Biological Sciences Faculty Publications   Biological Sciences

2020

Small-Scale Population Connectivity and Genetic Structure in Canada Thistle *(Cirsium arvense)*

Mahboubeh Hosseinalizadeh Nobarinezhad

Lavanya Challagundla

Lisa E. Wallace
Old Dominion University, lewallac@odu.edu

SMALL-SCALE POPULATION CONNECTIVITY AND GENETIC STRUCTURE IN CANADA THISTLE (*CIRSIUM ARVENSE*)

Mahboubeh Hosseinalizadeh Nobarinezhad,[1,]* Lavanya Challagundla,† and Lisa E. Wallace[2,]‡

*Department of Biological Sciences, Mississippi State University, Mississippi State, Mississippi 39762, USA; †Department of Data Science, John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA; and ‡Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA

Editor: Oscar Rocha

Premise of research.   Population connectivity, the exchange of genes among geographically separated subpopulations, is thought to be a key process for the maintenance of genetic diversity and the survival of invasive species in newly colonized areas. Plant populations' degree of genetic connectivity, which occurs via pollen and seed dispersal, leads to different degrees of genetic admixture and genetic structure. Environmental barriers and differential selection pressures that are variable across time and space tend to alter genetic structure within and among populations via restriction or facilitation of gene flow. Canada thistle, an invasive species of the United States and Canada, is well known for production of high numbers of seeds, asexual reproduction, and wide environmental tolerance. These factors may influence its success as an invader by facilitating population persistence.

Methodology.   In this study we evaluated genetic connectivity of 12 Canada thistle populations across a 75-km area using 10 microsatellite loci, estimated the spatial scale of genetic exchange between populations, and tested for an association between genetic structure and variation in landscape characteristics.

Pivotal results.   All loci were highly polymorphic within populations, and populations were significantly differentiated from one another ($F_{ST} = 0.21$, $p = 0.001$), but environmental, geographic, and climatic factors were found to have little explanatory power for the observed genetic structure. Bayesian clustering analysis suggested the presence of two distinct genetic groups and admixture in several populations.

Conclusions.   We conclude that, for this species, genetic admixture and co-occurrence of genetically distinct types may play an important role in rapid adjustment to diverse environments and persistence of populations across the landscape.

Keywords: Canada thistle, gene flow, genetic structure, landscape genetics, microsatellites.

Online enhancement: supplementary table.

Introduction

Knowledge of an invasive species' capacity for dispersal, colonization, and range expansion is of considerable importance for understanding how it will respond to biological and physical controls and for preventing further spread across the landscape (Sakai et al. 2001; Hoffmeister et al. 2005; Lawson Handley et al. 2011). For example, by altering population genetic structure through the interruption or facilitation of gene flow among populations, one can potentially reduce adaptive variation and introduce maladaptive genes into populations (Sakai et al. 2001). Highly diverse populations, which are hypothesized to harbor adaptive potential to respond to future environmental changes,

can be targeted for extirpation. Because population connectivity is dictated by gene dispersal, which is strongly influenced by the complexity of the landscape, consideration of landscape variables is important for fully understanding population structure (Manel et al. 2003; Storfer et al. 2007). Landscape genetic approaches provide a context for understanding dispersal patterns by considering population genetic structure within the context of landscape features (Manel et al. 2003; Guillot et al. 2005; Holderegger and Wagner 2006). By studying genetic and geographic variation in a spatially explicit framework, landscape genetic studies can contribute to understanding metapopulation dynamics, the barriers to gene flow, and the scale of gene flow. Such studies can also be useful for predicting impacts of management regimes on control and spread of invasive species (Storfer et al. 2007).

Canada thistle (*Cirsium arvense* (L.) Scop.) is a dioecious diploid (2n = 34) perennial species of Asteraceae that was introduced into Canada (Quebec and Ontario) by French settlers via impure crop seeds in the early seventeenth century (Stevens

---

[1] Author for correspondence; email: mh2329@msstate.edu.
[2] Author for correspondence; email: lewallac@odu.edu.

1847; Hodgson 1968). It is also believed that this species was independently introduced into Canada by French settlers and into New England by English and Dutch settlers (Hansen 1918). Indeed, Guggisberg et al. (2012) confirmed the presence of distinct gene pools consistent with multiple introductions to North America, first from western Europe and later from eastern Europe. These patterns of introduction are associated with the increase in agriculture in the Midwest. From Canada, Canada thistle was then likely propagated southward and westward through human activities (Hansen 1918; Evans 2002). Since its arrival in North America, Canada thistle has spread extensively. By 1899 it reached the Pacific coast and occupied nearly every North American herbaceous community (Dewey 1901; Hodgson 1968; Evans 2002). Because of weed legislation in 21 states in the United States, Canada thistle was declared noxious in 1896 (Guggisberg et al. 2012).

Invasiveness of Canada thistle is linked to its production of copious amounts of seed, asexual reproduction, high degree of morphological plasticity, and tolerance for a wide array of habitats (Moore 1975; Donald 1994; Heimann and Cussans 1996). The most common pollinator of *C. arvense* is the honeybee (*Apis mellifera*; Theis 2006). The seeds are 4–5 mm long, with a feathery pappus that assists in wind dispersal (Kay 1985; Blamey and Grey-Wilson 1989). Under ideal conditions, an individual plant can produce up to 100 heads per shoot and one to five flower heads per branch. Average seed production per plant has been estimated at 1530, but more seeds are expected when male and female plants are interspersed (Moore 1975). The high fecundity allows this species to be reintroduced frequently, which is a key component for invasive species to become established (Nadeau and Vanden Born 1989; Donald 1994). This species also reproduces asexually through root buds that form adventitiously on the thickened roots and give rise to new shoots when stems are buried or cut via mowing (Donald 1994).

Canada thistle is shade intolerant and frequently invades open disturbed sites such as pastures, ditches, and bottomlands in the Midwest and Great Plains and riparian areas in the intermountain West (Nuzzo 1997; Hoffmann et al. 2008). Previous studies have identified a positive association for the presence of Canada thistle with elevation (47.9 m) and soil clay content (5.4%; Hoffman et al. 2008). The species has a substantial ecological impact on natural and agricultural ecosystems because it crowds out native species, reduces crop and forage yields, and potentially hybridizes with native congeneric species (Nuzzo 1997). Although some studies (Tipping 1993; Louda and Potvin 1995; Kluth et al. 2001; Eber and Brandl 2003; Skuhrovec et al. 2008; Abela-Hofbauerova et al. 2011) demonstrated a negative effect on plant growth and reproduction from insect herbivory, biological control has not been highly successful at the population level, perhaps because the species is not limited by natural enemies in its home range (McClay 2002).

