W. Shoreline Dr., Muskegon, MI 49441 USA

Telephone: (440) 488-8042 (no fax)

Email addresses: hailee.leimbachmaus@gmail.com

Running title: Leimbach – Maus et al. – invasive baby's breath population structure

Word count: 7,145.

ABSTRACT

Coastal sand dunes are dynamic ecosystems with elevated levels of disturbance, and as such they are highly susceptible to plant invasions. One such invasion that is of major concern to the Great Lakes dune systems is the perennial baby's breath (*Gypsophila paniculata*). The invasion of baby's breath negatively impacts native species such as the federal threatened Pitcher's thistle (*Cirsium pitcheri*) that occupy the open sand habitat of the Michigan dune system. Our research goals were to (1) quantify the genetic diversity of invasive baby's breath populations in the Michigan dune system, and (2) estimate the genetic structure of these invasive populations. We analyzed 12 populations at 14 nuclear and 2 chloroplast microsatellite loci. We found strong genetic structure among populations of baby's breath sampled along Michigan's dunes (global $F_{ST} = 0.228$), and also among two geographic regions that are separated by the Leelanau peninsula. Pairwise comparisons using the nSSR data among all 12 populations yielded significant F_{ST} values. Results from a Bayesian clustering analysis suggest two main population clusters. Isolation by distance was found over all 12 populations ($R = 0.755$, $P < 0.001$) and when only cluster 2 populations were included ($R = 0.523$, $P = 0.030$); populations within cluster 1 revealed no significant relationship ($R = 0.205$, $P = 0.494$). The results suggest the possibility of at least two separate introduction events to Michigan. These results provide an understanding of the invasion history and factors contributing to invasion success.

Key words: Invasive species, genetic diversity, genetic structure, invasion history, microsatellites, *Gypsophila paniculata*.

INTRODUCTION

Coastal sand dunes are dynamic ecosystems. Both the topography and biological community are shaped by disturbance from fluctuations in water levels, weather patterns, and storm events (Arbogast and Loope 1999, Everard et al. 2010, Blumer et al. 2012). In these primary successional systems, vegetation plays an imperative role in trapping sand and soil, both of which accumulate over time and result in sand stabilization and dune formation (Cowles 1899, Olson 1958, Arbogast 2015). Much of the vegetative community native to coastal dune systems is adapted to the harsh conditions posed by the adjacent coast, and some species require early successional, open habitat to thrive (Albert 2000; Everard et al. 2010). For example, dune species such as Marram grass (*Ammophila brevigulata*), Lake Huron tansy (*Tanacetum huronense*), and Pitcher's thistle (*Cirsium pitcheri*) are adapted to sand burial and will continue to grow above the sand height as it accumulates (Albert, 2000). It is the heterogeneous topography and successional processes due to continuous disturbance that makes dune systems so unique (Everard et al. 2010).

Because coastal dune ecosystems have naturally elevated levels of disturbance, they are highly susceptible to plant invasions (Jorgensen and Kollman 2009, Carranza et al. 2010, Rand et al. 2015). Invasive plant species are known to be adept at colonizing disturbed areas, and in sparsely-vegetated dune systems that are often in early stages of succession, the opportunities for invasive colonizers are great (Cowles 1899, Grimes 1979, Baker, 1986, Sakai et al. 2001). Coastal dune systems also typically have a gradient of increasing stages of succession (Cusseddu et al. 2016) and this heterogeneous structure can further promote various stages of an invasion, such as colonization, dispersal, and range expansion (With 2001, Theoharides and Dukes 2007).

Within the Michigan dunes system, these successional processes have resulted in a patchwork pattern with alternating areas of open dune habitat, interdunal swales, shrub-scrub, and forested pockets scattered across the landscape (Cowles 1899, Albert 2000, Blumer et al. 2012). This landscape structure can play an important role in shaping species migration, invasive spread, and population demographics (With 2001, Theoharides and Dukes 2007, Jorgensen and Kollman 2008), thus potentially driving patterns of population structure for invasive species. However, management of dune communities can also have a strong impact on invasive populations, as well as the native plant community and the landscape they are invading. Invasive beach grasses *Ammophila breviligulata* and *A. arenaria,* and the management practices used to reduce their impact, led to changes in the morphology of the coastal dune ecosystem by decreasing the maximum dune elevation (Zarnetske et al 2010). Thus, just as a landscape can shape invasive populations, a species invasion can significantly alter the dune landscape (Grimes 1979, Cowles 1899, Sakai et el. 2001, Zarnetske et al. 2010).

In addition to the landscape, demographic processes during a species invasion also shape the genetic structure observed in contemporary populations. Multiple separate introduction events can result in contemporary populations that are genetically different from one another and from the native range (Dlugosch and Parker 2008, Crosby et al. 2014, Hagenblad et al. 2015). Bottleneck events during an introduction can further limit the genetic variation in the invasive range, though this has not necessarily been found to limit the success of an invader (Dlugosch and Parker 2008, Xu et al. 2015). Additionally, genetic admixture and inbreeding can lead to highly-structured populations, and the effect of these processes can be further influenced by the landscape structure and habitat heterogeneity (Crosby et a. 2014, Nagy and Korpelainen 2014, Moran et al. 2017, Bustamante et al. 2018). Observed levels of genetic variation and population

structure can be used to better understand a species' invasion history and the factors contributing to successful biological invasions (Crosby et a. 2014, Piya et al. 2014, Sakata et al. 2015, Moran et al. 2017).

Perennial baby's breath (*Gypsophila paniculata*) has been identified as a species of concern due to its impact on the integrity of the Michigan dune system (DNR 2015). A perennial iteroparous forb native to the Eurasian steppe region (Darwent and Coupland 1966), baby's breath has been found to negatively impact the coastal dune community in Michigan by crowding out sensitive species such as Pitcher's thistle (*Cirsium pitcheri*) through direct competition for limited resources, forming monotypic stands in the open dune habitat, preventing the reestablishment of native species, and limiting pollinator visits to native species (Baskett et al. 2011, Jolls et al. 2015, Emery and Doran 2013). Baby's breath dispersal is thought to be primarily wind-driven (Darwent and Coupland 1966), which is also the mechanism that shapes the dunes. Following seed maturity, the stems of baby's breath individuals become dry and brittle, breaking at the caudex and forming tumbleweed masses that can disperse roughly 10,000 seeds per plant up to 1 km (Darwent and Coupland 1966). Due to the topography and the heterogeneous habitat of the dune systems, the wind patterns of this landscape have the potential to shape the structure of invasive baby's breath populations. Wind can drive the direction and distance that baby's breath tumbleweeds are dispersing, and it is possible that wind patterns could both promote gene flow or limit it by driving tumbleweeds into undesirable habitat. Additionally, the steep topography in parts of the dunes could be preventing the tumbleweeds from dispersing significant distances. With these interactive processes in mind, this study explored the genetic structure of invasive populations of baby's breath within the Michigan coastal dune system. The goals of this research were to (1) quantify the genetic diversity of

invasive baby's breath populations in the Michigan dune system, and (2) estimate the genetic structure of these invasive populations. By estimating the genetic variation and structure of these invasive populations, we can better understand the impact of the landscape and its dynamic processes on the invasion history and success of this plant invasion.

METHODS

Study area and sample collection

To determine population structure on a regional scale in Michigan, we collected leaf tissue samples from plants at 12 different sites in the summers of 2016 – 2017. All sites were located in areas of known infestation along the dune system of Michigan (Figure 2.1), and the majority have a history of treatment primarily by The Nature Conservancy, the Grand Traverse Regional Land Conservancy, and the National Park Service (TNC 2013). Eleven sites were located along Lake Michigan in the northwest lower peninsula of Michigan, and one was located on Lake Superior in the upper peninsula. We collected leaf tissue samples (5-10 leaves per individual) from a minimum of 20 individuals per site (maximum of 35), and stored them in individual coin envelopes in silica gel until DNA extractions took place (total n = 313). Site locations in Michigan (Table 2.1) were separated by a minimum of 10 km and a maximum of 202 km. We subjectively chose individuals to be sampled by identifying a visibly infested area, selecting individuals regardless of size, and walking a minimum of ca. 5 meters in any direction before choosing another plant to minimize the chance of sampling closely related individuals. We observed that the number of individuals at the Petoskey State Park site was smaller and patchier than the others (~60 individuals total), so we conducted sampling more opportunistically. This opportunistic sampling involved collecting tissue from individuals that were less than 5 m apart,

and in some areas sampling from all individuals (ca. 3 – 4 individuals) within a small patch (ca. 5m x 5m).

Microsatellite genotyping

We extracted genomic DNA from all samples using DNeasy plant mini kits (QIAGEN, Hilden, Germany) and followed supplier's instructions with minor modifications, including an extra wash step with AW2 buffer. We then ran the extracted DNA twice through Zymo OneStep PCR Inhibitor Removal Columns (Zymo, Irvine, CA) and quantified the concentrations on a Nanodrop 2000 (Waltham, Massachusetts, USA). We included deionized water controls in each extraction as a quality control for contamination.

We amplified samples at 14 polymorphic nuclear microsatellite loci (hereafter nSSRs) that were developed specifically for analysis of *G. paniculata* using Illumina sequencing technology (Table 2.2) (Leimbach-Maus et al. in prep). We conducted polymerase chain reactions (PCR) using a forward primer with a 5'-fluorescent labeled dye (6-FAM, VIC, NED, or PET) and an unlabeled reverse primer. PCR reactions consisted of 1x KCl buffer, 2.0-2.5 mM $MgCl_2$ depending on the locus, 300 µM dNTP, 0.08 mg/mL BSA, 0.4 µM forward primer fluorescently labeled with either FAM, VIC, NED, or PET, 0.4 µM reverse primer, 0.25 units of Taq polymerase, and a minimum of 50 ng DNA template, all in a 10.0 µL reaction volume (Leimbach-Maus et al. in prep). The thermal cycle profile consisted of denaturation at 94˚C for 5 minutes followed by 35 cycles of 94˚C for 1 minute, annealing at 62˚C for 1 min, extension at 72˚C for 1 min, and a final elongation step of 72˚C for 10 minutes.

Each sample was also amplified at 2 universal chloroplast microsatellite loci (hereafter cpSSRs) previously developed for *Nicotiana tabacum* L. (Chung and Staub 2003) (ccssr4,

ccssr9) (Table 2.2). PCR reaction details and fragment lengths from Calistri et al. (2014) were used for *G. paniculata*. PCR reactions were conducted using a forward primer with a 5'-fluorescent labeled dye (Applied Biosystems, Foster City, CA) and an unlabeled reverse primer. PCR reactions for the cpSSRs are the same as detailed above for the nuclear loci. The thermal cycler profile for cpSSRs is as follows: denaturation at 94˚C for 5 minutes followed by 30 cycles of 94˚C for 1 minute, annealing at 52˚C for 1 minute, extension at 72˚C for 1 minute, and a final elongation step of 72˚C for 8 minutes.

We determined successful amplification by visualizing the amplicons on a 2% agarose gel stained with ethidium bromide. We multiplexed PCR amplicons according to dye color and allele size range (Table 2.2), added LIZ Genescan 500 size standard, denatured with Hi-Di Formamide at 94˚C for four minutes, and then performed fragment analysis on an ABI3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA) following instrument protocols. We genotyped individuals using the automatic binning procedure on Genemapper v5 (Applied Biosystems, Foster City, CA), and constructed bins following the Genemapper default settings. To account for the risk of genotyping error when relying on an automated allele-calling procedure, we visually verified that all individuals at all loci were correctly binned to minimize errors caused by stuttering, low heterozygote peak height ratios, and split peaks (DeWoody et al. 2006, Guichoux et al. 2011).

Quality control

Prior to any analysis, we used multiple approaches to check for scoring errors (DeWoody et al. 2006). We checked nSSR genotypes for null alleles and potential scoring errors due to stuttering and large allele dropout using the software Micro-Checker v2.2.3 (Van Oosterhout et al. 2004,

Van Oosterhout et al. 2006). Prior to marker selection, the loci used in this study were previously checked for linkage disequilibrium (Leimbach-Maus et al. in prep). We checked for heterozygote deficiencies in the package STRATAG in the R statistical program. We then screened our data for individuals with more than 20% missing loci and for loci with more than 10% missing individuals (Gomes et al. 1999; Archer et al. 2016). We found none, so all individuals and loci remained for further analyses. In addition, we genotyped 95 individuals twice to ensure consistent allele calls. For the nSSR dataset, we used Genepop 4.2 (Raymond and Rousset 1995, Rousset 2008) to perform an exact test of Hardy-Weinberg Equilibrium (HWE) with 1000 batches of 1000 Markov Chain Monte Carlo iterations (Gomes et al. 1999). We also checked for loci out of HWE in more than 60% of the populations; however, there were none.

nSSR genetic diversity

We calculated the total number of alleles per sampling location, private alleles, observed and expected heterozygosity in GenAlEx 6.502 (Peakall and Smouse 2006, 2012), and estimated the inbreeding coefficient (F_{IS}) in Genepop 4.2 (Raymond and Rousset 1995, Rousset 2008). We used the package diverSity in the R statistical program to calculate the allelic richness at each sampling location (Keenan et al. 2013).

nSSR genetic structure

To test for genetic differentiation between all pairs of sampling locations, we calculated Weir and Cockerham's (1984) pairwise F_{ST} values for 9999 permutations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012). In the R statistical program, we corrected the p-values using a false discovery rate (FDR) correction (Benjamini and Hochberg 1995). To test how much of the

genetic variance can be explained by within and between population variation, we ran an analysis of molecular variance (AMOVA) for 9999 permutations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012).

To examine the number of genetic clusters among our sampling locations, we used the Bayesian clustering program STRUCTURE v2.3.2 (Pritchard et al. 2000). Individuals were clustered assuming the admixture model both with and without a priori sampling location for a burnin length of 100,000 before 1,000,000 repetitions of MCMC for 10 iterations at each value of *K* (1 – 16). The default settings were used for all other parameters. We identified the most likely value of *K* using the Ln Pr(X|K) from the STRUCTURE output and the ΔK method from Evanno et al. (2005) in CLUMPAK (Kopelman et al. 2015). This initial STRUCTURE analysis of all 12 populations identified strong genetic structure patterns, so to test for more subtle population structure that may be present, we ran a separate STRUCTURE analysis within each of the previously identified clusters. These next two analyses followed the same parameters as the first STRUCTURE run, assuming the admixture model without a priori sampling location for a burnin length of 100,000 before 1,000,000 repetitions of MCMC for 10 iterations at each value of *K* (1–4 and 1–11, respectively). The default settings were used for all other parameters. We identified the most likely value of *K* using the Ln Pr(X|K) from the STRUCTURE output and the ΔK method from Evanno et al. (2005) in CLUMPAK (Kopelman et al. 2015).

To further explore the genetic structure of these populations, we ran a Principal Coordinates Analysis (PCoA) in GenAlEx 6.502, where the analysis was based on an individual pairwise genotypic distance matrix (Peakall et al. 1995, Smouse and Peakall 1999). To find and describe finer genetic structuring of the nSSR dataset, we performed a discriminant analysis of principal components (DAPC) in the R package *adegenet*, which optimizes among-group

variance and minimizes within-group variance (Jombart 2008, Jombart et al. 2010). To identify the number of clusters for the analysis, a Bayesian clustering algorithm was run for values of *K* clusters (1 – 16). We retained a *K*-value of 3 to explore any substructuring of the nSSR data. DAPC can be beneficial, as it can limit the number of principal components (PCs) used in the analysis. It has been shown that retaining too many PCs can lead to over-fitting and instability in the membership probabilities returned by the method (Jombart et al. 2010). Therefore, we performed the cross-validation function to identify the optimal number of PCs to retain.

To assess the effect of isolation by distance (IBD), we used a paired Mantel test based on a distance matrix of Slatkin's transformed F_{ST} ($D = F_{ST}/(1 - F_{ST})$) (Slatkin 1995) and a geographic distance matrix for 9999 permutations in GenAlEx 6.502, and the analysis follows Smouse et al. (1986) and Smouse and Long (1992). The mean geographic center was generated for each sampling location in ArcGIS software (ESRI™ 10.4.1), and the latitude and longitude of these points was then used to construct a matrix of straight line distances in km between each sampling location. The reported p-values are based on a one-sided alternative hypothesis (H_1: $R > 0$). A Mantel test was run for all sampling locations together, and a test was also run separately for populations within each cluster identified in the STRUCTURE analysis.

cpSSR genetic diversity

For the cpSSR dataset, we used the program HAPLOTYPE ANALYSIS v1.05 (Eliades and Eliades 2009) to calculate the number of haplotypes, haplotype richness, private haplotypes, haploid diversity. To visualize patterns in the cpSSR dataset, we created a minimum spanning network in the R package poppr (Kamvar et al. 2014). Nei's genetic distance was used as the basis for the network with a random seed of 9,999.

nSSR and cpSSR genetic structure

In order to compare the population structure of the nSSR and cpSSR data, we used the Φ_{ST} distance matrix for both datasets and ran an AMOVA. The population pairwise Φ_{ST} matrix facilitates comparison of molecular variance between codominant and dominant data by suppressing within individual variation, thus allowing for the comparison between varying mutation rates (Weir and Cockerham 1984, Excoffier et al. 1992). To test how much genetic variation could be explained by within populations, between populations, and between regions (genetic clusters identified through STRUCTURE analysis) for both datasets, we ran an AMOVA for 9, 999 iterations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012).

RESULTS

Microsatellite genotyping and genetic diversity

We genotyped 313 individuals from 12 locations at 14 nSSR loci (Table 2.2). No loci showed evidence for null alleles across all populations, there were no loci with more than 4 populations significantly out of HWE (less than 30% of populations) (Table 2.3), and no loci significantly deviated from linkage equilibrium across all populations. The nSSR loci were moderately polymorphic, and the number of alleles per locus per population ranged from 1 – 11, with a total of 85 alleles across 14 loci. Allelic richness (A_R) ranged from 2.32 – 4.21 per population with a mean of 3.53, and GM, PS, and TC populations exhibited lower levels of A_R than the other populations. Of the 6 private nSSR alleles identified, 5 were at low frequencies – occurring in five or fewer individuals, but the private nSSR allele in the GM population occurred in over 60% of individuals. Overall, the observed heterozygosity (H_O) averaged over loci for each population

ranged from 0.25 – 0.56 with a mean of 0.46, and the 3 northernmost populations (GM, PS, TC) had lower diversity in general. Expected heterozygosity (H_E) ranged from 0.30 – 0.57 across populations, with a mean of 0.49. GM and AD populations deviated significantly from HWE ($P < 0.05$). GM had a higher inbreeding coefficient (Table 2.3), but this could be attributed to our limited area in which to sample.

Both cpSSR loci were polymorphic, with 3 alleles per locus for a total of 6 alleles, and the number of alleles per population ranged from 2 – 4 with an average of 2.50 (Table 2.3). All alleles together resulted in 5 haplotypes. There were between 1 – 3 haplotypes per population for a haplotype richness ranging from 0.00 – 2.00, with a mean of 0.41 per population. Haploid diversity ranged from 0.00 – 0.58 with a mean of 0.10 per population. One allele and haplotype 2 were both unique to the SB and ZP sampling locations, and another allele and haplotype 4 were both private to five individuals sampled in GM, which occurred in a separate sampling location from the rest of the individuals in GM.

Genetic structure

The nSSR data suggested that there is strong genetic structure among the populations and regions of baby's breath sampled along the dunes of western and northern Michigan (global $F_{ST} = 0.228$). Pairwise comparisons using the nSSR data among all 12 populations yielded significant F_{ST} values after a FDR correction (Benjamini and Hochberg 1995) (Table 2.4). However, all pairwise comparisons of populations within Sleeping Bear Dunes National Lakeshore (hereafter Sleeping Bear Dunes or SBDNL) (GHB, SBP, DC, DP, EB, PB, SB) and nearby ZP displayed relatively lower pairwise F_{ST} values (Table 2.4), suggesting that there is some gene flow among these populations. The AMOVA based on the nSSR data also found that a significant amount of

the genetic variation could be explained by differences between populations in the northern region (GM, PS, TC) and populations in the southern region (GHB, SBP, DC, DP, EB, PB, SB, ZP, AD) (F_{CT} = 0.144, P < 0.0001), as well as among populations within regions (F_{SC} = 0.097, P < 0.0001). However, the majority of the genetic variance was explained by among population differences (F_{ST} = 0.228, P < 0.0001).

The Bayesian clustering analysis from the program STRUCTURE (Pritchard et al. 2000) partitioned the population into two clusters (K = 2) (Figure 2.2), inferred from both Ln Pr(X|K) and Evanno's ΔK (Supplemental Figure C). This analysis was run without inferring any prior information on sampling location, and then again with sampling information as prior. No differences were observed between the two results (without priors shown in Figure 2.2). Cluster 1 is comprised of the northernmost populations (GM, PS, TC), and cluster 2 includes all other populations. However, five individuals in the GM population (cluster 1) were assigned to cluster 2 (assignment probability > 90%), and these individuals were located at a separate sampling location from the rest in GM. In addition, though there is little admixture overall, several individuals in the GM, TC, EB, and AD populations showed a higher proportion of admixture among the two clusters. We ran subsequent STRUCTURE analyses on each of the two original identified clusters to further explore population substructuring (Supplemental Figures A and B). Results of the analysis of cluster 2 (Supplemental Figure A) suggest that it can be further broken down into two additional population groups (K = 2), inferred from both Ln Pr(X|K) and Evanno's ΔK (Supplemental Figure D). This suggests that the main cluster 2 can be further separated, as the individuals in the AD population were assigned to a separate additional cluster (Supplemental Figure A). Results of the analysis of cluster 1 (Supplemental Figure B) suggest that it can be further broken into three additional population groups (K = 3), inferred from both

Ln Pr(X|*K*) and Evanno's Δ*K* (Supplemental Figure E). The GM population clustered separately from the rest, and the five individuals that were sampled within the GM site also clustered separately from the rest of the individuals, with individuals of the PS and TC populations clustering together (Supplemental Figure B).

The Principal Coordinates analysis (PCoA) based on an individual pairwise genotypic distance matrix highlighted population substructuring (Figure 2.3). Individuals in the AD population expanded along both principal coordinates away from individuals assigned to the original STRUCTURE cluster 2 (Figure 2.2). In addition, the scatterplot supported the strong grouping of individuals in GM, PS, and TC together.

A Discriminant Analysis of Principal Components (DAPC) scatterplot (Figure 2.4a) grouped individuals into three clusters along two axes, supporting the substructuring illustrated in the PCoA. While the PCoA illustrated global diversity found in the nSSR dataset, the DAPC optimizes between group variance. DAPC can benefit from not using too many principal components (PCs), and we ran a cross-validation to identify how many PCs to retain. However, out of 69 total PCs, the cross-validation function suggested we retain 60 PCs (Jombart et al. 2010). We ran the DAPC using the recommended 60 PCs, but also checked if the general patterns remained with fewer PCs used by running the analysis with incrementally less PCs (45 and 30 PCs). All general patterns of the data in the scatterplots remained consistent despite the decreased PCs; therefore, we chose to use the scatterplot based on 30 PCs, as the benefit of the DAPC for our purposes is to show that the main patterns remain, despite minimization of within population variation (Jombart et al. 2010). Figure 2.4b shows the overlap between the distributions of individuals in DAPC clusters 2 and 3 along the first discriminant function, suggesting little distance between them. The membership of individuals of each population to the

three illustrated clusters can be seen in Figure 2.4c. This visualization of the data further highlights the more subtle structure of baby's breath populations in the dunes system of Michigan.

A Mantel test for isolation by distance (IBD) performed over all populations found a significant correlation between genetic and geographic distances ($R = 0.755$, $P < 0.001$) (Figure 2.5a). Upon further exploration of this correlation through separate Mantel tests within each identified STRUCTURE cluster, we found a significant correlation within cluster 2 (Figure 2.5c) ($R = 0.523$, $P = 0.030$), but no significant correlation within cluster 1 (Figure 2.5b) ($R = 0.205$, $P = 0.494$).

The AMOVA based on $\mathbf{\Phi}_{ST}$ distance (Table 2.5) facilitated the comparison between the nSSR and cpSSR data, which resulted in a significant amount of the genetic variation explained by differences among regions ($\mathbf{\Phi}_{CT}$), among populations within regions ($\mathbf{\Phi}_{SC}$), and within populations ($\mathbf{\Phi}_{ST}$) for both data sets ($P < 0.0001$). Both datasets also showed that most of the variation was explained by within population differences (nSSR $\mathbf{\Phi}_{ST} = 0.355$, cpSSR $\mathbf{\Phi}_{ST} = 0.736$, $P < 0.0001$).

For the cpSSR markers, the minimum spanning network illustrates the distribution of haplotypes across the 12 populations (Figure 2.6). Five haplotypes were found; Haplotype 1 was the most common, but only occurred in the SBDNL and ZP populations (GHB, SBP, DC, DP, EB, PB, SB, ZP). Haplotype 2 was private to the SB and ZP populations, but rare, occurring in one and two individuals, respective to the populations. Haplotype 3 was private to the five GM individuals located separately. Haplotype 4 was private to SB, ZP, and AD populations, and occurred in all AD individuals, but was less common in the SB and ZP populations. Haplotype 5 was private to GM, PS, and TC populations, occurring in all individuals. Due to the limited

number of polymorphic cpSSRs found for this study, caution should be used when interpreting these results. Genetic distance of the cpSSR dataset was estimated primarily to use as a complement to the nSSR analysis, to better understand gene flow among populations.

DISCUSSION

The natural disturbance regime of dynamic sand dune systems can result in a pattern of fragmented habitat and often sparse vegetative cover, making dune ecosystems highly susceptible to plant invasions (Jorgensen and Kollman 2009, Carranza et al. 2010, Rand et al. 2015). The topography, geographic distribution of preferred habitat, and disturbance regime in an ecosystem can influence various stages of a species invasion, including where the plant establishes, its dispersal patterns, and how densely it grows (With 2001, Theoharides and Dukes 2007). In addition, the demographic processes of an introduction event can shape contemporary population dynamics (Dlugosch and Parker 2008, Estoup and Guillemaud 2010). The invasion of baby's breath in the Michigan dune system is an opportunity to better understand the genetic structure of an invasive species, and how the dynamic landscape of a dune system may be shaping it. Our results indicate variation in genetic diversity among populations, as well as strong genetic structure that clusters individuals into two distinct groups. These two groups are separated by a peninsula that could be limiting gene flow between the two groups, causing this genetic separation.

We observed moderate levels of nuclear and chloroplast genetic diversity across populations of baby's breath throughout the dune system of Michigan (Table 2.5). However, genetic diversity in our northern-most populations (Grand Marais, Petoskey State Park, and Traverse City) was typically lower compared to that found in the populations in Sleeping Bear

Dunes, Zetterberg Preserve, and Arcadia Dunes. Differences in the level of genetic diversity among these regions could be due to differences in population size. Sleeping Bear Dunes is a largescale infestation and has some of the highest densities of baby's breath found within the MI coastal dunes (TNC 2013), consisting of up to 80% of the vegetation and covering hundreds of acres in some areas. The Grand Marais, Petoskey State Park, and Traverse City populations are much smaller than those found in Sleeping Bear Dunes, with continuous populations often limited to less than 45 acres (TNC 2012 internal report). These smaller populations could be more affected by the impact of genetic drift and potential inbreeding, resulting in the observed lower levels of genetic diversity (Ellstrand and Elam 1993, Young et al. 1996, Keller and Waller 2002).

The level of isolation between Grand Marais, Petoskey State Park, and Traverse City could also be contributing to the lower levels of genetic diversity observed in these areas compared to other populations. F_{ST} values among these three populations ranged from 0.121 – 0.221, which is much higher than the F_{ST} range observed between the Sleeping Bear Dunes, Zetterberg Preserve, and Arcadia Dunes populations (0.041 – 0.137). This suggests that our northern-most populations may have less gene flow between neighboring populations. This could be the result of larger geographic distances between these locations. For example, Grand Marais is located in Michigan's upper peninsula while Petoskey State Park and Traverse City are located in the lower peninsula. Higher levels of isolation could also be a result of decreased availability of suitable habitat, which may be more limited between these areas. Sleeping Bear Dunes and nearby surrounding areas make up a large contiguous amount of land that has been preserved by the National Parks Service, The Nature Conservancy and other local land conservancies. Thus, the dune habitat is often continuous, with limited human development. On the other hand,

Traverse City, Petoskey State Park and Grand Marais areas have more human development along the lakeshore, which may provide additional barriers to gene flow among these populations.

Management histories could also be contributing to the differences in genetic variation seen among these populations of baby's breath. The entire Petoskey State Park population was treated with herbicide or manual removal from 2007 – 2012 by The Nature Conservancy. At this time, managers considered the population to be at a desirable management level, and it has been unmanaged since (TNC 2012 internal report). It is possible that the intensive management resulted in a population bottleneck, and the population rebound following 2012 came from a reduced number of individuals, leading to the reduced genetic variation that we observe today. However, this is probably not the only reason for the lower levels observed. The Arcadia Dunes and Zetterberg Preserve populations have also been regularly managed since 2004 and 2007, respectively, so if management is solely driving these patterns we would expect Arcadia Dunes and Zetterberg Preserve to also have reduced genetic diversity. Although the Arcadia Dunes population does have the lowest allelic richness and heterozygosity of all the populations in cluster 2 (Figure 2.2), both populations have relatively high genetic variation despite over ten years of management. It is possible that higher levels of gene flow between these populations and those in Sleeping Bear Dunes may be helping to maintain genetic diversity. F_{ST} values between Zetterberg Preserve and other populations in Sleeping Bear Dunes range from 0.017 – 0.090, suggesting some gene flow, particularly with the population at the southern boundary (SB) of Sleeping Bear Dunes. Furthermore, infestations on private properties adjacent to Zetterberg Preserve have presumably buffered the population sizes. Given Petoskey State Park's geographic distance from Sleeping Bear Dunes, limited gene flow between them would prevent the maintenance of high genetic diversity after intense management.

The topography and habitat heterogeneity of the dune system likely contributes to the pattern of population structure of baby's breath throughout the Michigan dunes. Habitat heterogeneity can drive population structure, with variation in habitat type within the dunes acting as barriers to dispersal (Henry et al. 2009, Fant et al. 2014). Baby's breath is typically found in open back dune habitat, but has also been found in the fore dunes close to the lake beach and on steep dune sides. However, forested areas that are part of the back dunes have been identified by land managers as barriers between populations, preventing population spread of baby's breath (personal communication, Shaun Howard and Jon Throop). This can lead to populations in relatively close proximity to one another showing high levels of genetic differentiation. For example, the Empire Bluff population is located on the tip of a dune bluff: a small visitor outlook point surrounded by forest, and seems to be isolated from nearby populations. Despite its geographic proximity to Platte Bay (8.22 km), it is more genetically similar to Sleeping Bear Point, a population 12.73 km away (F_{ST} = 0.106 and F_{ST} = 0.069 respectively).

A Mantel test for isolation by distance (IBD) revealed a moderate positive relationship between nSSR genetic distances (based on transformed pairwise F_{ST} values) and geographic distances (straight-line distances in km) of all populations (Figure 2.5a), and was also found for the analysis of populations in STRUCTURE cluster 2 (Figure 2.5c). However, when examining the IBD relationship within cluster 1 from the STRUCTURE analysis, this positive relationship is not significant (Figure 2.5b). If IBD best described the distribution of genetic variation along the coast of Michigan, we would expect to see a gradient of genetic similarity following the coastal dunes along a North-South axis within each cluster. We attribute the overall significant relationships found (Figure 2.5a and 2.5c) to the strong genetic differences between populations

in the two main clusters. Additionally, the IBD found in cluster 2 is likely driven by the genetic difference of the Arcadia Dunes population in comparison to the Zetterberg Preserve and Sleeping Bear Dunes populations. Though geographic distance could be influencing more subtle structuring of these populations, the isolating effect of the topography within the dunes could have an effect that overrides that of geographic distance. These results further support the strong regional differences between the two clusters identified in the Bayesian analysis (Figure 2.2). Furthermore, the admixture patterns among the two clusters, specifically related to the Grand Marais, Traverse City, Empire Bluff, and Arcadia Dunes populations, could be contributing to the lack of IBD found when all populations were included.

The tumbleweed mechanism of dispersal that baby's breath employs could be an effective means to disperse seeds, but it is possible that the strong topographical structure, habitat heterogeneity and variable weather patterns within the dunes impact seed dispersal for gene flow more so than pollination. Baby's breath has been found to attract a diverse array of pollinator species (Baskett et al. 2011, Emery and Doran 2013), sometimes at the expense of native plant pollination, while seed dispersal is primarily limited to wind driven tumbleweeds that form once the upper foliage senesces. The variation in $\mathbf{\Phi}_{ST}$ values between the two marker types (nSSR $\mathbf{\Phi}_{ST}$ = 0.355, cpSSR $\mathbf{\Phi}_{ST}$ = 0.736) indicates that barriers to seed dispersal may be more limiting for gene flow than pollination. Darwent and Coupland (1966) also suggested that though seeds could be dispersed up to 1 km, many of the seeds were released from the fruit capsules near the parent plant prior to the stems tumbling. This could result in strong population structure due to a lower frequency of migrants. Therefore, the elements of the dune ecosystem that make it so dynamic could be impacting gene flow through seed dispersal by further limiting the plant's ability to spread throughout the landscape. However, the comparison of cpSSR to nSSR results should be

taken with some caution, as we had a limited number of polymorphic cpSSR markers. Though we chose to use microsatellites within the chloroplast genome to increase the likelihood of polymorphism, we still found these regions to be well-conserved and with limited variation in our dataset. Therefore, we cannot rule out the possibility of fragment size homoplasy confounding results of low genetic diversity in some populations (Bang and Chung 2015).

Structure analysis

Our results from the Bayesian clustering analysis in STRUCTURE (Pritchard et al. 2000, Evanno et al. 2005) separate the populations of baby's breath along the Michigan coastal dunes separate into two genetic clusters ($K = 2$) (Figure 2.2). A similar pattern was found when the nSSR dataset was analyzed using a PCoA (Figure 2.3) and a DAPC (Figure 2.4), with the exception that the individuals of the Arcadia Dunes population further separated from the Sleeping Bear Dunes and Zetterberg Preserve individuals in these analyses. The clusters are mainly divided into the Traverse City, Petoskey State Park, and Grand Marais cluster (cluster 1) and the Sleeping Bear Dunes populations, Zetterberg Preserve, and Arcadia Dunes cluster (cluster 2). The distribution of cpSSR haplotypes (Figure 2.6) across populations further illustrates the strong genetic clusters present in this dataset. Specifically, some haplotypes are only found in certain populations and in within each main population cluster. Haplotypes 1, 2, and 4 only occur in populations in cluster 2. Haplotypes 3 and 5 only occur in cluster 1, with haplotype 3 being private to the five individuals in Grand Marais that were located separately from the rest of those sampled at this location (Figure 2.6). The two distinct population clusters are separated by the Leelanau peninsula, which may be helping to limit gene flow among these clusters. This partitioning of cpSSR haplotypes could be due to seed dispersal limitations from

habitat fragmentation, unsuitable habitat, and land use, as the peninsula is comprised mainly of private residential properties along the narrow shoreline.

Understanding the invasion history of a species can help shed light on the factors and processes that contributed to the success of the species establishment. For baby's breath, it has been assumed that invasive populations were the result of ornamental plants escaping from gardens or being purposely planted for horticultural means (personal communication with TNC managers). Whether the clusters we observed for our dataset are the result of independent introductions or the result of one introduction followed by serial invasions is not known. Given that populations along coastal Michigan cluster into two distinct groups, either scenario is possible (Lombaert et al. 2018). In the serial invasion scenario, a founder population would have colonized one site in the Michigan coastal dunes, and then migrants from that population would have invaded subsequent areas (Lombaert et al. 2018). Over time, with limited gene flow, these populations could have become distinct and structured. However, we think this scenario may not be the best explanation for this invasion. Based upon herbarium records, the first occurrence of baby's breath in northwest Michigan was recorded in 1913 in Emmet County where Petoskey State Park is located (Emmet Co., 1913, *Gleason s.n.*, MICH). Records from Leelanau and Benzie counties, where Sleeping Bear Dunes, Zetterberg Preserve and Arcadia Dunes are located, were not collected until the late 1940's (Leelanau Co., 1947, *P.W. Thompson L-302*, MICH). If Petoskey State Park was the founding population for this invasion, we would expect higher genetic diversity in this population relative to those in Sleeping Bear Dunes, Zetterberg Preserve, and Arcadia Dunes, since a serial introduction would result in additional bottlenecks from the founding population. However, we observed the opposite pattern of genetic diversity.

Additionally, there are private cpSSR haplotypes to each of these clusters (Figure 2.6), a pattern we would not expect to see if all the populations came from one introduction event.

The other invasion scenario describes at least two independent introductions to the Michigan coastal dunes (Lombaert et al. 2018). In this scenario, we would expect strong genetic differentiation between the two or more founding populations. Our data supports this, as we observed both nSSR and cpSSR alleles privately shared only between populations within the same cluster. In addition, for the cpSSR markers, distinct haplotypes were found between the two regions, with haplotype 5 only observed in the Grand Marais, Petoskey State Park, and Traverse City cluster while haplotypes 1, 2, and 4 were only found in the Sleeping Bear Dunes, Zetterberg Preserve, and Arcadia Dunes cluster. There was also a high proportion of nSSR alleles common to both clusters, but this could be the result of limited genetic diversity in the initial source populations (Allendorf and Lundquist 2003). This scenario is particularly plausible, as the source populations would likely be a type of horticultural strain, given the popularity of perennial baby's breath in the floral industry (Vettori et al. 2013, Calistri et al. 2014). This hypothesis of at least two independent introductions also agrees with the herbarium record: a potential introduction event could have occurred in the early 1910's, leading to cluster 1 (GM, PS, TC), and a separate introduction event could have occurred in the late 1940's, leading to the establishment of the populations in Zetterberg Preserve and Sleeping Bear Dunes (cluster 2).

In addition to supporting the identified patterns in the nSSR dataset produced from the STRUCTURE analysis, the PCoA and DAPC (Figures 2.3 and 2.4) allowed us to identify more subtle population structuring. Specifically, the PCoA (Figure 2.3) illustrates the Arcadia Dunes population separating from the other populations along principal coordinate 2. The DAPC (Figure 2.4) also shows the subtler variation among populations within the Sleeping Bear Dunes

populations (specifically Figure 2.4c), and continues to support the segregation of the Grand Marais, Petoskey State Park, and Traverse City populations from the rest that we see in the STRUCTURE analysis (Figure 2.2). Variation in allele frequencies and decreased allelic richness are two factors that could explain the divergence of the Arcadia Dunes population in the PCoA (Figure 2.3). There are no private alleles or other obvious patterns causing this population to cluster separately from nearby populations (Zetterberg Preserve and South Boundary in Sleeping Bear Dunes). The higher rates of admixture between the two main clusters in Arcadia Dunes individuals (Figure 2.2) could also be a reason for its slight divergence from cluster 2. However, what is driving this potential higher level of admixture in the Arcadia Dunes population compared to others is currently unknown. Arcadia Dunes is a popular recreation area among locals and tourists (personal communication Jon Throop, Grand Traverse Regional Land Conservancy). Additionally, the autumn season brings about a high volume of foot traffic through all the dune areas of Michigan. It is possible that people may be accidentally transporting baby's breath seeds between these otherwise isolated populations, as the seed phenology coincides with the autumn senescence. While human transport of seeds may be occurring at other locations as well, Arcadia Dunes is a small enough population that newly introduced genotypes could have a higher likelihood of being detected from sampling relative to other larger populations, such as one in Sleeping Bear Dunes.