Because of the clonal nature of Canada thistle, it was long thought that stands comprised relatively few clones and that seed dispersal and/or establishment rarely occurred (reviewed in Moore 1975). Additionally, the majority of seeds produced through sexual reproduction may not move very far from the maternal plant (Becker et al. 2008). The reproductive strategy of Canada thistle may tend to keep achenes local, and the function of the pappus may be to protect achenes from dampness rather than aid in dispersal (Dandeno 1905; Sheldon and Burrows

1973). It is also believed that the dioecious nature of this species guarantees pollen-limited seed production, and distance to nearest pollen donor correlates negatively with fertilization (Lalonde and Roitberg 1994). Nevertheless, studies of genetic structure in the introduced and native ranges of Canada thistle (Hettwer and Gerowitt 2004; Solé et al. 2004; Bodo-Slotta et al. 2006, 2010; Guggisberg et al. 2012) have found high levels of genetic variation within populations. For example, Solé et al. (2004) reported a mean population genotypic diversity of 0.73 on the basis of amplified fragment length polymorphisms (AFLP) of European populations and found that 86% of the sampled plants contained a unique genotype. Bodo-Slotta et al. (2006) reported total genetic diversity (H_t) of 0.24 using four microsatellites in invasive populations from North Dakota, and Bodo-Slotta et al. (2010) reported total diversity of 0.183 for invasive populations across North America using seven microsatellite loci. Also using microsatellite loci, Guggisberg et al. (2012) reported mean expected heterozygosity (H_E) of 0.76 for six loci in North American populations and suggested that admixture of distinct introductory gene pools has contributed to the maintenance of diversity in the invasive range. Thus, maintenance of genetic variation within populations and presence of gene flow across populations through sexual reproduction may contribute to the persistence of invasive genotypes of Canada thistle in North America. Identification of patterns of genetic diversity, and specifically identification of highly successful genotypes, could be used to prevent the formation and spread of novel genotypes through sexual reproduction (Ward et al. 2008; Bommarco et al. 2010).

In this study, genetic connectivity across populations of Canada thistle was evaluated over a small geographic area to (1) determine whether genetic diversity at small spatial scales is a subset of that found at regional and continental scales, (2) determine genetic similarity among populations arranged in a linear pattern along roadsides, (3) estimate the extent of genetic exchange between populations arranged in this manner, and (4) determine whether genetic structure is associated with variation in landscape characteristics. We predicted that genetic relatedness, an indicator of gene flow, is negatively correlated with geographic distance. We also predicted that one or more landscape variables (i.e., latitude, longitude, elevation, slope, monthly precipitation, and distance from a primary road) are significantly associated with the pattern of genetic structure in the study area. The landscape variables were chosen because they are expected to reflect variation in factors that could influence dispersal, colonization, and range expansion of this species in North America. The results of this study quantify potential genetic discontinuities of Canada thistle across its invasive range at a smaller spatial scale than has been addressed in previous studies of multiple states (Bodo-Slotta et al. 2006) and across North America (Bodo-Slotta et al. 2010; Guggisberg et al. 2012). The results of this study also promote our understanding of microevolutionary processes that could generate genetic structure across natural landscapes.

Material and Methods

Leaf samples (16–24 per site; table 1) and voucher specimens (one per site) were collected from 12 sites along the Lewis and Clark Trail (US Route 93) and State Highway 75 in Idaho across a total distance of 75 km (fig. 1) in July 2007. Sites were selected

Table 1

Genetic Diversity of Populations of *Cirsium arvense*

Population	N	G	A	$\%P$	H_O	H_E	F_{IS}
1	24	20	4.1	100	.575	.512	−.119
2	24	22	4.9	100	.618	.574	−.095
3	16	7	2.0	90	.774	.418	−.789
4	24	22	4.8	100	.696	.619	−.102
5	22	22	4.9	100	.787	.658	−.216
6	24	18	3.7	100	.633	.542	−.184
7	20	20	4.0	100	.720	.561	−.280
8	24	20	3.4	90	.661	.470	−.405
9	24	22	4.7	100	.877	.656	−.365
10	24	22	4.7	100	.664	.629	−.048
11	24	22	3.9	100	.726	.604	−.184
12	24	12	2.1	60	.453	.286	−.499
Mean	21.8	19.08	3.9	95	.682	.544	−.260

Note. N = number of individuals sampled; G = number of unique multilocus genotypes; A = mean number of alleles per locus; $\%P$ = percentage of polymorphic loci; H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient.

if at least 50 stems of Canada thistle were present; areas between sampled sites contained few Canada thistle plants. At each collection site, individuals were haphazardly sampled at least 2 m apart to reduce the likelihood of sampling clonal ramets. The range and mean distance separating pairs of populations were 1.4–75.2 km and 23 km, respectively. After a maximum of 24 plants were collected, sampling ceased. Most plants were growing within 5 m of the road edge. GPS coordinates were recorded at the center of each sampling site. Leaf samples were stored on ice in the field and then at −80°C in the lab before DNA was extracted. Voucher specimens are deposited in the Mississippi State University herbarium (appendix). DNA was extracted from leaf samples
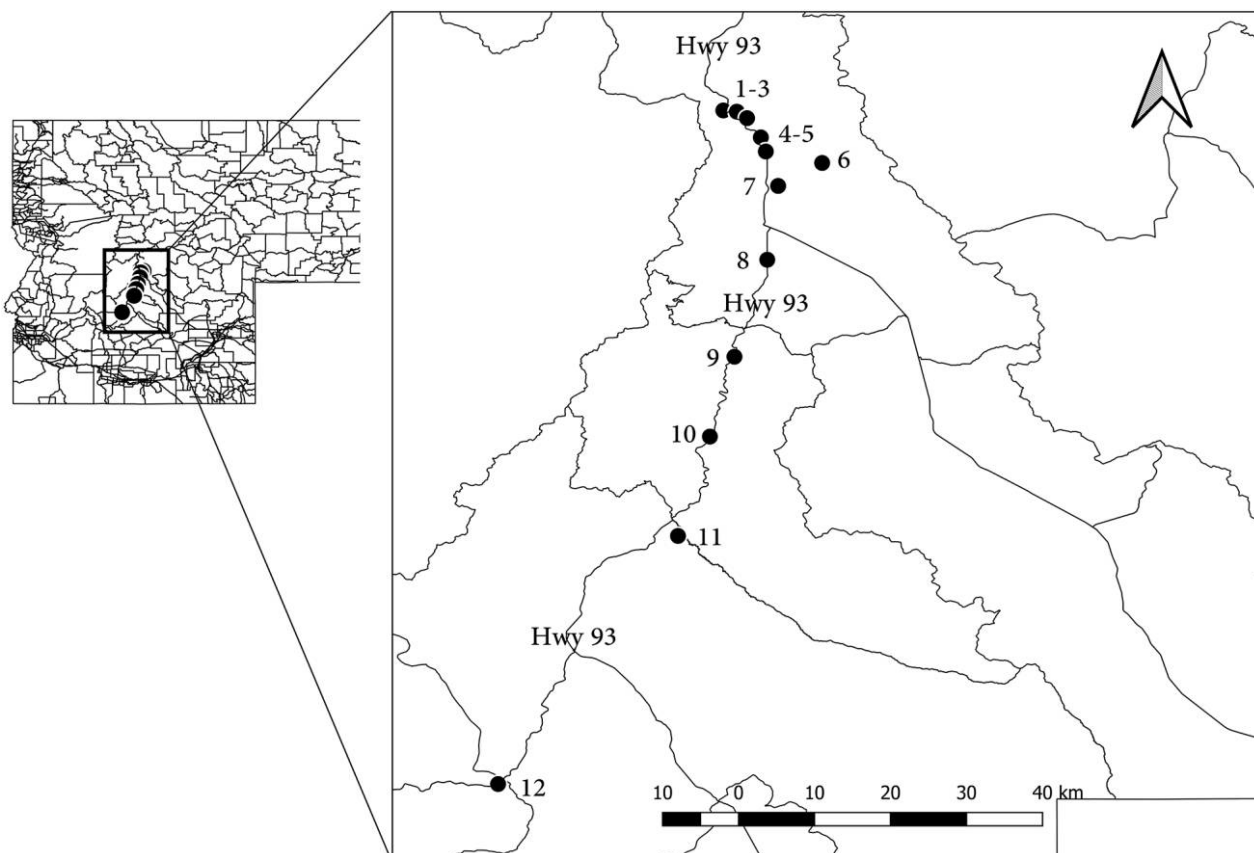