The invasion of baby's breath to the Great Lakes has the potential to disrupt the dynamism of the dune landscape and biological community in the northwest Michigan, and this threat has led to increased concern over its pervasiveness regionally and nationally. Estimating the genetic structure of invasive populations can lead to a better understanding of the invasion history and the factors influencing the success of an invasion (Crosby et al. 2014, Piya et al.

2014, Sakata et al. 2015). Through population level analysis, we found strong genetic structure present that separates the invasion in the Michigan dunes into two main regions. Based on these results, we suggest that the contemporary baby's breath population within the Michigan coastal dune system is the result of at least two separate introduction events. The genetic structure identified for these baby's breath populations probably results from a combination of demographic processes –multiple introductions, bottleneck events, isolation, and admixture, along with landscape level processes. The topography of the dunes is heterogeneous but also constantly shifting, and the baby's breath invasion is one example of how this dynamic system can shape the establishment, gene flow, and spread of invasive plant populations.

ACKNOWLEDGEMENTS

The authors thank the Environmental Protection Agency – Great Lakes Restoration Initiative Grant (C.G.P., Grant #00E01934) for financial support. We thank Dr. James McNair for assistance with statistical analyses, Dr. Timothy Evans for assistance with herbarium data expertise and submission, and Kurt Thompson for assistance with geographic data collection and visualization. Finally, we thank Emma Rice for all her assistance in field data collection, Syndell Parks for her assistance in laboratory data collection, and Matthew Kienitz for his assistance in field and laboratory data collection.

REFERENCES

Albert DA (2000). *Borne of the Wind, Michigan Sand Dunes*. University of Michigan Press: Michigan.

Allendorf FW, Lundquist LL (2003). Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* **17**: 24 – 30.

Arbogast AF (2015). *The Emerging Science of Coastal Sand Dune Age and Dynamics: Implications for Regulation and Risk Management in Michigan*.

Arbogast AF, Loope WL (1999). Maximum-limiting ages of Lake Michigan coastal dunes: Their correlation with Holocene lake level history. *Journal of Great Lakes Research* **25**: 372-382.

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

Baker HG (1986). Patterns of plant invasion in North America. In: Mooney HA, Drake JA (eds) *Ecology of Biological Invasions of North America and Hawaii*. Springer, New York, pp 44–57.

Baskett CA, Emery SM, Rudgers JA (2011). Pollinator Visits to Threatened Species Are Restored Following Invasive Plant Removal. *International Journal of Plant Sciences* **172**: 411–422.

Benjamini Y, Hochberg Y (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal of the Royal Statistical Society* Series B. **57**: 289 - 300.

Blumer BE, Arbogast AF, Forman SL (2012). The OSL chronology of eolian sand deposition in a perched dune field along the northwestern shore of Lower Michigan. *Quaternary Research* **77**: 445 – 455.

Bustamante RO, Durán AP, Peña-Gómez FT, Véliz D (2018). Genetic and phenotypic variation, dispersal limitation and reproductive success in the invasive herb *Eschscholzia californica* along an elevation gradient in central Chile. *Plant Ecology & Diversity* **10**: 419 – 429.

Calistri E, Buiatti M, Bogani P (2014). Characterization of *Gypsophila* species and commercial hybrids with nuclear whole-genome and cytoplasmic molecular markers. *Plant Biosystems* **150**: 11 – 21.

Carranza ML, Carboni M, Feola S, Acosta ATR (2010). Landscape-scale patterns of alien species on coastal dunes: the case of iceplant in central Italy. *Applied Vegetation Science* **13**: 135-145.

Chung S-M, and Staub JE (2003). The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theoretical and Applied Genetics*. 107: 757–767.

Cowles HC (1899). The Ecological Relations of the Vegetation on the Sand Dunes of Lake Michigan. Part I.-Geographical Relations of the Dune Floras. *Botanical Gazette*. **27**: 95 – 117.

Crosby K, Stokes TO, Latta RG (2014). Evolving California genotypes of *Avena barbata* are derived from multiple introductions but still maintain substantial population structure. *PeerJ* **2**: e633.

Cusseddu V, Ceccherelli G, Bertness M (2016). Hierarchical organization of a Sardinian sand dune plant community. *PeerJ* **4**: e2199.

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

Darwent AL, Coupland RT (1966). Life history of *Gypsophila paniculata*. *Weeds* **14**: 313–318.
Dewoody J, Nason JD, Hipkins VD (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* **6**: 951–957.

DiTomaso J, Kyser GB (2013). Weed Report: Gypsophila paniculata L. In *Weed Control in Natural Areas in the Western United States*. UC Davis Weed Research and Information Center.

Dlugosch KM, Parker IM (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**: 431 – 449.

Department of Natural Resources (DNR) (2015). Michigan Invasive Species Grant Program Handbook. Website http://www.michigan.gov/documents/dnr/2015-MISGP-hndbk_491809_7.pdf?source=govdelivery [accessed 05 March 2018].

Earl DA, vonHoldt BM (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Eliades N-G, Eliades DG (2009). HAPLOTYPE ANALYSIS: Software for analysis of haplotype data. (available from http://www.uni-goettingen.de/en/134935.html)

Ellstrand NC, Elam DR (1993). Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annual Review of Ecology and Systematics* **24**: 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.

Environmental Systems Research Institute (ESRI). (2012). ArcGIS Desktop: Release 10.4.1. Redlands, CA.

Estoup A, Guillemaud T (2010). Reconstructing routes of invasion using genetic data: why, how and so what. *Molecular Ecology* **19**: 4113 – 4130.

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

Everard M, Jones L, Watts B (2010). Have we neglected the societal importance of sand dunes? An ecosystem services perspective. *Aquatic Conservation: Marine and Freshwater Ecosystems* **20:** 476–487.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites. *Genetics* **131**: 479 – 491.

Fant JB, Havens K, Keller JM, Radosavljevic A, Yates ED (2014). The influence of contemporary and historic landscape features on the genetic structure of the sand dune endemic, *Cirsium pitcheri* (Asteraceae). *Heredity* **112**: 519 – 530.

Gomes I, Collins A, Lonjou C *et al.* (1999). Hardy–Weinberg quality control. *Annals of Human Genetics* **63**: 535–538.

Gleason HA, Cronquist A (1963). Manual of vascular plants of northeastern United States and adjacent Canada. D. Van Nostrand Company, New York, New York.

Grime, J. P. 1979. *Plant Strategies and Vegetation Processes*. John Wiley and Sons: Chichester, UK.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591–611.

Hagenblad J, Hülskötter J, Acharya KP, Brunet J, Chabrerie O, Cousins SA, Dar PA, Diekman M, De Frenne P, Hermy M, Jamoneau A, Kolb A, Lemke I, Plue J, Reshi ZA, Graae BJ (2015). Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. *BMC Genetics* **16**:103.

Henry P, Le Lay G, Goudet J, Guisan A, Jahodová S, Besnard G (2009). Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Molecular Ecology* **18**: 2819–2831.

Jolls CL, Marik JE, Hamzé SI, Havens K (2015). Population viability analysis and the effects of light availability and litter on populations of *Cirsium pitcheri*, a rare, monocarpic perennial of Great Lakes shorelines. *Biological Conservation* **187**: 82–90.

Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403 – 1405.

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.

Jorgensen RH, Kollman J (2009). Invasion of coastal dunes by the alien shrub *Rosa rugosa* is associated with roads, tracks and houses. *Flora – Morphology, Distribution, Functional Ecology of Plants* **204**: 289-297.

Kamvar ZN, Tabima JF, Grünwald NJ (2014). Poppr: An R Package For Genetic Analysis Of Populations With Clonal, Partially Clonal, And/or Sexual Reproduction. *PeerJ* **2**: e281.

Karamanski TJ (2000). A nationalized lakeshore: the creation and administration of Sleeping Bear Dunes National Lakeshore. *National Park Service, Department of the Interior*. Washington, DC.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA, O'Hara RB (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**: 782 – 788.

Keller LF, Waller DM (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution* **17**: 230 – 241.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**: 1179 – 1191.

Leimbach-Maus H, Parks S, Partridge C. Microsatellite primer development for invasive perennial herb, *Gypsophila paniculata* (Caryophyllaceae). Manuscript in preparation.

Lippincott Seeds. Gypsophila paniculata perennial. Minneapolis: Miss C.H. Lippincott, 1915. Henry G. Gilbert Nursery and Seed Trade Catalog Collection. Special Collections, National Agricultural Library.

Lombaert E, Guillemaud T, Deleury E (2018). Biases of STRUCTURE software when exploring introduction routes of invasive species. *Heredity* **120**: 485 – 499.

Mack RN, Lonsdale WM (2001). Humans as global plant dispersers: getting more than we bargained for. *BioScience* **51**: 95 – 102.

Michigan, Emmet Co.: Pellston, 16 July 1913, *H.A. Gleason s.n.* (MICH).

Michigan, Leelanau Co.: Leland Twp, 05 August 1947, *P.W. Thompson L-302* (MICH).

Moles AT, Gruber MA, Bonser SP (2008). A new framework for predicting invasive plant species. *Journal of Ecology* **96***:* 13–17.

Moran EV, Reid A, Levine JM (2017). Population genetics and adaptation to climate along elevation gradients in invasive *Solidago canadensis*. *PLoS ONE* **12**: e0185539.

Nagy AM, Korpelainen H (2014). Population genetics of Himalayan balsam (*Impatiens glandulifera*): comparison of native and introduced populations. *Plant Ecology & Diversity* **8**: 317 – 321.

Olson JS (1958). Rates of Succession and Soil Changes on Southern Lake Michigan Sand Dunes. *Botanical Gazette* **119**: 125 – 170.

Peakall R, Smouse PE, Huff DR (1995). Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides. Molecular Ecology* **4**: 135-147.

Smouse PE, Peakall R (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**: 561 – 573.

Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 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**: 2537 – 2539.

Pickering C, Mount A (2010). Do tourists disperse weed seed? A global review of unintentional human-mediated terrestrial seed dispersal on clothing, vehicles and horses. *Journal of Sustainable Tourism* **18**: 239 – 256.

Piya S, Nepal MP, Butler JL, Larson GE, Neupane A (2014). Genetic diversity and population structure of sickleweed (*Falcaria vulgaris*; Apiaceae) in the upper Midwest USA. *Biological Invasions* **16**: 2115–2125.

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

Raymond M, Rousset F (1995). GENEPOP (version 4.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.

Rand TA, Louda SM, Bradley KM, Crider KK (2015). Effects of invasive knapweed (*Centaurea stoebe* subsp. micranthos) on a threatened native thistle (*Cirsium pitcheri*) vary with environment and life stage. *Botany* **93**: 543–558.

Rowland J, Maun MA (2001). Restoration ecology of an endangered plant species: Establishment of new populations of *Cirsium pitcheri*. *Restoration Ecology* **9**: 60–70.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219 – 1228.

Rousset F (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**: 103 – 106.

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA et al (2001). The population biology of invasive species. *Annual Review Ecology and Systematics* **32**: 305–332.

Sakata Y, Itami J, Isagi Y, Ohgushi T (2015). Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range. *Journal of Plant Research* **128:** 909–921.

Sixth Annual Seed Catalogue, Floral and Western Cultivator's Guide. "Gypsophila." Chicago: Hovey and Nichols, 1868. Henry G. Gilbert Nursery and Seed Trade Catalog Collection. Special Collections, National Agricultural Library.

Stynes D (2011). *Economic Benefits to Local Communities from National Park Visitation and Payroll, 2010*. *NPS Natural Resource Report*. East Lansing, MI.

The Nature Conservancy and partners. (2013). Lake Michigan Coastal Dunes Restoration Report. The Nature Conservancy. Website https://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/michigan/lake-michigan-dune-report-2013.pdf [accessed 01 November 2017].

Theoharides KA, Dukes JS (2007). Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytologist* **176:** 256 – 273.

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

Van Oosterhout C, Weetman D, Hutchinson WF (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Resources* **6**: 255–256.

Vettori L, Schiff S, Tani C, Pasqualetto P, Bennici A (2013). Morphological and cytological observations of wild species and hybrids of *Gypsophila*. *Plant Biosystems* **150**: 1 – 11.

Weir B, Cockerham C (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**: 1358 – 1370.

With KA (2002). The landscape ecology of invasive spread. *Conservation Biology* **16**: 1192 – 1203.

Xu C-Y, Tang S, Fatemi M, Gross CL, Julien MH, Curtis C, van Klinken RD (2015). Population structure and genetic diversity of invasive *Phyla canescens*: implications for the evolutionary potential. *Ecosphere* **6**: 162.

Young A, Boyle T, Brown T (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**: 413 – 418.

Zarnetske PL, Seabloom EW, Hacker SD (2010). Non-target effects of invasive species management: beachgrass, birds, and bulldozers in coastal dunes. *Ecosphere* **1**: art13.

TABLES

Table 2.1 Sampling location names and geographic coordinates for baby's breath analyzed in this study. All locations are in Michigan. Location abbreviations are used in the main text and following tables and figures.

Sampling Location (Code)	*n*	Latitude	Longitude
Grand Marais (GM)	35	46.67825579	-85.97546860
Petoskey State Park (PS)	30	45.40288418	-84.91271857
Traverse City (TC)	30	44.74865647	-85.61882032
Good Harbor Bay (GHB)	20	44.93877954	-85.86802898
Sleeping Bear Point (SBP)	25	44.91095892	-86.04209863
Dune Climb (DC)	23	44.88285396	-86.04280635
Dune Plateau (DP)	30	44.87312491	-86.05846389
Empire Bluff (EB)	20	44.80154168	-86.07121955
Platte Bay (PB)	20	44.73111860	-86.10566158
South Boundary (SB)	20	44.72858265	-86.15892124
Zetterberg Preserve (ZP)	30	44.68665052	-86.25030285
Arcadia Dunes (AD)	30	44.53662395	-86.22527264

Notes: *n* number of individuals sampled.

Table 2.2 Characteristics of 14 nSSR loci developed for baby's breath and 2 universal cpSSR loci used in this study.

Locus	Primer sequences (5' - 3')	Repeat motif	Allele size range (bp)	Annealing temperature (°C)	Fluorescent label	Multiplex
nSSR Loci						
BB_21680	F: ACTACACACAGACTCGATCCTC R: CTTTGATTGTTTGGTGTAAGTTGC	$(AAAG)_5$	199 - 218	62	PET	PS1
BB_6627	F: CAAACTCAACCAACCAGACACC R: CACCTCAGCAACAACAGAGTG	$(AAAC)_5$	151 - 155	62	FAM	PS1
BB_3968	F: CATGGAGGACAATGAGAAGACG R: ACGGTGGTAATGAAGTTTGGTG	$(AGG)_6$	207 - 219	62	FAM	PS2
BB_5151	F: TCCACCTTATAACTCACCACCC R: TGAGGAAGGATAACAGCTCTCG	$(ACC)_5$	205 - 210	62	PET	PS2
BB_4443	F: TAGGGTGGGTGCTTGTACTAAC R: AAAGTGGTGCTGCAGAAGAATC	$(AAG)_{16}$	171 - 211	62	NED	PS2
BB_31555	F: TGTATAACTGAGATAACCCAGACG R: TTGTTACCTTGTTCCGGCAAAG	$(AC)_7$	150 - 156	62	VIC	PS2
BB_14751	F: CCTCAAACCCTAACAATGCTCC R: TCAGCCGATCCTCTAACACG	$(AAG)_{12}$	195 - 248	62	FAM	PS3
BB_3335	F: TCCACCAAACTCTTAAACTGCC R: CACAGACACAAAGGATCCAACC	$(AGG)_5$	215 - 244	62	NED	PS3
BB_4258	F: TCACAAGAGGCCCAATTTCTTC R: ACTTGAACCCGAACCTATACCC	$(AAT)_5$	178 - 195	62	VIC	PS3
BB_3913	F: GGCTGTCGGGTAATAAACACAG R: TCCCAACTCAAGTCATAGCCTAG	$(ACAG)_5$	159 - 171	62	PET	PS3
BB_2888	F: CTTCATTCATGTACAAGAGCGC R: AGAACTGGCTATGGATCGAAATG	$(AC)_{16}$	219 - 232	63	FAM	PS4
BB_5567	F: GGCTAGGGAAAGTAGGAAGACC R: CGTGTCCTGTTTCTCCATGATC	$(AAT)_5$	198 - 222	62	VIC	PS4
BB_7213	F: TTGCATTCCCACCATTTCATCC R: AGCCAACCTCGTATTAATTGCC	$(AC)_7$	161 - 248	62	PET	PS4
BB_8681	F: ATCTCCAGTTTCCGTGATTTGC R: TACGTCACAAGAGCTTTCAACC	$(ACC)_8$	204 - 222	62	NED	PS5
cpSSR Loci						
ccssr4	F: AGGTTCAAATCCTATTGGACGCA R: TTTTGAAAGAAGCTATTCARGAAC	$(T)_8$	204 - 219	52	NED	–
ccssr9	F: GAGGATACACGACAGARGGARTTG R: CCTATTACAGAGATGGTGYGATTT	$(A)_{13}$	199 - 215	52	PET	–

Notes : Locus ccssr4 and ccssr9 were developed by Chung and Staub (2003) using chloroplast sequences from *Nicotiana tabacum* L. Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

Table 2.3 Genetic diversity indices for baby's breath from each sampling location at 14 nSSRs and 2 cpSSRs.

	Sampling Locations											
	GM	PS	TC	GHB	SBP	DC	DP	EB	PB	SB	ZP	AD
nSSR Loci												
BB_21680												
N	35	30	30	20	25	23	30	19	20	20	30	30
N_A	3	3	4	3	3	4	3	4	3	4	4	3
H_O	0.286	0.667	0.300	0.450	0.440	0.522	0.500	0.421	0.500	0.400	0.500	0.700
H_E	0.411	0.539	0.534	0.540	0.582	0.481	0.540	0.676	0.524	0.510	0.548	0.546
F_{IS}	**0.3186**	−0.2198	**0.4517**	0.1915	0.2636	−0.0624	0.0909	**0.4000**	0.0709	**0.2400**	0.1040	−0.2661
BB_6627												
N	35	30	30	20	24	23	30	20	20	20	30	30
N_A	2	1	2	2	2	2	2	2	2	2	2	2
H_O	0.086	0.000	0.167	0.500	0.333	0.435	0.500	0.250	0.450	0.550	0.300	0.467
H_E	0.082	0.000	0.153	0.480	0.330	0.499	0.495	0.219	0.489	0.439	0.255	0.464
F_{IS}	−0.0303	–	−0.0741	−0.0160	0.0108	0.1506	0.0068	−0.1176	0.1047	−0.2294	−0.1600	0.0122
BB_3968												
N	35	30	30	20	25	23	30	20	20	20	30	30
N_A	3	2	1	4	3	3	4	4	3	4	3	2
H_O	0.143	0.067	0.000	0.400	0.240	0.304	0.367	0.450	0.550	0.400	0.500	0.133
H_E	0.207	0.064	0.000	0.476	0.246	0.334	0.414	0.475	0.509	0.345	0.418	0.180
F_{IS}	**0.3241**	−0.0175	–	0.1850	0.0432	0.1098	0.1320	0.0782	−0.0556	−0.1343	−0.1805	0.2750
BB_5151												
N	35	29	30	19	25	23	30	19	20	20	30	28
N_A	2	2	2	2	2	2	2	2	2	2	2	2
H_O	0.057	0.034	0.133	0.158	0.200	0.391	0.467	0.526	0.400	0.400	0.200	0.179
H_E	0.056	0.034	0.231	0.494	0.449	0.466	0.491	0.465	0.480	0.375	0.180	0.499
F_{IS}	−0.0149	−0.018	0.4369	**0.6949**	**0.5683**	0.1818	0.0667	−0.1043	0.1915	−0.0411	−0.0943	**0.6530**
BB_4443												
N	35	30	30	20	25	23	30	19	20	19	30	30
N_A	3	5	3	9	9	9^A	9	5	11	7	10^A	5
H_O	0.257	0.800	0.400	0.800	0.640	0.783	0.767	0.842	0.900	0.526	0.667	0.567
H_E	0.338	0.701	0.399	0.808	0.769	0.778	0.758	0.749	0.853	0.593	0.651	0.663
F_{IS}	**0.2537**	−0.1244	0.0156	0.0349	0.1804	0.0161	0.0052	−0.0971	−0.0301	0.2193	−0.0078	0.1623
BB_31555												
N	28	30	30	20	25	23	30	20	20	20	30	30
N_A	1	2	2	3	4	4	4	4	4	4	4	3
H_O	0.000	0.233	0.333	0.650	0.800	0.478	0.600	0.650	0.750	0.500	0.600	0.467
H_E	0.000	0.255	0.278	0.609	0.727	0.650	0.614	0.654	0.745	0.583	0.663	0.545
F_{IS}	–	0.1018	−0.1837	−0.0422	−0.0799	0.2851	0.0396	0.0314	0.0189	0.1667	0.1122	0.1603
BB_14751												
N	34	30	30	20	25	23	30	20	20	20	28	30
N_A	5^A	3	4	7	6	7	8	9	9	10	6	7
H_O	0.676	0.200	0.467	0.650	0.600	0.478	0.633	0.800	0.600	0.650	0.750	0.467
H_E	0.666	0.287	0.548	0.714	0.633	0.618	0.769	0.790	0.621	0.646	0.675	0.762
F_{IS}	**−0.0013**	0.3176	−0.0899	0.1099	0.0722	0.2473	0.1933	0.0130	0.0579	0.0159	−0.0925	0.2632
BB_3335												
N	33	30	30	20	25	23	30	20	20	20	30	30
N_A	5	3	4	8	8	9	7	5	9	10	8	6
H_O	0.242	0.500	0.433	0.800	0.760	0.826	0.667	0.800	0.750	0.850	0.767	0.600
H_E	0.403	0.562	0.369	0.818	0.732	0.827	0.817	0.694	0.808	0.815	0.789	0.709
F_{IS}	**0.4115**	**0.1265**	−0.1564	0.0470	−0.0179	0.0234	**0.2000**	−0.1280	0.0967	−0.0173	0.0451	0.1701

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), A_R allelic richness for each population averaged across loci, *H* haploid diversity, N_H number of haplotypes for each population averaged across loci, H_R haplotype richness for each population averaged across loci. Bold values indicate loci that deviated from Hardy-Weinberg equilibrium. ^{A} denotes a private allele, ^{P} denotes a private haplotype. Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

Table 2.3 (Continued) Genetic diversity indices for baby's breath from each sampling location at 14 nSSRs and 2 cpSSRs.

	Sampling Locations											
	GM	PS	TC	GHB	SBP	DC	DP	EB	PB	SB	ZP	AD
nSSR Loci												
BB_4258												
N	35	30	30	20	25	23	30	20	20	20	30	30
N_A	1	1	2	3^A	2	2	2	1	2	4^A	3	2
H_O	0.000	0.000	0.133	0.150	0.240	0.087	0.033	0.000	0.150	0.250	0.400	0.300
H_E	0.000	0.000	0.124	0.141	0.269	0.083	0.033	0.000	0.139	0.228	0.326	0.339
F_{IS}	–	–	−0.0545	−0.0364	0.1273	−0.0233	−0.017	–	−0.0556	−0.0734	−0.2104	0.1329
BB_3913												
N	35	30	30	20	25	23	30	20	20	20	30	30
N_A	3	2	3	4	4	4	4	4	4	4	4	2
H_O	0.171	0.133	0.300	0.400	0.480	0.565	0.667	0.550	0.500	0.550	0.600	0.467
H_E	0.207	0.124	0.292	0.471	0.537	0.619	0.578	0.614	0.494	0.621	0.638	0.444
F_{IS}	0.1873	−0.0545	−0.0166	0.1762	0.1259	0.1090	−0.1373	**0.1292**	0.0130	0.1399	0.0769	−0.0331
BB_2888												
N	35	29	30	20	25	23	30	20	20	20	30	30
N_A	4	2	3	4	6	6	5	6	5	6^A	6	5
H_O	0.657	0.586	0.600	0.800	0.680	0.913	0.833	0.750	0.800	0.800	0.900	0.667
H_E	0.594	0.498	0.651	0.734	0.724	0.772	0.793	0.768	0.746	0.734	0.786	0.589
F_{IS}	−0.0922	−0.1610	0.0953	−0.0648	0.0811	−0.1608	−0.0335	0.0484	−0.0465	−0.0648	−0.1282	−0.1154
BB_5567												
N	35	30	30	20	25	23	30	20	20	20	30	30
N_A	4	3	4	5	3	3	4	3	4	3	5	5
H_O	0.629	0.533	0.567	0.700	0.480	0.609	0.667	0.400	0.550	0.400	0.600	0.767
H_E	0.716	0.613	0.562	0.745	0.614	0.474	0.604	0.374	0.589	0.371	0.613	0.716
F_{IS}	**0.1368**	0.1463	0.0080	0.0859	0.2371	−0.2649	**−0.0872**	−0.0447	0.0913	−0.0519	0.0387	−0.0545
BB_7213												
N	35	30	30	20	25	23	30	20	20	20	30	30
N_A	3	2	3	3	4	3	3	3	3	5	5	3
H_O	0.229	0.367	0.100	0.500	0.640	0.391	0.500	0.400	0.650	0.500	0.633	0.667
H_E	0.359	0.375	0.096	0.434	0.642	0.638	0.565	0.386	0.611	0.499	0.599	0.633
F_{IS}	**0.3754**	0.0392	−0.0235	−0.1276	0.0241	**0.4054**	0.1317	−0.0100	−0.0378	0.0231	−0.0406	−0.0366
BB_8681												
N	35	28	30	20	24	22	30	19	20	20	30	30
N_A	3	3	2	3	3	3	4	2	3	3	4	3
H_O	0.114	0.357	0.333	0.500	0.500	0.136	0.400	0.211	0.250	0.300	0.467	0.600
H_E	0.109	0.523	0.320	0.395	0.469	0.206	0.456	0.188	0.265	0.464	0.502	0.438
F_{IS}	−0.0342	**0.3333**	−0.0247	−0.2418	−0.0455	0.3571	0.1397	−0.0909	0.0821	**0.3753**	0.0866	−0.3558
A_R across loci	2.660	2.320	2.540	3.970	3.750	3.920	3.990	3.660	4.070	4.210	4.190	3.120
cpSSR Loci												
ccssr4												
N	35	30	29	20	25	23	30	20	20	20	30	30
N_A	2	1	1	1	1	1	1	1	1	2	2	1
H	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.50	0.00
ccssr9												
N	35	30	29	20	25	23	30	20	20	20	30	30
N_A	2^A	1	1	1	1	1	1	1	1	2	2	1
H	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.12	0.00
N_H across loci	2^P	1	1	1	1	1	1	1	1	3	3	1
H_R across loci	0.991	0	0	0	0	0	0	0	0	2	1.897	0

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), A_R allelic richness for each population averaged across loci, *H* haploid diversity, N_H number of haplotypes for each population averaged across loci, H_R haplotype richness for each population averaged across loci. Bold values indicate loci that deviated from Hardy-Weinberg equilibrium. ^{A} denotes a private allele, ^{P} denotes a private haplotype. Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

Table 2.4 Pairwise F_{ST} values for nSSR data among all sampling locations based on Weir and Cockerham's (1984) estimate. Darker color – increasing F_{ST} value, lighter color – decreasing F_{ST} value.

	GM	PS	TC	GHB	SBP	DC	DP	EB	PB	SB	ZP	AD
GM	–	–	–	–	–	–	–	–	–	–	–	–
PS	0.221	–	–	–	–	–	–	–	–	–	–	–
TC	0.147	0.121	–	–	–	–	–	–	–	–	–	–
GHB	0.264	0.245	0.221	–	–	–	–	–	–	–	–	–
SBP	0.261	0.246	0.230	0.047	–	–	–	–	–	–	–	–
DC	0.321	0.320	0.277	0.063	0.041	–	–	–	–	–	–	–
DP	0.272	0.265	0.246	0.026	0.029	0.025	–	–	–	–	–	–
EB	0.220	0.240	0.175	0.081	0.069	0.093	0.083	–	–	–	–	–
PB	0.290	0.283	0.260	0.062	0.040	0.050	0.050	0.106	–	–	–	–
SB	0.253	0.266	0.207	0.094	0.057	0.077	0.069	0.068	0.066	–	–	–
ZP	0.240	0.238	0.211	0.071	0.040	0.090	0.055	0.071	0.082	0.017	–	–
AD	0.298	0.254	0.231	0.121	0.121	0.137	0.123	0.170	0.128	0.132	0.128	–

Notes: All values significant at $P \leq 0.005$ after FDR correction (Benjamini and Hochberg 1995). Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

Table 2.5 Analysis of Molecular Variance (AMOVA) for 14 nuclear and 2 chloroplast SSR loci in 12 populations of baby's breath. Regional differences identified in the Bayesian clustering analysis ($K = 2$) were included in AMOVA, and the analysis was based on **Φ** estimates to compare variance across both marker types following Excoffier et al. (1992) and Weir and Cockerham (1984).

	nSSRs				cpSSRs			
Source of Variation	df	% of Variation	**Φ** − statistics	P − value	df	% of Variation	**Φ** − statistics	P − value
Among regions	1	22.62	0.226	0.0001	1	26.27	0.263	0.0001
Among populations within regions	10	12.89	0.167	0.0001	10	47.37	0.643	0.0001
Within populations	301	64.49	0.355	0.0001	301	26.36	0.736	0.0001

FIGURES

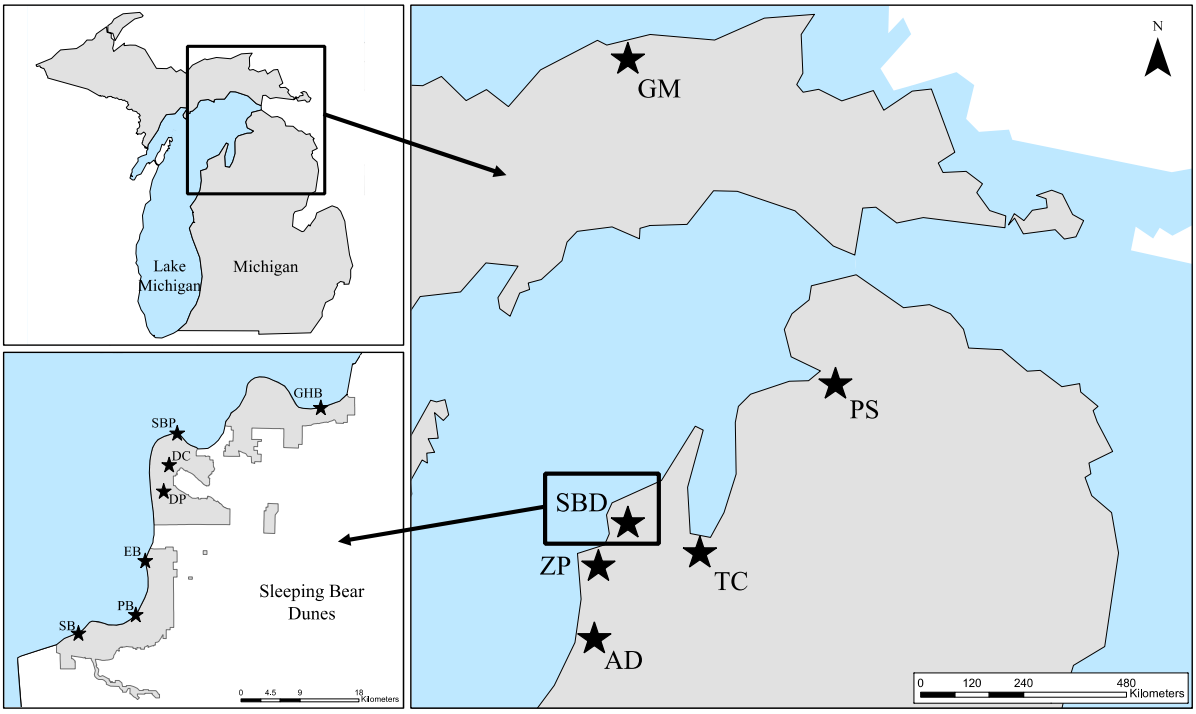


Figure 2.1 Map of baby's breath sampling locations in Michigan. Seven were located throughout Sleeping Bear Dunes National Lakeshore. Park boundary delineated by grey shading in bottom left panel. Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

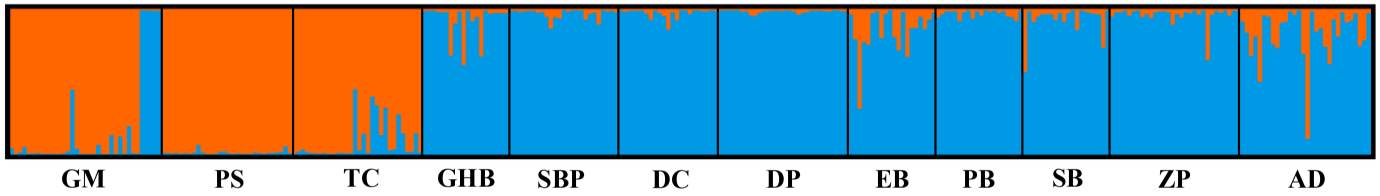


Figure 2.2 Results from Bayesian cluster analysis based on nSSR data using the program STRUCTURE (Pritchard et al. 2000) indicate (K = 2) population clusters of baby's breath (Pritchard et al. 2000, Evanno et al. 2005). Cluster 1 (left) includes the northernmost populations, and Cluster 2 (right) includes all other populations. Each individual (N = 313) is represented by a line in the plot, and individuals are grouped by population. Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

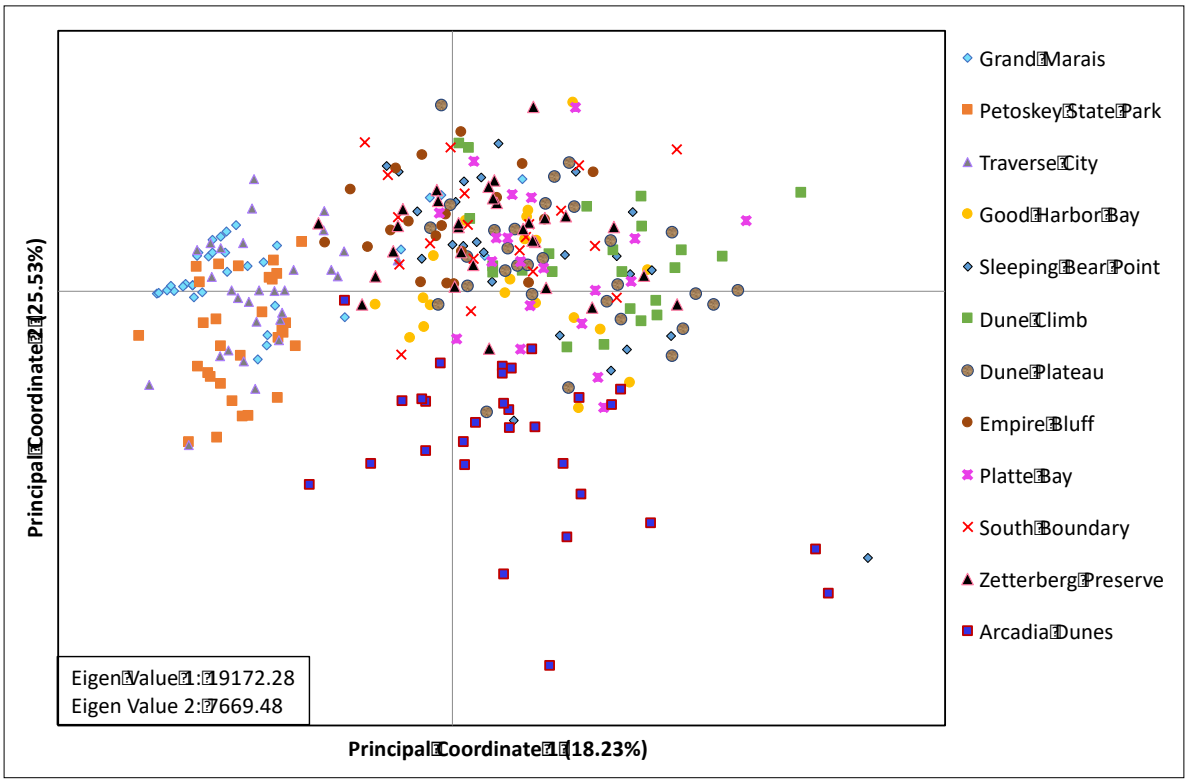
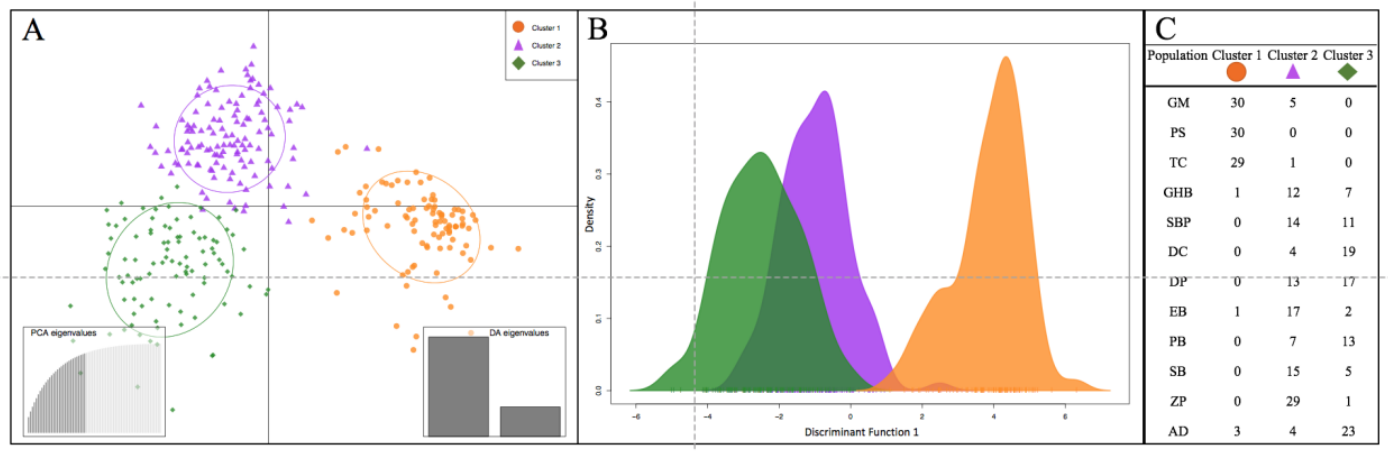


Figure 2.3 Principal Coordinates Analysis (PCoA) based on a genotypic distance matrix between all baby's breath individuals performed in GenAlEx 6.502 (Peakall and Smouse, 2006, 2012). Individuals labeled by sampling location.