Fig. 1 Sampling locations of populations of Canada thistle included in this study.

using the DNEasy Plant Mini Kit (Qiagen, Germantown, MD). All individuals were genotyped using primers from 11 microsatellite loci developed for Canada thistle or other *Cirsium* species (Jump et al. 2002; Bodo-Slotta et al. 2005). To permit multiplexed genotyping of multiple loci, one primer for each locus contained a tailing sequence (5′AGGAAACAGCTATGACCAT 3′) that matched a fluorescently labeled primer (6-FAM or HEX) with the same sequence. Single-locus reactions were conducted for each sample and contained 1.0 μL template DNA, 1X GoTaq Flexi Buffer (Promega, Madison, WI), 160 μM of each dNTP (Promega), 1.5–2 mM $MgCl_2$ (primers Caca05, Caca10, Caca22, C11, C120, C128 required 1.5 mM; primer C101 required 2 mM; and primers Caca01, Caca04, Caca07, D117 required 2.5 mM), 0.3 μM locus-specific primer, 0.3 μM fluorescent-labeled primer, 0.1–0.2 μM tailed locus-specific primer (primers Caca04, Caca07, Caca10, Caca22, C128 required 0.1 μM; primers Caca01, Caca05, C11, C101, C120, D117 required 0.2 μM), and 1 unit of GoTaq DNA polymerase (Promega). Fragments were amplified using the following thermal cycling parameters: denaturation at 94°C for 3 min, 35 cycles at 94°C for 40 s, at appropriate annealing temperature for 40 s (45°C for primer D117; 50°C for primer C11; 52°C for primers Caca05 and C128; 55°C for primers Caca04, Caca22, C101, C120; 60°C for primer Caca10; 64°C for primer Caca01 and Caca07), and at 72°C for 30 s, followed by an elongation step at 72°C for 4 min. Fragments from distinct loci that were of different sizes and did not contain overlapping fluorescent tags were combined, diluted, and mixed with 0.4 μL LIZ-600 internal size standard (Thermo Fisher Scientific, Grand Island, NY) before being run on a capillary electrophoresis instrument at the Arizona State University DNA Lab. GeneMarker software (SoftGenetics, State College, PA) was used to assign allele sizes for each locus.

The presence of null alleles for each locus and population was evaluated with MICRO-CHECKER 2.2.3 (van Oosterhout et al. 2004). Genetic diversity within populations was assessed in terms of number of alleles per locus, percent polymorphic loci (%P), observed heterozygosity (H_O), and expected heterozygosity (H_E) using GenAlEx version 6.503 (Peakall and Smouse 2012; table 1). Frequency of private alleles was detected using GenAlEx. The number of unique multilocus genotypes for each population was calculated by Poppr, an R package (Kamvar et al. 2014) that is specifically designed for genetic analysis of populations with clonal or sexual reproduction, using RStudio statistical software version 1.1.456 (RStudio Team 2012). Deviation from Hardy-Weinberg equilibrium (HWE) was tested by permuting alleles among individuals using GenAlEx. Statistical significance of differences between H_E and H_O was assessed using a sequential Bonferroni-corrected (Holm 1979) p value (0.0004) for multiple comparisons. A global test of deviation across all loci was also conducted using Fisher's (1954) method in GENEPOP version 3.2 (Raymond and Rousset 1995; Rousset 2008). Genotypic linkage disequilibrium for each pair of loci was estimated through log likelihood ratio statistic (G-test) in GENEPOP version 3.2 (Raymond and Rousset 1995; Rousset 2008) and the application of a Bonferroni-corrected (Holm 1979) p value of 0.0011 to test for significance. Genetic differentiation among populations was estimated by calculating multilocus F_{ST} based on the approach of Weir and Cockerham (1984) and using GENEPOP version 3.2 (Raymond and Rousset 1995; Rousset 2008). An analysis of molecular variance (AMOVA; Excoffier et al. 1992) was conducted using GenAlEx (Peakall and Smouse 2012), which allows the hierarchical partitioning of the variance components among populations and among individuals within populations. Statistical significance was assessed by comparison with 999 permutations of the data.

A Bayesian statistical framework provided by the program STRUCTURE version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009) was used to estimate the number of genetically cohesive groups. STRUCTURE uses a Bayesian Markov chain Monte Carlo (MCMC) method to find the minimum number of distinct genetic clusters in a data set. Run parameters included the admixture model, correlated allele frequencies, sampling location of the individuals as a prior, a burn-in period of 50,000 MCMC replicates, and a sampling period of 100,000 replicates. Twenty iterations of each K value were conducted, and K was evaluated from 1 to 10. Multiple posterior probability values for each K were generated, and the most likely K was evaluated by the ΔK method following Evanno et al. (2005) and using STRUCTURE HARVESTER (Earl and vonHoldt 2012). This method compares the rate of change in lnL between successive Ks and the corresponding variance of lnL of each K (Evanno et al. 2005). We also used Clumpak (Kopelman et al. 2015) to aid in assessing results across multiple independent runs at a single K value. Clumpak identifies sets of highly similar runs, separates groups of runs that represent distinct modes, and ultimately generates a consensus solution for each K value.

Isolation by distance, which is defined as a decrease in the genetic similarity between populations as the geographic distance between them increases, was investigated by comparing pairwise population genetic and geographic distances in a Mantel test (Mantel 1967). A matrix of pairwise F_{ST} between populations was created using option 6 suboption 7 in GENEPOP version 3.2 (Raymond and Rousset 1995; Rousset 2008), and a matrix of pairwise geographic distances between populations was calculated using GenAlEx and a modification of the haversine formula (Sinnott 1984) based on GPS coordinates of the sampled sites. A Mantel test was performed using GENEPOP version 3.2 (Raymond and Rousset 1995; Rousset 2008) option 6 suboption 9, which allows analysis of isolation by distance as described in Rousset (1997). Geographic distances were log transformed, pairwise F-statistics were converted to $F_{ST}/1 - F_{ST}$, minimum geographic distance was set to 1.0, and 1000 permutations were used to assess significance of a positive correlation indicative of isolation by distance.