Population	Cluster 1	Cluster 2	Cluster 3
GM	30	5	0
PS	30	0	0
TC	29	1	0
GHB	1	12	7
SBP	0	14	11
DC	0	4	19
DP	0	13	17
EB	1	17	2
PB	0	7	13
SB	0	15	5
ZP	0	29	1
AD	3	4	23

Figure 2.4 Discriminant analysis of principal components based on baby's breath nSSR data and calculated in the *adegenet* v2.1.0 (Jombart 2008, Jombart et al. 2010) package for R. (A) Scatterplot of both discriminant function axes; all individuals (n = 313) are included and represented by a dot. (B) Plot of DAPC sample distribution on discriminant function 1. (C) Table of cumulated variance explained by the eigenvalues of the PCA used in the DAPC, based on all 69 identified principal components.
Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

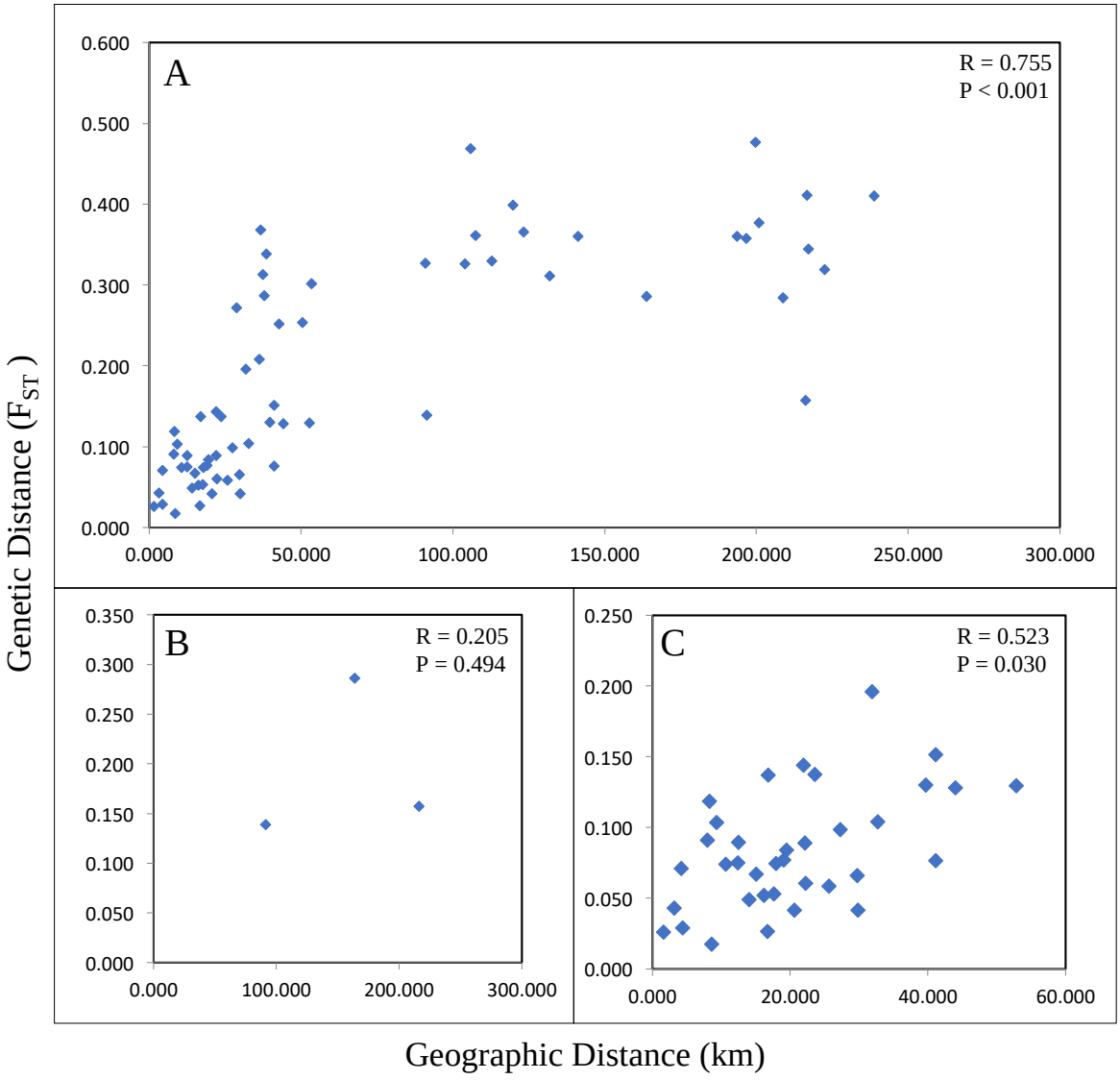


Figure 2.5 Mantel tests using transformed pairwise population F_{ST} values of nSSR data (Slatkin 1995) and straight-line distances (km) between populations based on the mean center latitude and longitude of each location. (A) Between all populations, (B) between populations in the Northern region (cluster 1), and (C) between populations in the Southern region (cluster 2) identified from the Bayesian clustering analysis. Reported p-values based on the one-sided alternative hypothesis (H^1: $R > 0$).

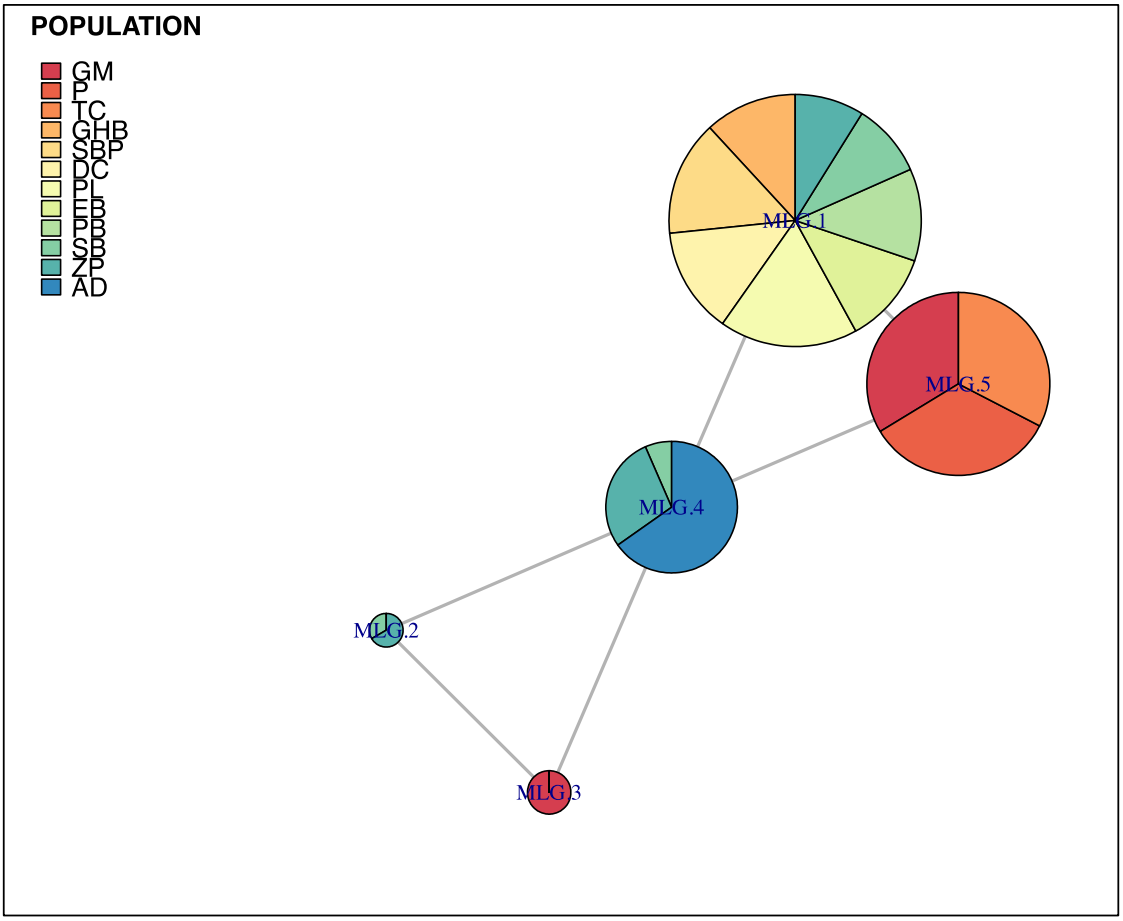
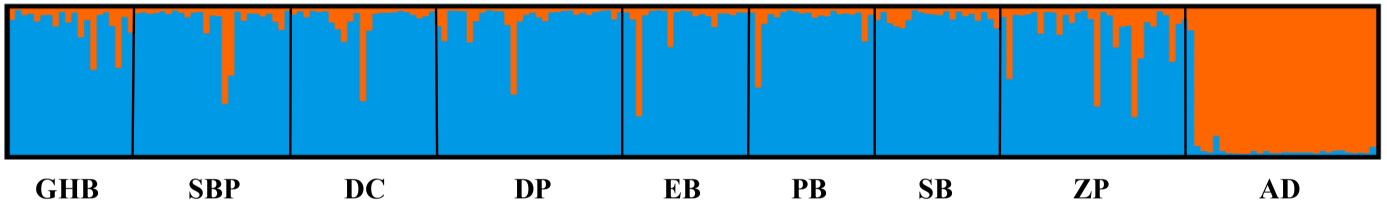
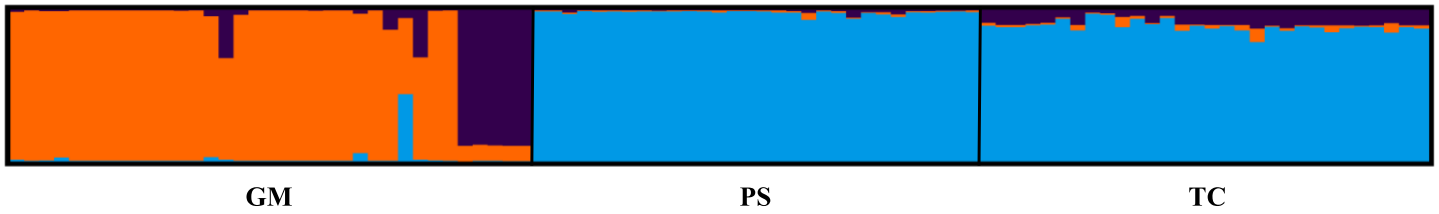


Figure 2.6 Minimum spanning network based on Nei's genetic distance (Nei 1972) matrix of baby's breath cpSSR data. Created in the *poppr* v2.8.0 package (Kamvar et al. 2014) for R. Illustrates the distribution of haplotypes across the 12 populations. Haplotype size indicates frequency in populations.
Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), Traverse City (TC), Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluffs (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).



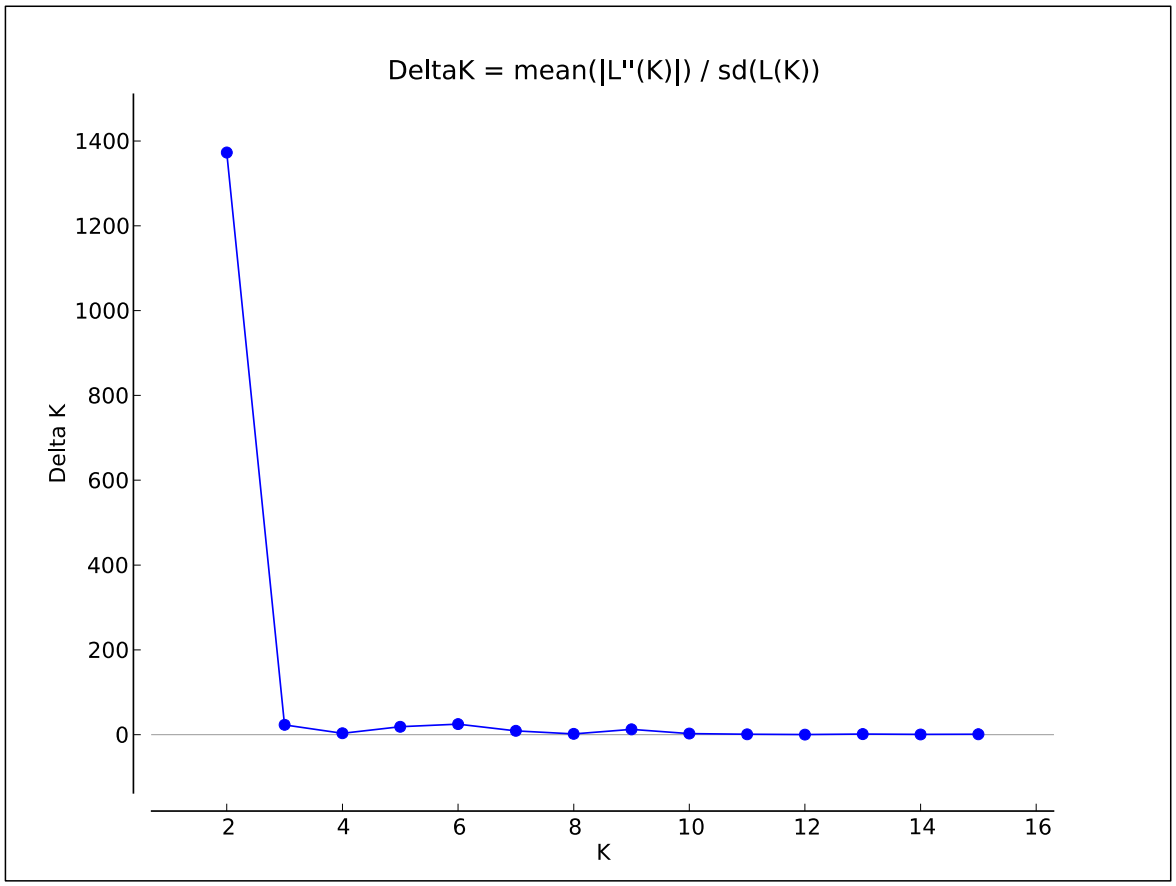
Supplemental Figure A Results from Bayesian cluster analysis based on the nSSR data of the previously identified cluster 2 using the program STRUCTURE (Pritchard et al. 2000). Two (K = 2) population clusters of baby's breath were also suggested for this analysis (Pritchard et al. 2000, Evanno et al. 2005). Cluster 1 (left) includes all the populations in Sleeping Bear Dunes and Zetterberg Preserve, and Cluster 2 (right) includes Arcadia Dunes. Each individual (N = 218) is represented by a line in the plot, and individuals are grouped by population.

Sampling location codes: Good Harbor Bay (GHB), Sleeping Bear Point (SBP), Dune Climb (DC), Dune Plateau (DP), Empire Bluff (EB), Platte Bay (PB), South Boundary (SB), Zetterberg Preserve (ZP), Arcadia Dunes (AD).

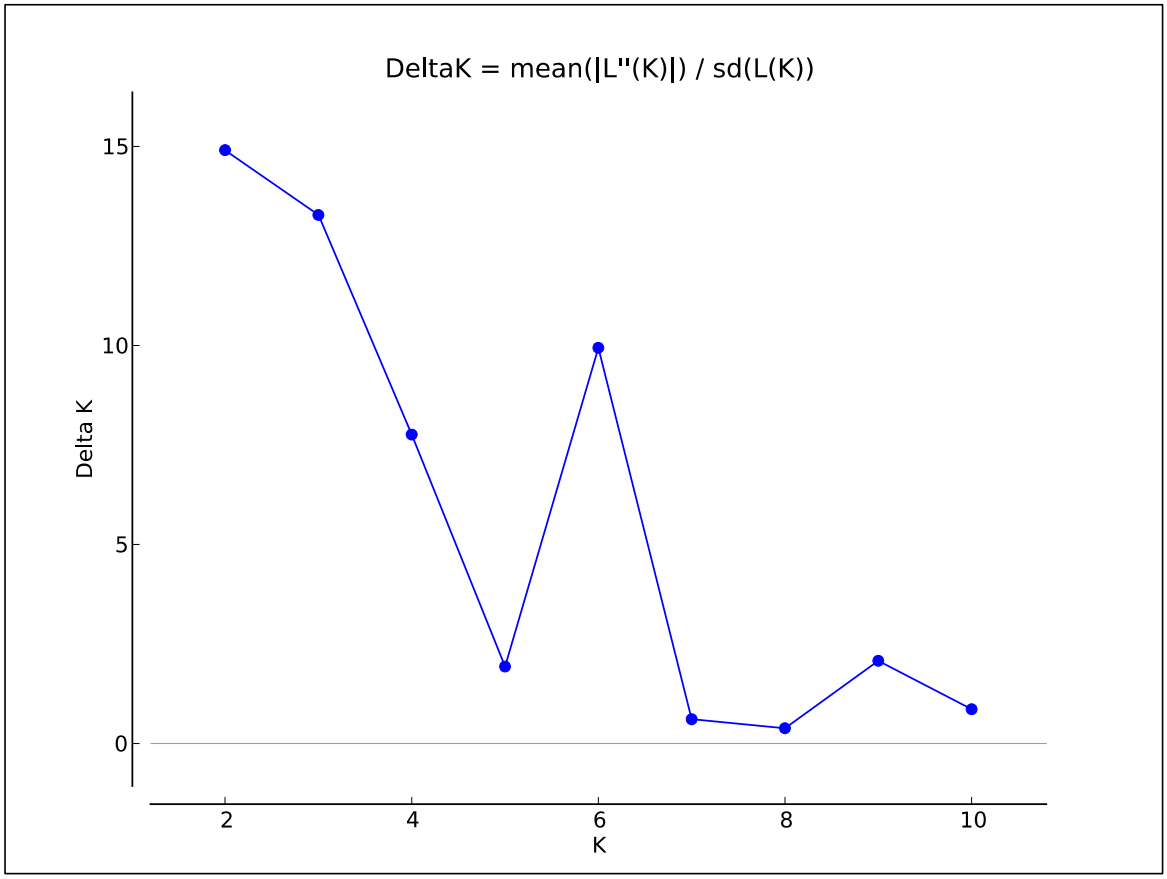


Supplemental Figure B Results from the Bayesian cluster analysis based on the nSSR data of the previously identified cluster 1 using the program STRUCTURE (Pritchard et al. 2000). Three ($K = 3$) population clusters of baby's breath were suggested for this analysis (Pritchard et al. 2000, Evanno et al. 2005). Cluster 1 includes 25 of the individuals in Grand Marais, Cluster 2 includes the five Grand Marais individuals that were in a slightly separate location at that site, and Cluster 3 includes individuals in Petoskey State Park and Traverse City. Each individual ($N = 95$) is represented by a line in the plot, and individuals are grouped by population.