Data were collected for the following landscape variables: latitude, longitude, elevation, slope, monthly precipitation, and distance from a primary road. Elevation values for each set of coordinates were determined using a web-based tool (http://mapcoordinates.net/en, Vivid Planet Software, Salzburg). Slope values were extracted for each set of coordinates using a digital elevation model, ASTER GDEM (ASTER GDEM Validation Team 2009), from USGS Earth Explorer (http://earthexplorer.usgs.gov) by ArcMap, ArcGIS desktop 10.6.1 (ESRI 2018). Monthly precipitation was extracted for each population in ArcMap, ArcGIS desktop 10.6.1 (ESRI 2018), using data from the WorldClim climate database (Fick and Hijmans 2017). The distance between the set of coordinates and an adjacent point on US Route 93 was measured in QGIS 3.2.3 Bonn (QGIS Development Team 2018) using the measure line tool. Using RStudio version 1.1.456 (RStudio Team 2012), we conducted

principal component analysis (PCOA) on the environmental variables of slope, distance from road, elevation, and precipitation to determine whether collection sites differ in a multivariate analysis. The hierarchical Bayesian method of Foll and Gaggiotti (2006) implemented in GESTE version 2 was used to assess whether landscape features may influence population genetic structure of Canada thistle. In GESTE, the reversible-jump MCMC method was used with 10 pilot runs with 5000 burn-in, and then samples were drawn from a chain that was 50,000 in length and separated by a thinning interval of 50. All combinations of variables were considered, and models were evaluated using estimates of posterior probability and the 95% highest probability density interval. By individually calculating the cumulative probability for each factor, the program compares the importance of each factor in explaining the observed F_{ST} values. GESTE was run under three different scenarios: (1) gradual population expansion using the variables latitude and longitude; (2) geographical suitability using the variables slope, elevation, and the distance between a primary road and each population; and (3) importance of local climatic factors using the variable mean monthly precipitation.

Results

The number of unique multilocus genotypes within a population varied from 7 to 22, with a mean of 19.08 (table 1). Duplicate multilocus genotypes were removed from each population before subsequent analyses. Possible null alleles were detected at 5 loci and in 8 of the 12 populations studied. Five populations were predicted to contain null alleles for locus C11, with expected frequencies of 0.17–0.32. Therefore, we removed this locus from the data set. We also accounted for null alleles at two other loci with predicted null allele frequency greater than 0.2, which is a suggested cutoff for null alleles that can influence analyses of genetic structure (Dakin and Avise 2004). For these loci (i.e., D117 in populations 9 and 11 and Caca05 in population 10), we recoded homozygous genotypes that had been identified by MICRO-CHECKER as having a potential null allele as heterozygous for the observed allele size and a missing allele.

All loci were highly polymorphic within populations (mean %P = 95%), and the number of alleles per locus ranged from 2.0 to 4.9, with a mean value of 3.9 (table 1). Observed and expected heterozygosity ranged from 0.453 to 0.877 and from 0.286 to 0.658, respectively (table 1). The inbreeding coefficient (F_{IS}) varied among populations but was consistently negative, indicating heterozygote excess across all populations (table 1). Private alleles were detected in seven populations (1, 3, 4, 5, 9, 10, and 12), with frequencies ranging from 0.023 to 1. After Bonferroni correction was applied, 33 out of 120 comparison tests showed significant deviation from HWE. The expected number of deviations was six, which was calculated by setting up the alpha threshold at 5% of the total pairs of comparisons. The global test of deviation across all loci indicated a significant excess of heterozygous genotypes ($p < 0.001$). After Bonferroni correction was applied, 118 out of 540 tests of linkage disequilibrium, all in different populations and involving different pairs of loci, were significant.

In accordance with the high levels of variation observed at the 10 loci, significant genetic differentiation was detected among the sampled populations ($F_{ST} = 0.21$, $p = 0.001$). AMOVA indicated that variation was partitioned primarily within populations (82%) rather than among populations (18%; table 2). Bayesian analysis in STRUCTURE based on the highest ΔK value (fig. 2A) suggested the presence of two distinct genetic clusters ($K = 2$). The proportion of membership of each population to each of the two genetic clusters is shown in figure 2B. Cluster 1 contains populations 1, 2, and 4–8, whereas cluster 2 contains populations 3 and 9–12. The mean pairwise F_{ST} is 0.139 within cluster 1, 0.261 within cluster 2, and 0.281 between clusters 1 and 2 (table 4). For populations 4 and 5, which exhibit mixed ancestry, mean pairwise F_{ST} is 0.199 and 0.14, respectively, with other populations in cluster 1 and 0.224 and 0.207, respectively, for other populations in cluster 2.

The Mantel test revealed a positive but not significant correlation between pairwise population genetic and geographic distances ($r = 0.44$, $p = 0.057$; fig. 3). Values for landscape factors used in this study are provided in table S1 (available online). Elevation values ranged from 1122 to 1633 m, slope ranged from 2.97 to 9.89, average monthly precipitation ranged from 16.9 to 24.1 mm, and distance from a major highway ranged from 11.76 to 7080 m. Despite the small geographic area sampled, PCOA indicated variation in these variables among the collection sites (fig. 4). Nevertheless, analyses in GESTE indicated that these landscape traits had little importance for population genetic differentiation at the scale surveyed in this study. The constant model showed the highest posterior probability (0.645), the model including longitude and the constant showed a posterior probability of 0.22, and the model containing latitude and the constant achieved a much lower posterior probability (0.11). Considering the geographic factors of slope, elevation, and distance from US Route 93, the model including the constant term again had the highest posterior probability (0.7429), while the model containing constant and elevation had a lower probability (0.11). When testing for the influence of precipitation on population differentiation, the model that included only the constant term had a higher posterior probability (0.925) compared with the model that included mean monthly precipitation (0.09; table 3).

Discussion

Genetic Diversity of Canada Thistle in North America

Canada thistle is known to grow extensively via creeping lateral roots. However, levels of within-population diversity

Table 2

Hierarchical Structure of Genetic Variation in Canada Thistle Populations according to an Analysis of Molecular Variance