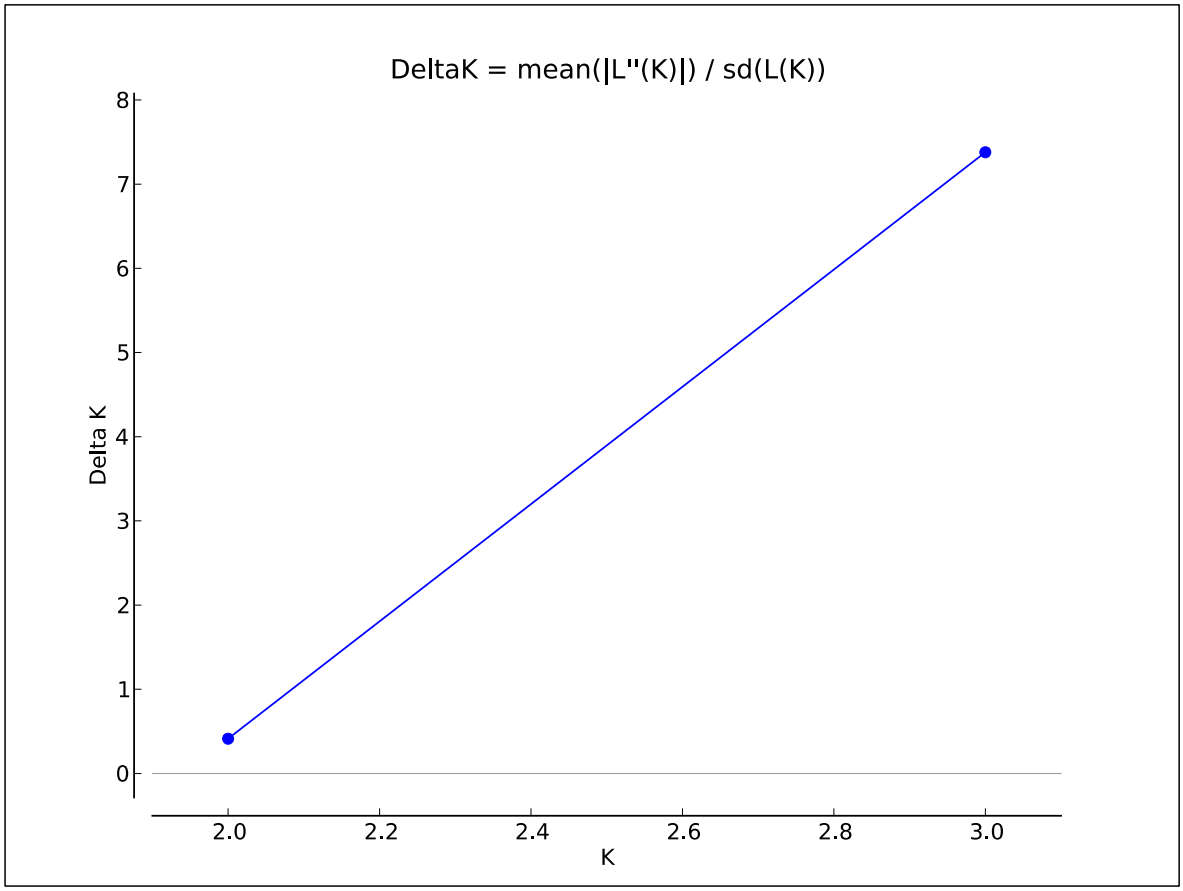
Sampling location codes: Grand Marais (GM), Petoskey State Park (PS), and Traverse City (TC).



Supplemental Figure C Plot of Evanno's ΔK method (Evanno et al. 2005) based on a Bayesian clustering analysis of all 12 populations from the program STRUCTURE (Pritchard et al. 2000). The largest rate of change suggests the highest likelihood of cluster number. This analysis was run without inferring any prior information on sampling location. Two genetic clusters were inferred from this data and
Ln Pr(X|K).



Supplemental Figure D Plot of Evanno's ΔK method (Evanno et al. 2005) for the southern cluster of the first Bayesian clustering analysis from the program STRUCTURE (Pritchard et al. 2000). The largest rate of change suggests the highest likelihood of cluster number. This analysis was run without inferring any prior information on sampling location. Two genetic clusters were inferred from this data and Ln Pr(X|K).



Supplemental Figure E Plot of Evanno's ΔK method (Evanno et al. 2005) for the northern cluster of the first Bayesian clustering analysis from the program STRUCTURE (Pritchard et al. 2000). The largest rate of change suggests the highest likelihood of cluster number. This analysis was run without inferring any prior information on sampling location. Three genetic clusters were inferred from this data and LnPr(X|K).

Chapter IV

Extended Review of Literature

Research overview

Invasive species pose a worldwide threat to the biodiversity of native communities, but often little is known about an invasion, or how it should be best managed. One particular ecosystem under duress of biological invasions is the Great Lakes dune system, where increased disturbance has made it more susceptible to invasive species colonization. Baby's breath (*Gypsophila paniculata*) is one plant species that is of concern to this dune system, as it jeopardizes populations of several threatened and endangered species. However, it is invasive throughout North America (Albert, 2000; Darwent, 1975). The population structure and genetic diversity of invasive populations in northwest Michigan are unclear. This information could aid in targeted management and help protect biodiversity by identifying subpopulations with high adaptive potential, and with increased likelihood of population spread. What follows is a literature review covering all the background information relevant to my thesis.

Invasion biology

The Problem with invasive species

Biological invasions are widely seen as a contributor to environmental change (Sakai et al. 2001; Chown et al. 2015). The effects of climate change can increase the effects of biological invasions, and the potential consequences of such events occurring has led to an increased focus in research on biological invasions (Chown et al. 2015). By reducing biodiversity and the productivity of ecosystem services, invasive species threaten the integrity of native systems, which in turn negatively impacts the myriad of ways in which humans benefit from natural

systems, such as water and air filtration (Pejchar and Mooney, 2009). Subsequently, the financial cost associated with managing invasive populations is high: estimates range from millions to billions of dollars annually (Sakai et al. 2001). With invasions predicted to become more frequent and more destructive in the advent of climate change, it is important to understand the mechanisms driving these invasions in order to improve current management strategies (Chown et al, 2015).

Field of invasion genetics

Molecular ecologists and evolutionary biologists have a complicated relationship with nonnative invasive species. They recognize that invaders can threaten biodiversity and productivity of ecosystem services (Sakai et al. 2001). At the same time, they relish in the prime opportunity invasive species present to test many ecological and evolutionary hypotheses, due to the swift response many invasive populations have in evolving to persist in novel abiotic and biotic circumstances (Sakai et al. 2001; Bock et al. 2015; Lawson-Handley et al. 2011). It is argued that the field of invasion genetics began in 1964, when a group of scientists came together to discuss the evolutionary changes that might occur as nonnative species colonize novel environments (Barrett, 2015). This symposium, and the book that became the product of the ideas discussed at the symposium, *The Genetics of Colonizing Species*, set into motion a set of research questions that have shaped the field of invasion genetics that we know today (Barrett, 2015). Barrett (2015) defines invasion genetics as

"the study of the historical, ecological, and demographic processes responsible for the patterns of genetic diversity in populations and their influence on invasion success and contemporary evolution during biological invasion."

Mechanisms driving invasion success

Though it has been over 50 years since the formalization of the field of invasion genetics, the research being performed today attempts to answer many of the same questions asked at that first symposium. Several mechanisms have been found to promote invasion success: pre-invasion and post-invasion adaptation, phenotypic plasticity, intraspecific admixture and interspecific hybridization, propagule pressure, genetic diversity, and gene flow (Whitney and Gering, 2014; Barrett, 2015). H.G. Baker's (1965) hypothesis of a "general-purpose genotype" suggests that certain genotypes give an individual the ability to respond to a variety of different environmental conditions due to increased phenotypic plasticity in that genotype. Baker argued therefore, that phenotypic plasticity increased invasiveness (Baker, 1965). However, identifying plasticity in fitness-related traits, and the degree to which plasticity facilitates invasion success can be difficult.

Just like H.G. Baker's (1965) "general-purpose genotype," other scientists have tried to identify factors that reliably predict invasiveness, including phenotypic traits, environmental conditions, phenotypic plasticity, and the history of the invasion process itself. Several phenotypic traits such as sexual and asexual reproductive capabilities, rapid development, and large seed crops have been associated with invasive species, but these traits do not always confer invasion success (Sakai et al, 2001; Darwent, 1975). Meanwhile, Moles et al. (2008) argue that there is not a definitive set of traits that invasive species exhibit, and they also argued that the

environmental conditions at the time of the nonnative species colonization play a large role in what results in a successful species invasion.

The history of the invasion process and the factors that lead up to the actual invasion event itself can also influence the success of an invasion. For example, the number of individuals in a colonizing population, or propagule pressure, has been found to correlate with colonization success (Wilson et al. 2009). However, it is still unclear whether colonizing population size is the primary factor influencing invasion success, or whether consequences of the population size are driving success. Increased genetic diversity and the resulting increased adaptive potential, which is positively correlated with founding population, are two consequences of colonizing population size that could actually be the drivers of successful invasion (Whitney and Gering, 2015).

The publishing of *The Genetics of Colonizing Species* by Bakers and Stebbins (1964), and the research that has followed, addresses the evolutionary changes that can influence whether a biological invasion is successful or not. However, it was Elton's (1958) publication of *The ecology of invasions by animals and plants* that addressed the ecological mechanisms that can also influence the success or failure of an invasion. Ecologists suggest that hypotheses such as Evolution of Increased Competitive Ability (Blossey and Notzold, 1995), Enemy Release Hypothesis (Keane and Crawley, 2002) and Escape From Enemy Hypothesis (Wolfe, 2002) should not be underestimated when considering mechanisms of invasion. These hypotheses address tradeoffs in energy expenditure and the biotic interactions introduced species are released from (e.g. predators) in novel environments. Moles et al. (2008) suggests a framework that uses traits of native species, traits of invading species, and environmental conditions to predict successful invasions, arguing that abiotic conditions play a large role in predicting a successful species invasion.

Use of genetic and genomic tools

While current research questions are similar to those asked back in the 1960's when the field of invasion genetics emerged, the ways scientists approach these questions has changed. Due to the increasing development of genetic and genomic methods, researchers have many more tools with which to identify the mechanisms facilitating species invasions. Using genetic tools, we can calculate the genetic diversity and structure of recently colonized invasive populations, and in using genomic tools, we can identify what genes influence fitness in a newly colonized population.

It is unknown to what extent admixture and genetic diversity drive invasion success (Dlugosch et al. 2015), but genetic diversity can be an effective way to quantify population differences, structure, and general invasion history. Selectively neutral markers like amplified fragment length polymorphisms (AFLPs) and microsatellites are useful for identifying population-level genetic differences, and in comparing genetic structure of the native and introduced ranges (Stout et al. 2014; Stabile et al. 2016). By using a combination of molecular markers, such as nuclear and chloroplast markers in plants, modes of reproduction and movement patterns can be identified in invasive populations (Sakata et al. 2015; Piya et al. 2014). *Falcaria vulgaris* (sickleweed) is an invasive species in the Midwestern United States, and Piya et al. (2014) used nuclear microsatellites, and nuclear and chloroplast DNA sequences to identify the population genetic diversity and structure. By using both maternal and biparentally inherited markers, they were able to identify that while diversity between populations was high, there was no logical structure. This led them to conclude that several human-mediated introductions resulted in a more random distribution of genotypes and

chlorotypes (Piya et al. 2014). Though analysis of neutral markers will not provide information regarding evolutionary change of an invasive population, it can provide estimates of the genetic diversity present within populations, how that variation is distributed among populations, and how populations may be spreading.

Beyond the use of neutral genetic markers, the increased use of genomic tools such as genotype-by-sequencing (GBS) and quantitative trait locus (QTL) mapping has led to a finer-scale understanding of the evolutionary mechanisms facilitating invasion success (Chown et al. 2015). Using next-generation technology, a targeted portion of the genome is sequenced during GBS, and from there specific loci can be linked with phenotypic traits that increase fitness during high-throughput QTL mapping of that targeted region. Paterson et al. (1995) used restriction fragment length polymorphisms (RFLPs) to isolate and sequence a portion of the *Sorghum halepense* (Johnsongrass) genome in both wild and cultivated species. They found a locus thought to determine the number of vegetative buds that go on to become either rhizomes or tillers (Paterson et al. 1995). Therefore, by targeting specific genes, results from genomic analysis can lead the field of invasion genetics to a better understanding of how species can persist in novel environments.

It is difficult for scientists to define patterns in invasive species, and predict invasions that are applicable across a wide ecosystem range (Lawson-Handley et al. 2011). By combining both branches of invasion biology (evolutionary drivers and ecological drivers) a more holistic understanding of invasive populations can be achieved. However, it can take years of research to produce answers robust enough for management implications. A continuing challenge to this field will be to drive research that has applicability to the management of nonnative invasions (Chown et al. 2015).

Great Lakes dune systems

Brief history

Due to their dynamic nature and continual disturbance, dune ecosystems are constantly under the stress of plant invasions (Karamanski, 2000). Though coastal dune systems are most prevalent along oceanic coasts, there is one particularly unique coastal dune system residing in the midwestern United States. The Great Lakes dune system is the largest freshwater complex of its kind, and is even said to rival marine coastal dune systems in terms of sheer size (Arbogast 2015). The majority of the Great Lakes dunes are present along the eastern shore of Lake Michigan, the southeastern shore of Lake Superior, and western shore of lake Huron. The geographic positioning of the dunes is attributed to the abundance of fine grain sand leftover from glacial retreats, the position of the shoreline that is privy to direct winds off the lakes, and the substantial fetch, or distance the wind is able to travel over both lakes prior to reaching the shore (Arbogast 2015).

As glaciers retreated and the prehistoric lake Nipissing was left in the area that would become Lakes Michigan, Huron and Superior approximately 6000 years ago, the formative years of dune creation began along the shores of the Great Lakes (Albert, 2000; Arbogast 2015). As the glaciers retreated, glacial drift (mixture of boulders, cobbles, sand, and clay) was eroded by lake wave action and deposited onshore, becoming the sand source that would over time become the substrate of the dunes (Albert, 2000).

Since then (~6000 years ago), the dunes have been maintained by winds gusting up from the adjacent lake and both shifting the existing dunes by blowing sand around, and also by drying previously-deposited sand particles from the lake and carrying them up onto shore, where vegetation plays an imperative role in trapping sand, causing accumulation over time resulting in

dune formation (Cowles 1899, Arbogast 2015). This dune system could not exist without the influence of the adjacent lakes, and the growth and stability of the dunes is thought to be connected to the fluctuations in lake levels, weather patterns, and storm events (Arbogast and Loope, 1999; Blumer et al. 2012).

Economic benefits

Some biophysical processes inherent to coastal dune systems also have socio-economic value. Coastal sand dunes act as a natural buffer, absorbing and dissipating the energy from wave and wind action off the adjacent water, and it is also hypothesized that the dunes help filter and purify water (Everard et al. 2010). This buffering can largely benefit human infrastructure by protecting against adverse storm conditions. It is thought that this importance will increase in the advent of climate change, to protect inland areas from severe storms and erosion (Cochard et al. 2008). Sand dunes can also be a major source for sand mining. The Michigan Department of Environmental Quality reported that from 1978 - 2015, between 1 million and 3 million tons of sand was mined annually from Michigan. While many of the mining sites have since closed, and legislation has been created to decrease the negative effects of mining on the ecosystem, as of 2017, nine sites are still active, and they all exist within 2 miles of the lake Michigan shoreline (MI DEQ).

Despite the natural socio-economic benefits associated with dune systems, in the Great Lakes, the main economic value of the dune systems is the tourism opportunities it offers (Karamanski, 2000). With a rich natural and anthropological history of the region, Sleeping Bear Dunes National Lakeshore (SBDNL) and other state and national parks attract millions of visitors every year, making the area economically important to northwest Michigan

(Karamanski, 2000). With tourism brings increased real estate market and patronage to the businesses and along the coasts, and thousands of jobs have been created due to the increased tourism in the area (Karamanski, 2000).

Ecological importance

Though formed thousands of years ago, coastal dune systems are very dynamic, and continue to change at various rates depending upon their age (Arbogast 2015). It is the heterogeneous topography and successional processes due to continuous disturbance that makes the dune system so unique, and the ecological niches that occupy dune systems are equally diverse (Everard et al, 2010). Many of the flora and fauna native to the Great Lakes dune system are adapted to the harsh conditions posed by the adjacent coast, and some of them require sparsely vegetated successional habitats with open sand to thrive (Albert, 2000; Everard et al. 2010). For example, some species found within the Great Lakes dunes are adapted to the strong winds, extreme temperatures, intense light, and low soil moisture and nutrients (Albert, 2000).

Ammophila brevigulata (Marram grass), *Tanacetum huronense* (Lake Huron tansy), *Cirsium pitcheri* (Pitcher's thistle), and other native plant species are especially adapted to sand burial, and will continue to grow above the sand height while their roots grow into the sand (Albert, 2000). The species *Cirsium pitcheri* (Pitcher's thistle), *Trimerotropis huroniana* (Lake Huron locust), and *Charadrius melodus* (Piping Plover) are all either state or federally threatened or endangered species, to which the Great Lakes dune system is important habitat. While the first two are native to the Great Lakes, *Charadrius melodus* uses the Great Lakes beaches, foredunes and gravelly bluffs for nesting habitat (Albert, 2000).

Though environmental disturbance is inherent to this system, the increase in human foot traffic due to growing real estate and tourism exponentially increases disturbance to the biological community. For example, Rowland and Maun (2001) concluded that disturbance and habitat reduction was a contributing factor to the decline in abundance of *C. pitcheri*, which is endemic to the Great Lakes dune system. Similarly, *C. melodus* nests were being destroyed, and nesting pairs killed, by the increased presence of dogs and raccoons associated with developing real estate along the beaches (Karamanski, 2000; Albert, 2000).

With increased disturbance comes an increased risk of biological invasion. Invasive species are often adept at colonizing disturbed areas, which makes dunes highly susceptible to biological invasions (Sakai et al. 2001; Albert, 2000). Specifically, spotted knapweed (*Centaurea stoebe*), garlic mustard (*Alliaria petiolate*), baby's breath (*Gypsophila paniculata*), and purple loosestrife (*Lythrum salicaria*) have had deleterious effects to various habitats within the Great Lakes dune system, which has led to costly management efforts in attempts to curb their effects (Albert, 2000; Karamanski, 2000). Their effects include inhibiting native plant growth due to allelopathy, outcompeting native vegetation, eliminating habitat for other species, and thus altering the native plant and animal community (TNC, 2013). It is imperative that the Great Lakes dune system be managed carefully in order to maintain the natural dynamic ecosystem that supports a biodiversity found nowhere else in the world (Arbogast, 2015).

Baby's breath (*Gypsophila paniculata*)

Morphology and life history

Baby's breath (*Gypsophila paniculata*) is a dicotyledonous, herbaceous perennial forb of the family Caryophyllaceae. Mature individuals can grow up to 0.75 meters high, with many erect

stems (Darwent and Coupland 1966). In Darwent and Coupland's (1966) field experiment, they observed an average of 18 shoots per plant in the field. Baby's breath has many branches of panicle-like compound inflorescences, with 5 wedge-shaped white petals, and a spherical fruit capsule that holds long black seeds (Darwent, 1975). It is capable of producing up to 14,000 seeds per plant, and has a taproot that has been documented as extending up to 4 meters in depth and 4-7 cm in diameter. This large taproot contains abundant energy stores allowing baby's breath to overwinter in harsh environments (Darwent and Coupland 1966, Darwent 1975).

Darwent (1975) observed the plants as taking two years to mature, with rapid root growth during the first two years, and an average taproot length of 62 cm after one year of growth (Darwent and Coupland, 1966). They observed that as seeds were maturing, the stems became dry and brittle. Darwent and Coupland (1966) hypothesized that wind was the primary driver of dispersal, applying stress on the dried stems and causing them to eventually break off. As they break off, the winds push the tumbling stems across the landscape, scattering the loosely-held seeds. These seeds are dropped both near the parental plant and up to 1 kilometer from the parent plant (Darwent and Coupland, 1966).

Distribution

Baby's breath occurs from eastern Europe to western Asia, with the epicenter of the genus *Gypsophila* origin occurring in north Iraq and Iran, in the Black Sea and Caucasus regions (Darwent, 1975). As it is native to the semi-desert Steppe regions of Eurasia, it commonly grows in areas of low annual precipitation and extreme seasonal and day-night temperature differences (Darwent, 1975). The first documentation of baby's breath in North America was in Manitoba, Canada in 1887. Since then it has been identified throughout North America as an adventive and

weedy species, in Saskatchewan, British Columbia, and Alberta, as well as Washington, Oregon, California, and Michigan (Gleason and Cronquist, 1963; Darwent 1975; DiTomaso and Kyser et al. 2013).

It prefers sandy, calcium-rich, well-drained soil, and can persist in a variety of environmental conditions, including low soil moisture, extreme temperatures and high winds, conditions typical of its native Steppe region (Darwent, 1975). In North America, baby's breath is found invading several areas: agricultural fields and rangelands of the west, Pacific northwest and Canada; disturbed areas such as roadsides and ditches; and the Great Lakes dune system (Darwent, 1975; DiTomaso and Kyser et al. 2013; TNC, 2013). While it is listed as a noxious weed in California and Washington, and has been documented across the West as invasive on state, federal, and university-based weed lists, there has been little research into its invasive range.

Economic benefits

It is hypothesized that baby's breath was introduced to North America due to its popularity as a garden ornamental, and has since escaped cultivation and become invasive across North America (Darwent and Coupland, 1966; Darwent, 1975). While it is still a common garden ornamental, it is a very important species to the cut flower industry. Vettori et al. (2013) recently reported it as one of the top ten best selling cut flower species globally; Calistri et al. (2014) stated that it was in the top twenty most economically important ornamental species worldwide. Zvi et al. (2008) also claimed that baby's breath was the most important species to the commercial cut flower industry. There are large growing facilities in Ecuador, Columbia, Peru, Mexico, Costa Rica, and

Israel to produce cut flowers at a high volume to meet the high demand (Halevy, 1999; USDA – APHIS internal report 2012).

Due to the demand in the floral industry for heterogeneity in the floral organs, breeding programs have yielded many varieties of baby's breath (Calistri et al, 2014). However, this breeding has led to most plants used for cut flowers to be sterile. Wild species are used as sources of pollen, but desired varieties are primarily grown from vegetative propagation (Calistri et al, 2014). Several cytological and genetic studies have been done to better identify the causes of infertility of these varieties, as well as the relation of these species to each other (Calistri et al, 2014; Vettori et al, 2013). Despite its economic importance to the commercial cut flower industry, baby's breath can have deleterious effects as an invasive species once it has escaped into new habitats.

Ecological impacts

Invasive populations of baby's breath could be preventing or slowing the reestablish of native species (Emery et al 2013). With a few key morphological characteristics, namely its extensive taproot and high seed crop, baby's breath is able to outcompete native species for resources, which can lead to monotypic stands in the dunes (Darwent, 1975). It has been reported to occupy up to 80% of the vegetation in the Great Lakes dunes (Karamanski, 2000). Plant species native to dune ecosystems are often short-distance dispersers, and can be both dispersal-limited (French et al. 2011) and seed-limited (Leicht-Young et al. 2009), due in part to the continuous disturbance of the dune habitat. When the disturbance level is increased with a plant invasion, and an invasive population outcompetes natives plants locally, it can be difficult for native plants to recolonize the area due to their limitations (Emery et al. 2013).

In one of few studies performed on invasive baby's breath, Emery et al. (2013) set up research plots in the dunes of northwest Michigan to measure the impacts of baby's breath removal on native and non-native plant communities, as well as arbuscular mycorrhizal spore abundance, soil nutrients and sand movement. While they found no statistically significant association between baby's breath removal and changes in soil properties, native species (pitcher's thistle) abundance, and nonnative species cover, they did find that, prior to removal, baby's breath presence was associated with reduced native species abundance and increased presence of non-native species. Additionally, they did find a slight increase in pitcher's thistle abundance in areas where baby's breath removal took place (Emery et al. 2013).

There have been other accounts of the negative effects baby's breath has in its introduced range. The invasion of baby's breath could be altering the arthropod and pollinator communities and plant-pollinator interactions in the Michigan dune system (Emery and Doran 2013; Baskett et al. 2011). Emery and Doran (2013) analyzed baby's breath invasions and management in the Michigan dune system and how they might affect the arthropod community. They found that management of baby's breath did not have an effect on the arthropod community. However, there was an increase in herbivore arthropod dominance, as well as pollinator and predator mean abundance associated with the presence of baby's breath (Emery and Doran, 2013). Another study in the Michigan dunes by Baskett et al. (2011) found that invaded plots had higher pollinator species richness than plots where baby's breath was removed or naturally uninvaded. However, they also found that following removal, there was an increase in pollinator visits to native species Pitcher's thistle (*Cirsium pitcheri*), leading the authors to the conclusion that the baby's breath invasion is limiting pollinator visits to threatened native species (Emery and Doran, 2013).

Management of baby's breath in Michigan dune system

History of management

Invasive baby's breath was first documented in Emmet County, Michigan in 1913 (Emmet Co., 1913, *Gleason s.n.*, MICH), and became prevalent around SBDNL in the 1950's (Emery and Doran, 2013). The Nature Conservancy (TNC) began treating in Northwest Michigan in the mid 1990's (TNC, 2013). However, with funding from Meijer Corporation, The Nature Conservancy partnered with The National Park Service via SBDNL in 2007 to create the "Lake Michigan Dune Restoration Project," a joint effort to address the baby's breath invasion in the dune system (TNC, 2013). The Michigan Dune Alliance was created in 2009, and initiated a coordinated effort between the various state, non-profit, and regional land agencies to protect the Great Lakes dune system. Then in 2015, the Michigan Department of Natural Resources identified baby's breath as a main threat to the coastal dune systems along Lake Michigan.

Management types

The Nature Conservancy primarily uses herbicide, but at times this kind of treatment is not feasible, they implement manual treatment with shovels (TNC, 2013). Glyphosate is an aromatic amino acid inhibitor commonly known as Roundup, and is mixed at a concentration of 2% and sprayed directly onto the vegetative part of the plant. Management crews of TNC and SBDNL employ "manual" removal by using a sharp spade to sever the taproot below the woody caudex (TNC, 2013). Though TNC and SBDNL have found herbicide treatments to be both time efficient and effective at killing baby's breath, the weather patterns in the dunes (high temperatures, wind, rain), and the difficult terrain often make this method difficult. There is also considerable variation in how each individual applies the herbicide, often leading to inconsistent

results and the need for retreatment. This can lead to very high financial costs to treatment, as herbicide is a costly expense. Manual removal efforts have also been effective at killing the plants, but it is very time consuming and tedious. However, it is the more versatile of methods when considering weather and terrain.

The financial cost of managing baby's breath invasions is unknown, but it has the potential to become a financial drain on farmers, ranchers, and resource managers, due to its persistence in a variety of conditions (Darwent, 1975; Ditomaso and Kyser et al. 2013). Though several studies analyzing the biology and ecology of baby's breath invasions were highlighted here, there are many gaps in knowledge that need filling in order to fully understand the ecological impacts associated with its invasion in North America.

Conclusions

Gaps in Knowledge

Invasive baby's breath can outcompete species such as pitcher's thistle for resources, alter the plant community that the Lake Huron locust (*Trimerotropis huroniana*) is adapted to foraging on, and occupy the once-open gravelly bluffs where piping plover (*Charadrius melodus*) traditionally nests (Darwent and Coupland, 1966; Albert, 2000; Emery et al. 2013; Baskett et al. 2011). The threat baby's breath poses to fragmenting and limiting the already tenuous population of pitcher's thistle in the lake Michigan dune system has driven an increase in the concern over the invasion of baby's breath.

Resource managers have put in mammoth efforts to control the invasion and reinvasion of baby's breath throughout Northwest Michigan. However, to date, no appropriate molecular markers have been developed to identify the connectivity of the northwest Michigan populations,

a region where the invasion could have deleterious effects. Data revealing how populations of baby's breath are spreading through the region could benefit restoration efforts, and decrease the amount of time they spend retreating invaded areas by targeting populations that are sources for spread throughout the area. Currently, resource managers attempt to treat as many infested areas as they can through the Lake Michigan dune system. By estimating the genetic variation present in each population, management crews can identify areas to target to optimize habitat preservation and restoration, and prevent further spread or population admixture.

We are using molecular markers called microsatellites to estimate the population structure of baby's breath in northwest Michigan. Microsatellites are molecular markers that are affordable to develop, selectively neutral, and have high mutation rates, which allows us to identify population level differences in genetic variation.

Research Questions

1. What is the population genetic structure and variation of baby's breath in northwest Michigan?
2. What is the main source influencing regrowth in areas managed by The Nature Conservancy and the National Park Service?

Through our research, we found strong genetic structure among all the populations of baby's breath sampled along the dunes of Michigan, and we also found that these populations segregated into two genetic clusters. These clusters are separated geographically by the Leelanau peninsula, which could be acting as a barrier, limiting gene flow among the two clusters and causing the strong structure. The land use of the Leelanau peninsula is much more residential and

varied, unlike the large protected areas of Zetterberg Preserve and Sleeping Bear Dunes, and this varied land use could be preventing gene flow of baby's breath between these two population clusters. By using both nuclear and chloroplast markers, we found that gene flow through seed dispersal could be more limited between populations than it is by pollination, a result that suggests the dynamic topography and habitat connectivity could be influencing the population structure of baby's breath, even on a small scale. The strong genetic distinction we found among these two population clusters suggests the possibility of at least two separate introduction events of baby's breath to Michigan. However, some admixture between the two clusters could be explained by human-mediated dispersal, a mechanism that would cause gene flow between otherwise isolated populations. These results together provide a better understanding of the invasion history of baby's breath and factors contributing to its invasion success.

The molecular markers we developed can be used throughout the native and invasive range of baby's breath, and in similar species. The population analysis of baby's breath in northwest Michigan has contributed to an adaptive management plan to aid TNC in conserving resources as they target populations to manage. Due to the size and amount of seeds each individual can produce (Darwent and Coupland 1966), management technicians should take precautionary measures to avoid spreading seeds and increasing gene flow between disparate populations. Additionally, education of the public should be a priority to prevent further gene flow between populations through accidental spread of seeds. Finally, due to the high genetic diversity and structure of populations, we recommend targeting specific populations for management, as some are more related to one another, and could be increasing the rate of regrowth and spread through continued gene flow. Due to economic viability of the dune system in the Great Lakes, the dynamic nature of the ecosystem, and the biodiversity it harbors, it is

important to the health of the economy and ecology of Michigan to manage this dune system appropriately. By supporting the management of invasive baby's breath, we are contributing the protection of a dynamic dune ecosystem and the economy that relies on it.

Extended Methodology

Microsatellite library development

The microsatellite library development and sequence analysis protocols (unpublished data) were provided to us by the Cornell University Ecology and Evolutionary Biology department (CU-EEB). The following is a summary of these methods. The five baby's breath samples were pooled and the genomic DNA was digested in three separate reactions using the following restriction enzymes: *AluI, RsaI,* and *HYP166II*. These reactions were combined and Klenow and dATP were used to adenylate the blunt ends. After being supplemented to 1 mM with ATP, T4 DNA ligase was used to ligate an Illumina Y-adaptor to the fragments of the combined digests. Fragments were hybridized to 3' biotinylated oligonucleotide repeat probes to enrich for microsatellites, and were then captured using streptavidin-coated magnetic beads. Enriched fragments were amplified using an Illumina primer pair (one universal and one indexed), and the PCR product was quantified using a Qubit 2.0 fluorometer. Libraries were then pooled and Ampure magnetic beads (Beckman Coulter) were used to target and recover fragments of 300-600 base pairs (bp). Libraries were then submitted to the Sequencing and Genotyping Facility at the Cornell Life Science Core Laboratory Center (CLC), and sequenced using a 2x250 paired end format on an Illumina MiSeq.

Assembly and microsatellite identification

The paired fastq files were imported into SeqMan NGen (v 11) and assembled *de novo*. Prior to assembly, adaptor sequences were removed and low quality ends (Q<20) were trimmed. The

assembly was constructed with minimum k-mer = 99, a gap penalty of 10, and the max gap set to 50 bp/kb. Default options were used for the remaining parameters.

Msatcommander (v 1.0.3) was used to search the fasta sequence files for dimeric, trimeric, tetrameric and pentameric microsatellite loci and potential primer pairs (Faircloth, 2008). The minimum length for each repeat unit was 6 for dimeric loci and 5 for all other repeat motifs. For primer design, we selected primers that produced products of 150-450 bp, that had a GC content between 30-70%, and that had a Tm between 58-62ºC, with an optimum of 60ºC. This produced roughly 3,893 unique primer pairs to be tested.

Primer optimization

Prior to PCR optimization, contigs containing potential primers were aligned using ClustalOmega to ensure they were targeting unique microsatellite regions (Sievers et al., 2011). We focused on 107 primer pairs that consisted of either tetrameric, trimeric, or dimeric motifs, and yielded products between 150-300 bp. Of the 107 primer pairs that were tested, 73 successfully amplified, and 16 were determined to be polymorphic and easily scored (Appendix S3). DNA from leaf tissue collected from three populations (Zetterberg Preserve, SBDNL, Petoskey State Park) along eastern Lake Michigan in 2016 was used for primer optimization (population geographic details included as footnote on Table 1.2). A minimum of 30 tissue samples were collected from each population. Tissue storage and DNA extraction methods are the same as previously stated.

PCR reactions consisted of 1x KCl buffer (Thermo Fisher), 2.0-2.5 mM $MgCl_2$ depending on the locus (Thermo Fisher), 300 µM dNTP (New England BioLabs, Ipswich, Massachusetts, USA), 0.08 mg/mL BSA (Thermo Fisher), 0.4 µM forward primer fluorescently labeled with

either FAM, VIC, NED, or PET (Applied Biosystems, Foster City, California, USA), 0.4 µM reverse primer (Integrated DNA Technologies, Coralville, Iowa, USA), 0.25 units of Taq polymerase (Thermo Fisher), and a minimum of 50 ng DNA template, all in a 10.0 µL reaction volume. The thermal cycle profile consisted of 94˚C for 5 minutes followed by 35 cycles of 94˚C for 1 minute, primer-specific annealing temperature (Table 1.1) for 1 min, 72˚C for 1 min, and a final elongation step of 72˚C for 10 minutes. Successful amplification was determined by visualizing the amplicons on a 2% agarose gel stained with ethidium bromide. Fragment analysis of the amplicons was performed on an ABI3130xl Genetic Analyzer (Applied Biosystems).

Study area and sample collection

To determine population structure on a regional scale in Michigan, we collected leaf tissue samples from 12 sites in the summers of 2016-2017. All sites were located in areas of known infestation along the dune system of Michigan (Figure 2.1), and all with a history of treatment primarily by The Nature Conservancy, the Grand Traverse Regional Land Conservancy, and the National Park Service (TNC 2013). Eleven sites were located along Lake Michigan in the northwest lower peninsula of Michigan, and one was located on Lake Superior in the upper peninsula. We collected leaf tissue samples (5-10 leaves per individual) from a minimum of 20 individuals per site (maximum of 35), and stored them in individual coin envelopes in silica gel until DNA extractions took place (n = 313). Site locations in Michigan (Table 2.1) were separated by a minimum of 10 km and a maximum of 202 km. We subjectively chose individuals to be sampled by identifying a visibly infested area, selecting individuals regardless of size, and walking a minimum of ~5 meters in any direction before choosing another plant to minimize the chance of sampling closely related individuals. We observed that the number of individuals at

the Petoskey State Park site was smaller and patchier than the others (~60 individuals total), so we conducted sampling more opportunistically. This opportunistic sampling involved collecting tissue from individuals that were less than 5 m apart, and in some areas sampling from all individuals (~3 – 4 individuals) within a small patch (~5m x 5m).

Microsatellite genotyping

We extracted Genomic DNA from all samples using DNeasy plant mini kits (QIAGEN, Hilden, Germany) and followed provided instructions with minor modifications, including an extra wash step with AW2 buffer. We then ran the extracted DNA twice through Zymo OneStep PCR Inhibitor Removal Columns (Zymo, Irvine, CA) and quantified the concentrations on a Nanodrop 2000 (Waltham, Massachusetts, USA). We included deionized water controls in each extraction as a quality control for contamination.

We amplified samples at 14 polymorphic nuclear microsatellite loci (hereafter nSSRs) that were developed specifically for analysis of *G. paniculata* using Illumina sequencing technology (Table 2.2) (Leimbach-Maus et al. in prep). We conducted polymerase chain reactions (PCR) using a forward primer with a 5'-fluorescent labeled dye (6-FAM, VIC, NED, or PET) and an unlabeled reverse primer. PCR reactions consisted of 1x KCl buffer, 2.0-2.5 mM $MgCl_2$ depending on the locus, 300 µM dNTP, 0.08 mg/mL BSA, 0.4 µM forward primer fluorescently labeled with either FAM, VIC, NED, or PET, 0.4 µM reverse primer, 0.25 units of Taq polymerase, and a minimum of 50 ng DNA template, all in a 10.0 µL reaction volume (Leimbach-Maus et al. in prep). The thermal cycle profile consisted of denaturation at 94˚C for 5 minutes followed by 35 cycles of 94˚C for 1 minute, annealing at 62˚C for 1 min, extension at 72˚C for 1 min, and a final elongation step of 72˚C for 10 minutes.

Each sample was also amplified at 2 universal chloroplast microsatellite loci (hereafter cpSSRs) previously developed for *Nicotiana tabacum* L. (Chung and Staub 2003) (ccssr4, ccssr9) (Table 2.2). PCR reaction details and fragment lengths from Calistri et al. (2014) were used for *G. paniculata*. PCR reactions were conducted using a forward primer with a 5'-fluorescent labeled dye (Applied Biosystems, Foster City, CA) and an unlabeled reverse primer. PCR reactions for the cpSSRs are the same as detailed above for the nuclear loci. The thermal cycler profile for cpSSRs is as follows: denaturation at 94˚C for 5 minutes followed by 30 cycles of 94˚C for 1 minute, annealing at 52˚C for 1 minute, extension at 72˚C for 1 minute, and a final elongation step of 72˚C for 8 minutes.

We determined successful amplification by visualizing the amplicons on a 2% agarose gel stained with ethidium bromide. We multiplexed PCR amplicons according to dye color and allele size range (Table 2.2), added LIZ Genescan 500 size standard, denatured with Hi-Di Formamide at 94˚C for four minutes, and then performed fragment analysis on an ABI3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA) following instrument protocols. We genotyped individuals using the automatic binning procedure on Genemapper v5 (Applied Biosystems, Foster City, CA), and constructed bins following the Genemapper default settings. To account for the risk of genotyping error when relying on an automated allele-calling procedure, we visually inspected that all individuals at all loci were correctly binned to minimize errors caused by stuttering, low heterozygote peak height ratios, and split peaks (DeWoody et al. 2006, Guichoux et al. 2011).