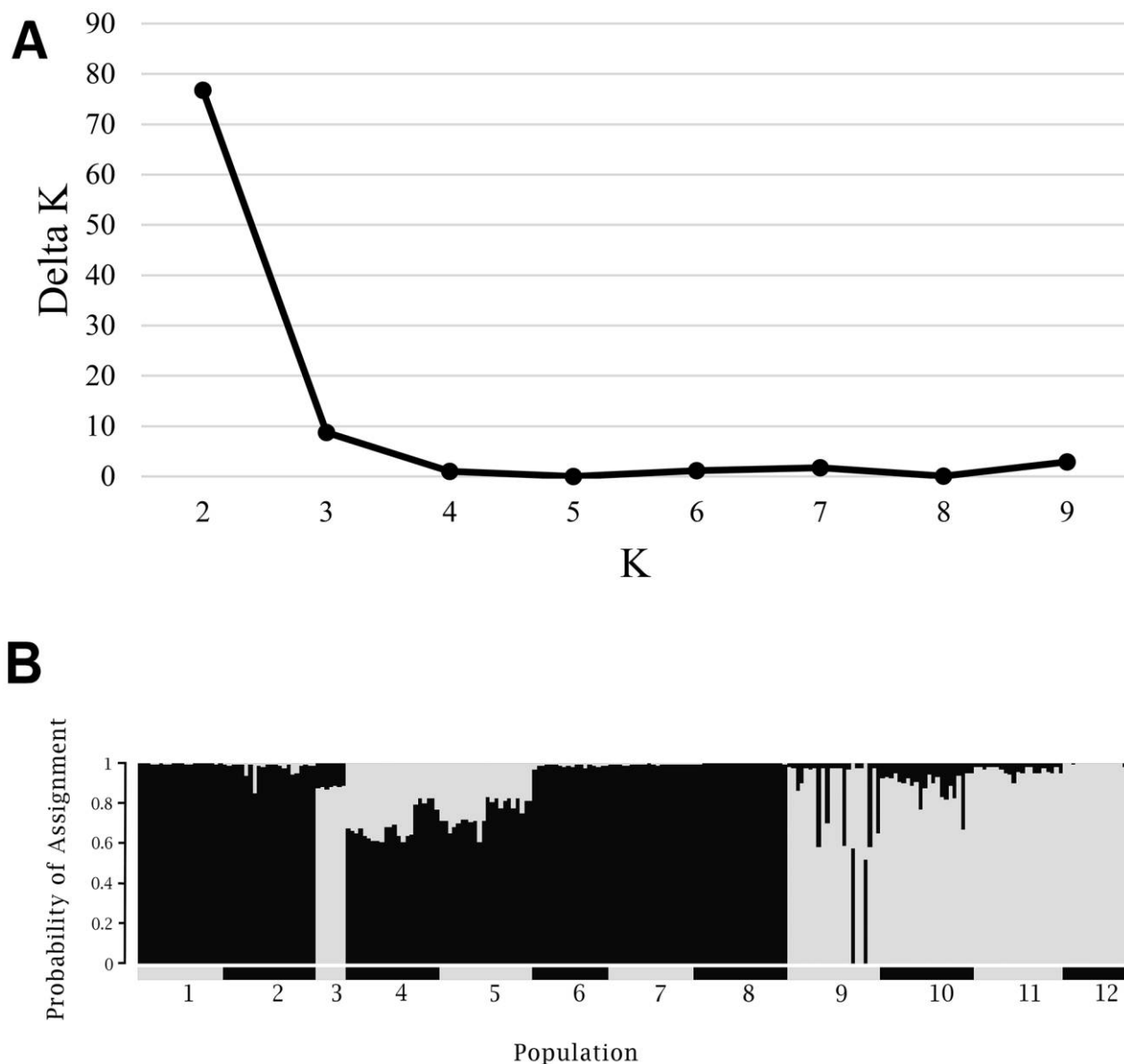| Level	df	Sum of squares	Variance component	Percent variation	Degree of differentiation (F_{ST})	p value
Among populations	11	350.33	.763	18	.21	.001
Within populations	233	788.812	3.385	82	...	...

Fig. 2   Results from Bayesian analysis in STRUCTURE. *A*, Δ*K* values based on the method of Evanno et al. (2005) for all values of *K* that were tested. *B*, Assignment probability of each sample into each of the two clusters.

observed in this study and previous studies (Hettwer and Gerowitt 2004; Solé et al. 2004; Bodo-Slotta et al. 2006, 2010; Guggisberg et al. 2012) suggest that genetic diversity within populations is influenced by both sexual and asexual reproduction. Similar to Bodo-Slotta et al. (2010), who used seven microsatellite loci, studied 93 populations in North America, and found that nearly 92% of individuals had unique multilocus genotypes, we found that more than 85% of the sampled individuals in most of the populations contained a unique multilocus genotype (table 1). In comparison with other invasive plant species of North America with similar characteristics, Canada thistle appears to have a higher level of genetic diversity. For example, Wu et al. (2018) reported mean observed and expected heterozygosity of 0.289 and 0.271, respectively, in the clonally reproducing wavyleaf basketgrass,

Oplismenus undulatifolius (Ard.) P. Beauv.; Poaceae, whereas Grimsby et al. (2007) reported observed and expected heterozygosity ranging from 0.321 to 0.534 and from 0.317 to 0.430, respectively, in invasive *Fallopia japonica* (Houtt.) Ronse Decr. (Polygonaceae). While asexual reproduction can be important for population expansion in Canada thistle (Heimann and Cussans 1996), the observed high number of individuals with unique multilocus genotypes is a strong indication that sexual reproduction is also common. Sexual reproduction can generate up to 5000 seeds per plant (Bodo-Slotta et al. 2010). Negative F_{IS} values observed in this and previous studies indicate an excess of heterozygotes and are consistent with low levels of inbreeding expected for dioecious species and/or selection for more heterozygous multilocus genotypes. Solé et al. (2004) hypothesized that
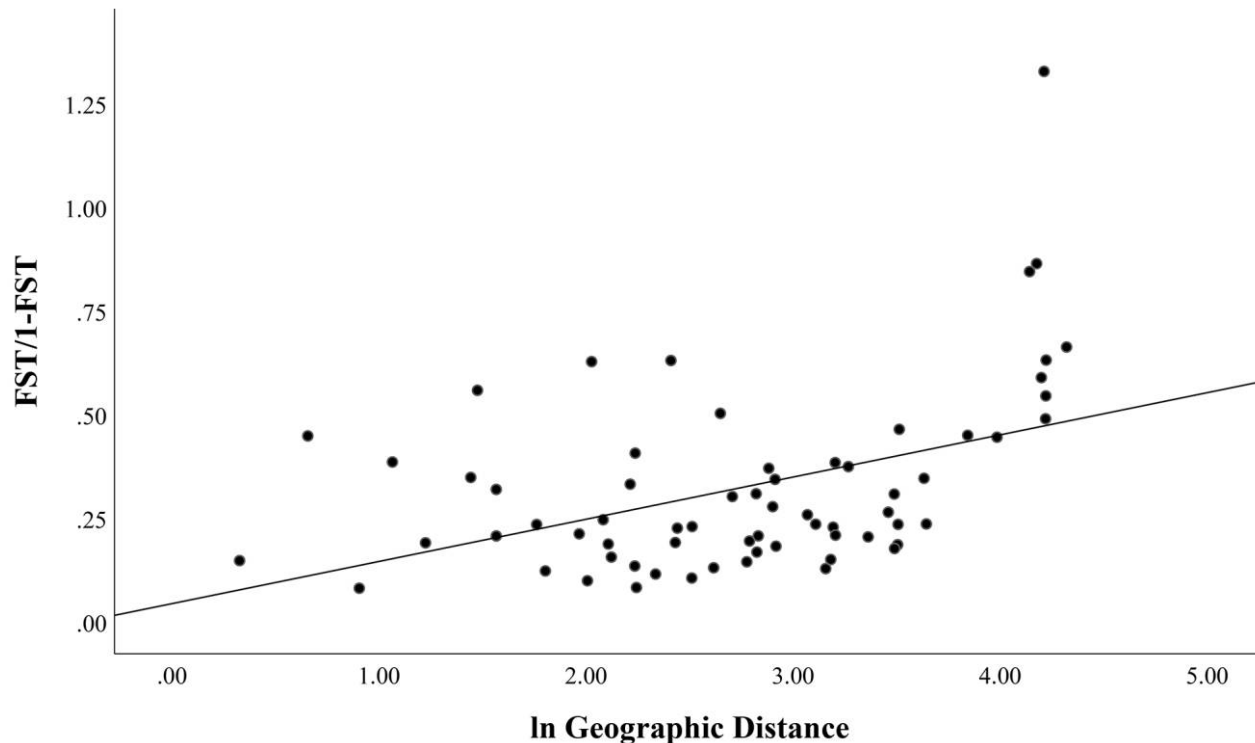
Fig. 3 Bivariate plot of pairwise genetic distances ($F_{ST}/1 - F_{ST}$) versus ln of geographical distances. A Mantel test indicated a nonsignificant correlation between the matrices ($p = 0.057$).

a combination of multiple introductions, continued gene flow within North American populations, and the lack of genetic bottlenecks has contributed to the success of Canada thistle in North America. The collective set of studies of this species in North America clearly supports the importance of sexual reproduction for maintaining diverse genotypes in this part of Canada thistle's distribution.

Although comparisons of genetic diversity among studies considering different populations and marker sets should be viewed cautiously, previous studies of Canada thistle suggest notable differences in population genetic structure between its native and introduced ranges. For example, the mean number of alleles reported by Jump et al. (2002) in 25 individuals sampled from their native European range was 6.2 across eight microsatellite loci (Caca01, Caca04, Caca05, Caca07, Caca10, Caca17, Caca22, Caca24). Bodo-Slotta et al. (2005) reported that the primers developed by Jump et al. (2002) were not polymorphic in populations from North Dakota and Minnesota, but they found even higher levels of allelic diversity among four microsatellite loci they developed (C101, C120, C128, D117; mean $A = 7.75$; Bodo-Slotta et al. 2005, 2006). Guggisberg et al. (2012) reported a remarkable mean of 41 alleles per locus across the six microsatellite loci they used to survey variation of Canada thistle populations across Europe and North America. Compared with populations sampled by Jump et al. (2002) and Bodo-Slotta et al. (2005), we found substantially lower levels of polymorphism in the sampled populations from the Rocky Mountains (mean $A = 3.9$; table 1), which may indicate genetic drift in these populations due to either founder effects or allelic

attrition associated with limited gene flow. Bodo-Slotta et al. (2010) and Guggisberg et al. (2012) reported significant genetic structure across populations of Canada thistle in North America and evidence of genetic connectedness across long distances. For example, the populations from the Rocky Mountains sampled by Bodo-Slotta et al. (2010) exhibited unique genetic differences for microsatellite loci compared with those sampled from the Midwest or the Pacific coast. Guggisberg et al. (2012) identified two genetic groups within North America, one that traces to eastern Europe and another that traces to western Europe, and strongly suggest two independent introductions, one to the East Coast and, later, another to the Midwest. Local genetic changes could also account for observed patterns of genetic structure in native and introduced ranges of Canada thistle, as Bommarco et al. (2010) found that this species has diverged genetically in Sweden according to environmental pressures from different habitat types and suggested that it is capable of rapid evolutionary change.

Spatial Scale of Genetic Differentiation

All studies of Canada thistle to date have reported significant genetic differentiation among populations at small and large geographic scales. Solé et al. (2004), who used AFLP markers in populations sampled within a 15-km² area in southern Germany, reported F_{ST} values that were nearly three times higher in those European populations ($F_{ST} = 0.64$) than the values we found in this study ($F_{ST} = 0.21$). Additionally, they reported considerable differentiation among populations even on a small geographic
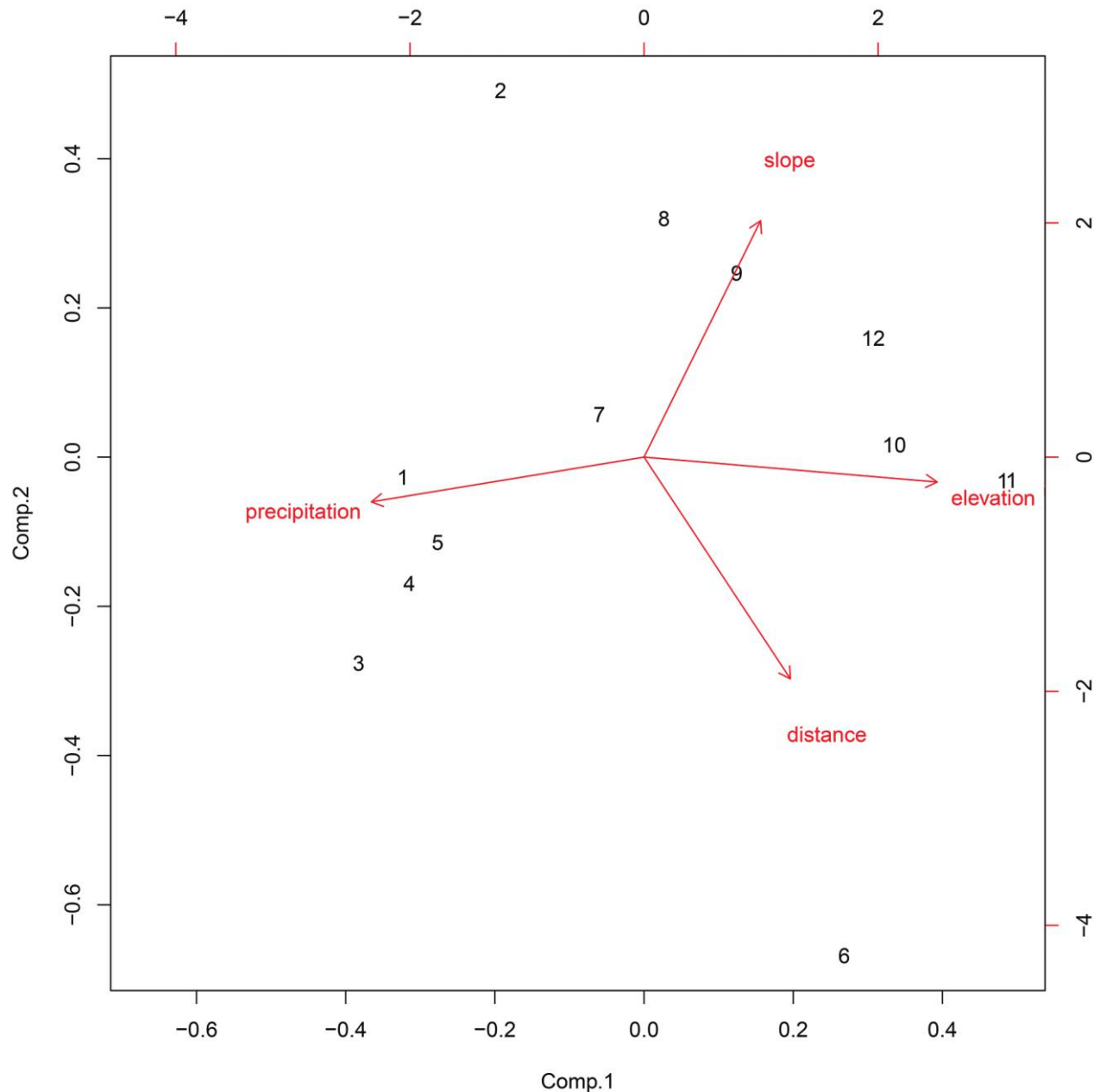
Fig. 4   Results from principal component analysis show variation in environmental variables among collection sites of Canada thistle. The importance of an environmental variable is indicated by the length of the arrow, and the direction of an arrow indicates the main change for that variable.