Quality control

Prior to any analysis, we used multiple approaches to check for scoring errors (DeWoody et al. 2006). We checked nSSR genotypes for null alleles and potential scoring errors due to stuttering and large allele dropout using the software Micro-Checker v2.2.3 (Van Oosterhout et al. 2004, Van Oosterhout et al. 2006). Prior to marker selection, the loci used in this study were previously checked for linkage disequilibrium (Leimbach-Maus et al. in prep). We checked for heterozygote deficiencies in the package STRATAG in the R statistical program, screening our data for individuals with more than 20% missing loci, and loci with more than 10% missing individuals (Gomes et al. 1999; Archer et al. 2016). We found none, so all individuals and loci remained for further analyses. In addition, we genotyped 95 individuals twice to ensure consistent allele calls. For the nSSR dataset, we used Genepop 4.2 (Raymond and Rousset 1995, Rousset 2008) to perform an exact test of Hardy-Weinberg Equilibrium (HWE) with 1000 batches of 1000 Markov Chain Monte Carlo iterations (Gomes et al. 1999). We also checked for loci out of HWE for more than 60% of the populations; however, there were none.

nSSR genetic diversity

We calculated the total number of alleles per sampling location, private alleles, observed and expected heterozygosity in GenAlEx 6.502 (Peakall and Smouse 2006, 2012), and estimated the inbreeding coefficient (F_{IS}) in Genepop 4.2 (Raymond and Rousset 1995, Rousset 2008). We used the package diverSity in the R statistical program to calculate the allelic richness at each sampling location (Keenan et al. 2013).

nSSR genetic structure

To test for genetic differentiation between all pairs of sampling locations, we calculated Weir and Cockerham's (1984) pairwise F_{ST} values for 9999 permutations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012). In the R statistical program, we corrected the p-values using a false discovery rate (FDR) correction (Benjamini and Hochberg 1995). To test how much of the genetic variance can be explained by within and between population variation, we ran an analysis of molecular variance (AMOVA) for 9999 permutations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012).

To examine the number of genetic clusters among our sampling locations, we used the Bayesian clustering program STRUCTURE v2.3.2 (Pritchard et al. 2000). Individuals were clustered assuming the admixture model both with and without a priori sampling location for a burnin length of 100,000 before 1,000,000 repetitions of MCMC for 10 iterations at each value of *K* (1 – 16). The default settings were used for all other parameters. We identified the most likely value of *K* using the Ln Pr(X|K) from the STRUCTURE output and the ΔK method from Evanno et al. (2005) in CLUMPAK (Kopelman et al. 2015). This STRUCTURE analysis of all 12 populations identified strong genetic structure patterns. To test for more subtle population structure that may be present within each initially identified cluster, we ran a STRUCTURE analysis for the populations included in each of these clusters. These next two analyses followed the same parameters as the first STRUCTURE run, assuming the admixture model without a priori sampling location for a burnin length of 100,000 before 1,000,000 repetitions of MCMC for 10 iterations at each value of *K* (1–4 and 1–11, respectively). The default settings were used for all other parameters. We identified the most likely value of *K* using the Ln Pr(X|K) from the

STRUCTURE output and the ΔK method from Evanno et al. (2005) in CLUMPAK (Kopelman et al. 2015).

To further explore the genetic structure of these populations, we ran a Principal Coordinates Analysis (PCoA) in GenAlEx 6.502, where the analysis was based on an individual pairwise genotypic distance matrix (Peakall et al. 1995, Smouse and Peakall 1999). To find and describe finer genetic structuring of the nSSR dataset, we performed a discriminant analysis of principal components (DAPC) in the R package *adegenet*, which optimizes among-group variance and minimizes within-group variance (Jombart 2008, Jombart et al. 2010). To identify the number of clusters for the analysis, a Bayesian clustering algorithm was run for values of *K* clusters (1 – 16). We retained a *K*-value of 3 to explore any substructuring of the nSSR data. DAPC can be beneficial, as it can limit the number of principal components (PCs) used in the analysis. It has been shown that retaining too many PCs can lead to over-fitting and instability in the membership probabilities returned by the method (Jombart et al. 2010). Therefore, we performed the cross-validation function to identify the optimal number of PCs to retain.

To assess the effect of isolation by distance (IBD), we used a paired Mantel test based on a distance matrix of Slatkin's transformed F_{ST} ($D = F_{ST}/(1 - F_{ST})$) (Slatkin 1995) and a geographic distance matrix for 9999 permutations in GenAlEx 6.502, and the analysis follows Smouse et al. (1986) and Smouse and Long (1992). The mean geographic center was generated for each sampling location in ArcGIS software (ESRI™ 10.4.1), and the latitude and longitude of these points was then used to construct a matrix of straight line distances in km between each sampling location. The reported p-values are based on a one-sided alternative hypothesis (H_1: $R > 0$). A Mantel test was run for all sampling locations together, and a test was also run separately for populations within each cluster identified in the STRUCTURE analysis.

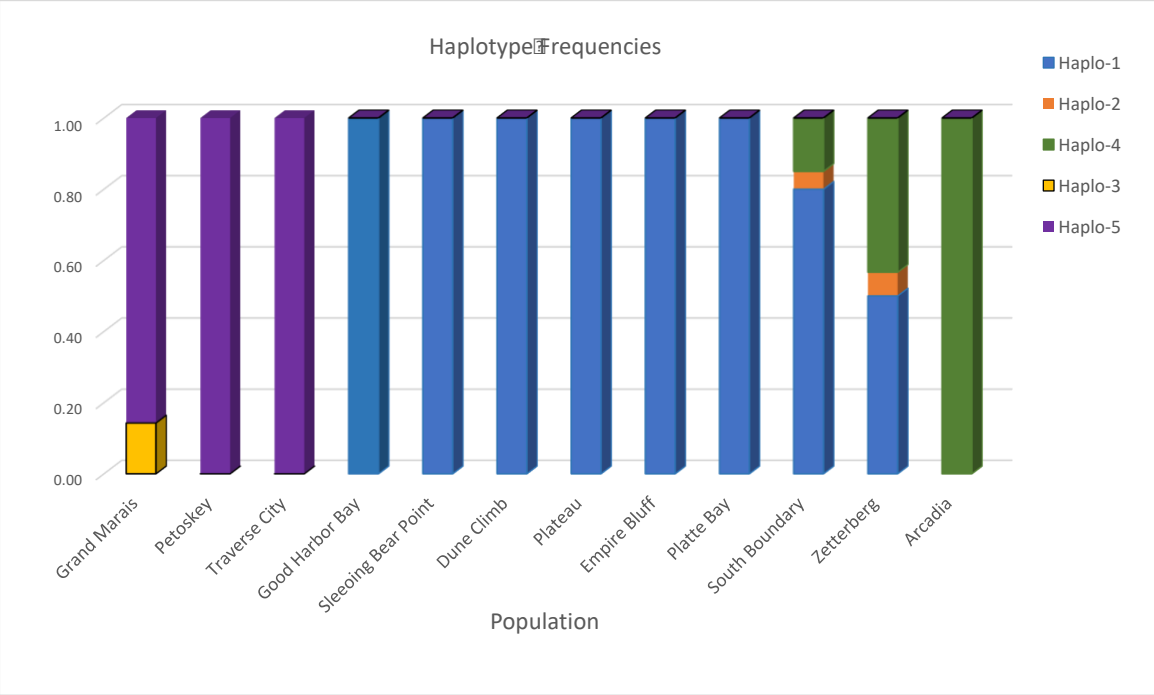
cpSSR genetic diversity

For the cpSSR dataset, we used the program HAPLOTYPE ANALYSIS v1.05 (Eliades and Eliades 2009) to calculate the number of haplotypes, haplotype richness, private haplotypes, haploid diversity. To visualize patterns in the cpSSR dataset, we created a minimum spanning network in the R package poppr (Kamvar et al. 2014). Nei's genetic distance was used as the basis for the network with a random seed of 9,999.

nSSR and cpSSR genetic structure

In order to compare the population structure of the nSSR and cpSSR data, we used the Φ_{ST} distance matrix for both datasets and ran an AMOVA. The population pairwise Φ_{ST} matrix facilitates comparison of molecular variance between codominant and dominant data by suppressing within individual variation, thus allowing for the comparison between varying mutation rates (Weir and Cockerham 1984, Excoffier et al. 1992). To test how much genetic variation could be explained by within populations, between populations, and between regions (genetic clusters identified through STRUCTURE analysis) for both datasets, we ran an AMOVA for 9, 999 iterations in GenAlEx 6.502 (Peakall and Smouse 2006, 2012).

Supplemental Table



Supplemental Figure F Haplotype frequencies across 12 populations of baby's breath calculated in program HAPLOTYPE NETWORK (Eliades and Eliades 2009).

References

Albert, D. A. 2000. *Borne of the Wind, Michigan Sand Dunes*.

Arbogast, A.F. 2015. *The Emerging Science of Coastal Sand Dune Age and Dynamics : Implications for Regulation and Risk Management in Michigan*.

Arbogast, A.F. and W.L. Loope. 1999. Maximum-limiting ages of Lake Michigan coastal dunes: Their correlation with Holocene lake level history. *Journal of Great Lakes Research* 25(2): 372-382.

Archer, F.I., Adams, P.E., and B.B. Schneiders. 2016. STRATAG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources*. 17: 5–11.

Baker, H. and G.L. Stebbins. 1965. Characteristics and modes of origin of weeds. In H. Baker & G. L. Stebbins (Eds.), *The Genetics of Colonizing Species* (pp. 147–168). New York, NY: Academic Press.

Barrett, S.C.H. 2015. Foundations of invasion genetics: The Baker and Stebbins legacy. *Molecular Ecology* 24(9): 1927–1941.

Baskett, C.A., Emery, S.M., and J.A. Rudgers. 2011. Pollinator Visits to Threatened Species Are Restored Following Invasive Plant Removal. *International Journal of Plant Sciences* 172(3): 411–422.

Benjamini, Y. and Y. Hochberg. 1995. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. Journal of the Royal Statistical Society, Series B. 57: 289 - 300. doi: 10.2307/2346101

Blossey, B. and R. Notzold. 1995. Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. *Journal of Ecology* 83: 887–889.

Blumer, B.E., Arbogast, A.F., and S.L. Forman. 2012. The OSL chronology of eolian sand deposition in a perched dune field along the northwestern shore of Lower Michigan *Quaternary Research* 77(3): 445 – 455.

Bock, D.G., Caseys, C., Cousens, R.D., Hahn, M.A., Heredia, S.M., Hübner, S., Turner, K.G., Whitney, K.D., and L.H. Rieseberg. 2015. What we still don't know about invasion genetics. *Molecular Ecology* 24(9): 2277–2297.

Calistri, E., Buiatti, M., and P. Bogani. 2014. Characterization of *Gypsophila* species and commercial hybrids with nuclear whole-genome and cytoplasmic molecular markers. *Plant Biosystems* 150(1): 1 – 11.

Chown, S.L., Hodgins, K.A., Griffin, P.C., Oakeshott, J.G., Byrne, M., and A.A. Hoffmann. 2015. Biological invasions, climate change and genomics. *Evolutionary Applications* 8(1): 23–46.

Chung S-M. and J.E. Staub. 2003. The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theoretical and Applied Genetics*. 107: 757–767. DOI 10.1007/s00122-003-1311-3

Cochard, R., Ranamukhaarachchi, S.L., Shivakoti, G.P., Shipin, O.V., Edwards, P.J., and K. T. Seeland. 2008. The 2004 tsunami in Aceh and Southern Thailand: a review on coastal ecosystems, wave hazards and vulnerability. *Perspectives in Plant Ecology Evolution and Systematics* 10:3–40.

Cowles, H.C. 1899. The Ecological Relations of the Vegetation on the Sand Dunes of Lake Michigan. Part I.-Geographical Relations of the Dune Floras. *Botanical Gazette*. 27 (2): 95 – 117.

Darwent, A. L. 1975. The Biology of Canadian Weeds. 14. Gypsophila paniculata L. *Canadian Journal of Plant Science* 55(4): 1049–1058.

Darwent, A. L. and R.T. Coupland. 1966. Life History of Gypsophila paniculata. *Weed Science Society of America* 14(4): 313–318.

Dewoody, J., Nason, J. D., and V.D. Hipkins. 2006. Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*. 6(4), 951–957.

DiTomaso, J., and G.B. Kyser. 2013. *Weed Control in Natural Areas in the Western United States*. UC Davis Weed Research and Information Center. Website https://www.cal-ipc.org/resources/library/publications/weedcontrol/ [accessed 10 October 2016]

Eliades N-G. and D.G. Eliades. 2009. HAPLOTYPE ANALYSIS: Software for analysis of haplotype data. (available from http://www.uni-goettingen.de/en/134935.html)

Elton, C. S. 1958. *The ecology of invasions by animals and plants*. Boston, MA: Springer.

Emery, S.M., Doran, P.J., Legge, J.T., Kleitch, M., and S. Howard. 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.

Emery, S.M., and P.J. Doran. 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.

Environmental Systems Research Institute (ESRI). 2012. ArcGIS Desktop: Release 10.4.1. Redlands, CA.

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

Everard, M., Jones, L., and B. Watts. 2010. Have we neglected the societal importance of sand dunes? An ecosystem services perspective. *Aquatic Conservation: Marine and Freshwater Ecosystems* 20(4): 476–487.

Excoffier L., Smouse, P.E., and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites. *Genetics*. 131: 479-491.

Faircloth, B.C. 2008. Msatcommander: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8: 92–94.

French, K., Mason, T.J., and N. Sullivan. 2011. Recruitment limitation of native species in invaded coastal dune communities. *Plant Ecology* 212(4): 601–609.

Gleason, H.A., and A. Cronquist. 1963. Manual of vascular plants of northeastern United States and adjacent Canada. D. Van Nostrand Company, New York, New York.

Gomes I, Collins A, Lonjou C *et al.* 1999. Hardy–Weinberg quality control. *Annals of Human Genetics* 63: 535–538.

Guichoux, E., Lagache L., Wagner S., Chaumeil P., Léger P., Lepais O., … R.J. Petit. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11(4): 591–611.

Hale M.L., Burg T.M. and T.E. Steeves. 2012. Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. *PLoS ONE* 7(9): e45170.

Halevy, A.H. 1999. Ornamentals: Where diversity is king—the Israeli experience. p. 404–406. In: J. Janick (ed.), *Perspectives on new crops and new uses.* ASHS Press, Alexandria, VA.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403 – 1405.

Jombart, T., Devillard, S. and F. Balloux. 2010. Discriminant analysis of principal components: Anew method for the analysis of genetically structured populations. *BMC Genetics* 11(1): 94.

Kamvar, Z.N., Tabima, J.F., and N.J. Grünwald. 2014. Poppr: An R Package For Genetic Analysis Of Populations With Clonal, Partially Clonal, And/or Sexual Reproduction. *PeerJ.* 2: e281.

Karamanski, T.J. 2000. A nationalized lakeshore: the creation and administration of Sleeping Bear Dunes National Lakeshore. *National Park Service, Department of the Interior*. Washington, DC.

Keane, R. and M. Crawley. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution* 17: 164–170.

Keenan K., McGinnity P., Cross T.F., Crozier W.W., Prodöhl, P.A., and R.B. O'Hara. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* 4: 782-788.

Kopelman, N.M., Mayzel J., Jakobsson, M., Rosenberg, N.A., and I. Mayrose. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15(5): 1179-1191.

Lawson – Handley, L.J., Estoup, A., Evans, D.M., Thomas, C.E., Lombaert, E., Facon, B., … and H.E. Roy. 2011. Ecological genetics of invasive alien species. *BioControl* 56(4): 409–428.

Leicht – Young, S.A., Pavlovic, N.B., Grundel, R., and K.J. Frohnapple. 2009. A comparison of seed banks across a sand dune successional gradient at Lake Michigan dunes (Indiana, USA). *Plant Ecology* 202(2): 299–308.

Leimbach-Maus H, Parks S, and C. Partridge. Microsatellite primer development for invasive perennial herb, *Gypsophila paniculata* (Caryophyllaceae). Manuscript in preparation.

MI DEQ. 2017. *Sand Dune Mining Site Status Summary*. Website https://www.michigan.gov/deq/0,4561,7-135-3311_4114-359904--,00.html [accessed 01 November 2017].

Michigan, Emmet Co.: Pellston, 16 July 1913, *H.A. Gleason s.n.* (MICH).

Moles, A.T., Gruber, M.A., and S.P. Bonser. 2008. A new framework for predicting invasive plant species. *Journal of Ecology* 96: 13–17.

Paterson, A.H., Schertz, K.F., Lin, Y.R., Liu, S.C., and Y.L. Chang. 1995. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92(13): 6127–6131.

Peakall, R. et al. 1995. Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss Buchloe dactyloides. *Molecular Ecology* 4: 135-147.

Peakall, R. and Smouse P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537 – 2539.

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

Piya, S., Nepal, M.P., Butler, J.L., Larson, G.E., and A. Neupane. 2014. Genetic diversity and population structure of sickleweed (*Falcaria vulgaris*; Apiaceae) in the upper Midwest USA. *Biological Invasions* 16(10): 2115–2125.

Pritchard, J. K., Stephens, M., and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics 155*(2): 945–959.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 4.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Robertson, A., and W.G. Hill. 1984. Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics* 107: 703–718.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145 (4): 1219 – 1228.

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

Rowland, J., and M.A. Maun. 2001. Restoration ecology of an endangered plant species: Establishment of new populations of *Cirsium pitcheri*. *Restoration Ecology* 9(1): 60–70.

Sakai, A. K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., … and S.G. Weller. 2001. The population biology of invasive species. *Annual Review Ecology and Systematics* 32: 305–332.

Sakata, Y., Itami, J., Isagi, Y., and T. Ohgushi. 2015. Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range. *Journal of Plant Research* 128(6): 909–921.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W.,…and D.G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539.

Smouse, P.E. and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82: 561-573.

Stabile, J., Lipus, D., Maceda, L., Maltz, M., Roy, N., and I. Wirgin. 2016. Microsatellite DNA analysis of spatial and temporal population structuring of *Phragmites australis* along the Hudson River Estuary. *Biological Invasions* 18(9): 2517–2529.

Stout, J.C., Duffy, K.J., Egan, P., Harbourne, M., and T.R. Hodkinson. 2015. Genetic diversity and floral width variation in introduced and native populations of a long-lived woody perennial. *AoB PLANTS* 7: 1–11.

The Nature Conservancy, and partners. 2013. Lake Michigan Coastal Dunes Restoration Report. The Nature Conservancy. Website https://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/michigan/lake-michigan-dune-report-2013.pdf [accessed 01 November 2017].

USDA – APHIS internal report. 2012. Submission to the U.S. Department of Agriculture Animal and Plant Health Inspection Service Concerning the Status of Genetically Engineered Gypsophila paniculata Cut Flowers Under 7 CFR Part 340. Website https://www.aphis.usda.gov/biotechnology/downloads/reg_loi/Danziger_USDA_012012.pdf [accessed 15 May 2018].

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources* 4: 535–538.

Van Oosterhout, C., Weetman, D. and W.F. Hutchinson. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Resources* 6: 255–256.

Vettori, L., Schiff, S., Tani, C., Pasqualetto, P., and A. Bennici. 2013. Morphological and cytological observations of wild species and hybrids of *Gypsophila*. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 149(2): 322 – 328.

Weir, B. and C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38(6): 1358-1370.

Whitney, K.D., and E. Gering. 2015. Five decades of invasion genetics. *New Phytologist* 205(2): 472–475.

Wilson, J.R.U., Dormontt, E.E., Prentis, P.J., Lowe, A.J., and D.M. Richardson. 2009. Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution* 24(3): 136 – 144.

Wolfe, L. M. 2002. Why alien invaders succeed: support for the escape-from-enemy hypothesis. *The American Naturalist* 160(6): 705–11.

Zvi, M.B., Zuker, M., Ovadis, A., Shklarman, M., Ben-Meir, E., Zenvirt, H., and A. Vainstein. 2008. Agrobacterium-mediated transformation of gypsophila (*Gypsophila paniculata* L.). *Molecular Breeding* 22(4): 543–553.