scale (<5 km). Genetic differentiation estimated by Bodo-Slotta et al. (2006) using microsatellite loci for midwestern populations, including 20 populations from North Dakota, one from Minnesota, and one from Virginia (F_{ST} = 0.298), is similar to that estimated for Rocky Mountain populations in this study. Bodo-Slotta et al. (2010) reported a greater mean difference among North American populations (F_{ST} = 0.264) than among populations from England (F_{ST} = 0.246); they connected these findings to strong founder effects or restriction of gene flow in North America. By contrast, Guggisberg et al. (2012) found much less variation for microsatellite loci among populations sampled across Europe and North America, with mean F_{ST} of only 0.05 in the native and introduced ranges.

These differences could reflect the use of different loci that exhibit varying levels of polymorphism because of their different mutation rates and ancestral polymorphism, but many of the loci are consistent across studies of this species in North America. Alternatively, the varying estimates of genetic differentiation could reflect biological differences associated with the colonization history of the North American populations that were surveyed.

The degree of genetic differentiation among populations in this study is noteworthy given the relatively small geographic area that was sampled. Importantly, the F_{ST} value we estimated within a 75-km area is comparable to that estimated by Bodo-Slotta et al. (2010) for populations sampled at the continental scale. We

Table 3

Importance of Geographic and Environmental Factors in Explaining the Observed Genetic Structure of Canada Thistle Populations

Environmental scenario	Factor(s) included	Posterior probability
Spatial range expansion	*Constant*	*.645*
	Constant, latitude	.11
	Constant, longitude	.22
	Constant, latitude, longitude	.02
	Constant, latitude, longitude, latitude × longitude	.0001
Geographic suitability	*Constant*	*.742*
	Constant, slope	.06
	Constant, elevation	.11
	Constant, distance to road	.06
	Constant, slope, distance to road	.0001
	Constant, slope, elevation	.01
	Constant, distance to road, elevation	.01
	Constant, elevation, distance to road, slope	.0001
Climatic suitability	*Constant*	*.925*
	Constant, precipitation	.09

Note. Posterior probabilities were generated using GESTE. For each analysis, the model with the highest probability is shown in italics.

predicted that genetic structure in Rocky Mountain populations of Canada thistle would be determined by physical distance between populations as well as by resistance to movement associated with heterogeneity in the landscape. However, we did not find strong evidence of an isolation-by-distance pattern or evidence that environment and climatic factors explain a significant portion of the observed structure. Nevertheless, abiotic factors of the landscape could influence seed dispersal and pollinator behavior in ways that were not considered in our analyses. For example, the pollinators of Canada thistle are highly mobile bees (Walther-Hellwig and Frankl 2000), but there is a high level of traffic along US Route 93 that could influence pollinator abundance, thereby creating genetic connections among some populations and enhancing differentiation among other populations. For example, there is evidence that while bee and bird pollinators may have the ability to cross roads and railroads, human structures may restrict their movement to only one side (Bhattacharya et al. 2003; Geerts and Pauw 2011). Additionally, if vehicles hit pollinators, then pollinator numbers along busy roads could also influence gene flow among Canada thistle stands. Finally, the presence of ridges and intervening valleys in the study area could impede movement of pollen and seeds among populations if pollinators are unable or unwilling to move between these areas.

STRUCTURE analysis shows that considerable differentiation can exist at small and large spatial scales. Similar to Guggisberg et al. (2012), we identified two genetic groups within the study area and evidence of admixture of these groups. It is unclear whether the groups identified in the two studies represent the same gene pools, but Guggisberg et al. (2012) did identify admixture in populations from Idaho, Colorado, and South Dakota, which are close to the area studied here. Individuals in populations 4 and 5 exhibit mixed ancestry between physically close populations and populations at least 80 km away, and population 3 is clustered with distant populations. Pairwise genetic differentiation between populations (table 4) shows that average pairwise F_{ST} value within cluster 1 is less than between two clusters, which affirms the fact that populations grouped in one cluster are genetically more similar to each other than populations in different clusters are. However, this expectation was not met in cluster 2, which suggests substructure within this cluster that may arise from genetic differences, primarily in populations 3 and 12. Populations 3 and 12 have the highest mean F_{ST} values (0.340 and 0.382, respectively), highest F_{IS} values (−0.789 and −0.499, respectively), lowest number of alleles per locus (2.0 and 2.1, respectively), and lowest expected heterozygosity (0.418 and 0.286, respectively). Clustering of population 3

Table 4

Mean Population Pairwise Degree of Differentiation (F_{ST}) Observed in Canada Thistle

Population	1	2	3	4	5	6	7	8	9	10	11
2	.07432										
3	.358	.3092									
4	.1746	.1711	.2778								
5	.1345	.1087	.2579	.1281							
6	.1537	.1259	.3342	.1601	.1027						
7	.1838	.118	.3853	.2416	.1595	.1575					
8	.1148	.09498	.3863	.2488	.1971	.1863	.1896				
9	.2553	.2169	.2698	.1712	.1626	.2048	.2316	.2888			
10	.1854	.1724	.2771	.1301	.1132	.1695	.1901	.2356	.08956		
11	.2351	.1898	.3167	.1563	.1497	.1905	.2087	.272	.1435	.07581	
12	.4633	.3704	.5702	.3867	.3524	.3983	.3284	.4577	.3077	.3099	.2567

and admixture of populations 4 and 5 with populations 9–12 suggest that this geographic area has been colonized from distinct gene pools. The results presented here suggest that admixture can occur at very small spatial scales and that even genetically distinct types can successfully invade similar habitats. Admixed populations are expected in systems where geographical structure exists because of limited gene flow but dispersal barriers are not absolute. In such cases, admixed populations are expected to be geographically and genetically intermediate between the ranges occupied by each population. Admixed populations may also appear if microevolutionary processes within populations alter the frequency of certain genotypes and then those changes are maintained, at least over short periods, because of limited gene flow with neighboring populations.

Conclusion

Canada thistle is ranked as one of the most problematic invasive weeds in North America because it can cause habitat destruction and competes with native and agricultural plants, ultimately causing economic losses (Donald 1994; McClay 2002). Mosena et al. (2018) showed that the cumulative number of counties in North America invaded by this species has increased significantly since 1880, which indicates that this species is ge-

netically capable of relatively fast acclimatization to new environments. Results from this work and previous studies of genetic structure in North American populations (Bodo-Slotta et al. 2006, 2010; Guggisberg et al. 2012) are helpful for identifying areas of distinct and mixed genetic stocks. Genetic admixture is likely to be a major factor in the continued establishment and persistence of this species across temperate areas around the world in the future. Results from the present study reveal significant genetic structure across Canada thistle populations at a relatively small spatial scale of its invasive range and the potential for admixture between distinct gene pools. Given the potential for rapid local spread of populations suggested by Bommarco et al. (2010) and the novelty that admixed genotypes could contribute to invasiveness, we suggest that management practices should consider the importance of genetic diversity and structure in future control measures of this species.

Acknowledgments

Appendix

Voucher Numbers of Herbarium Specimens from Each Sampled Location

Vouchers are deposited at the Mississippi State University Herbarium (Missa).

Population 1: LEW288, Population 2: LEW289, Population 3: LEW290, Population 4: LEW291, Population 5: LEW292, Population 6: LEW293, Population 7: LEW294, Population 8: LEW282, Population 9: LEW283, Population 10: LEW284, Population 11: LEW285, Population 12: LEW286.

Literature Cited

Abela-Hofbauerova I, Z Münzbergova, J Skuhrovec 2011 The effect of different natural enemies on the performance of *Cirsium arvense* in its native range. Weed Res 51:394–403.

ASTER GDEM Validation Team 2009 ASTER global DEM validation summary report. Ministry of Economy, Trade, and Industry, Tokyo/NASA, Washington, DC. 28 pp.

Becker RL, MJ Haar, BD Kinkaid, LD Klossner, F Forcella 2008 Production and wind dispersal of Canada thistle (*Cirsium arvense* L.) achenes. Minnesota Department of Transportation, St. Paul. http://www.lrrb.org/pdf/200839.pdf.

Bhattacharya M, RB Primack, J Gerwein 2003 Are roads and railroads barriers to bumblebee movement in a temperate suburban conservation area? Biol Conserv 109:37–45.

Blamey M, C Grey-Wilson 1989 The illustrated flora of Britain and northern Europe. Lubrecht & Cramer, Port Jervis, NY.

Bodo-Slotta TA, ME Foley, S Chao, RA Hufbauer, DP Horvath 2010 Assessing genetic diversity of Canada thistle (*Cirsium arvense*) in North America with microsatellites. Weed Sci 58:387–394.

Bodo-Slotta TA, ME Foley, DP Horvath 2005 Development of polymorphic markers for *Cirsium arvense*, Canada thistle, and their amplification in closely related taxa. Mol Ecol Notes 5:917–919.

Bodo-Slotta TA, JM Rothhouse, DP Horvath, ME Foley 2006 Genetic diversity of Canada thistle (*Cirsium arvense*) in North Dakota. Weed Sci 54:1080–1085.

Bommarco R, M Lönn, U Danzer, KJ Pålsson, P Torstensson 2010 Genetic and phenotypic differences between thistle populations in response to habitat and weed management practices. Biol J Linn Soc 99:797–807.

Dakin EE, JC Avise 2004 Microsatellite null alleles in parentage analysis. Heredity 93:504–509.

Dandeno JB 1905 The parachute effect of thistle-down. Science 22:568–572.

Dewey HL 1901 Canada thistle. USDA Botany Circular 27. USDA, New York.

Donald WW 1994 The biology of Canada thistle (*Cirsium arvense*). Rev Weed Sci 6:77–101.

Earl DA, BM vonHoldt 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361.

Eber S, R Brandl 2003 Regional patch dynamics of *Cirsium arvense* and possible implications for plant-animal interactions. J Veg Sci 14:259–266.

ESRI (Environmental Systems Research Institute) 2018 ArcGIS geostatistical analyst release 10.6.1. ESRI, Redlands, CA. http://resources.arcgis.com.

Evanno G, S Regnaut, J Goudet 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620.

Evans CL 2002 The war on weeds in the Prairie West: an environmental history. University of Calgary Press, Toronto, ON.

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

Falush D, M Stephens, JK Pritchard 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

——— 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578.

Fick SE, RJ Hijmans 2017 WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37:4302–4315.

Fisher RA 1954 Statistical methods for research workers. 12th ed. Oliver & Boyd, Edinburgh.

Foll M, OE Gaggiotti 2006 Identifying the environmental factors that determine the genetic structure of populations. Genetics 174:875–891.

Geerts S, A Pauw 2011 Easy technique for assessing pollination rates in the genus *Erica* reveals road impact on bird pollination in the Cape fynbos, South Africa. Austral Ecol 36:656–662.

Grimsby JL, D Tsirelson, MA Gammon, R Kesseli 2007 Genetic diversity and clonal vs. sexual reproduction in *Fallopia* spp. (Polygonaceae). Am J Bot 94:957–964.

Guggisberg A, E Welk, R Sforza, DP Horvath, JV Anderson, ME Foley, LH Rieseberg 2012 Invasion history of North American Canada thistle, *Cirsium arvense*. J Biogeogr 39:1919–1931.

Guillot G, F Mortier, A Estoup 2005 GENELAND: a computer package for landscape genetics. Mol Ecol Notes 5:712–715.

Hansen AA 1918 Canada thistle and methods of eradication. USDA Farmers' Bulletin 1002. USDA, Washington, DC.

Heimann B, GW Cussans 1996 The importance of seeds and sexual reproduction in the population biology of *Cirsium arvense*—a literature review. Weed Res 36:493–503.

Hettwer U, B Gerowitt 2004 An investigation of genetic variation in *Cirsium arvense* field patches. Weed Resour 44:289–297.

Hodgson JM 1968 The nature, ecology, and control of Canada thistle. USDA Technical Bulletin 1386. USDA, Washington, DC.

Hoffmann JD, S Narumalani, DR Mishra, P Merani, RG Wilson 2008 Predicting potential occurrence and spread of invasive plant species along the North Platte River, Nebraska. Invasive Plant Sci Manag 1:359–367.

Hoffmeister TS, LEM Vet, A Biere, K Holsinger, J Filser 2005 Ecological and evolutionary consequences of biological invasion and habitat fragmentation. Ecosystems 8:657–667.

Holderegger R, HH Wagner 2006 A brief guide to landscape genetics. Landsc Ecol 21:793–796.

Holm S 1979 A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70.

Hubisz MJ, D Falush, M Stephens, JK Pritchard 2009 Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332.

Jump AS, DA Dawson, CM James, FI Woodward, T Burke 2002 Isolation of polymorphic microsatellites in the stemless thistle (*Cirsium acaule*) and their utility in other *Cirsium* species. Mol Ecol Notes 2:589–592.

Kamvar ZN, JF Tabima, NJ Grünwald 2014 *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. http://doi.org/10.7717/peerj.281.

Kay Q 1985 Hermaphrodites and subhermaphrodites in a reputedly dioecious plant, *Cirsium arvense* (L.) Scop. New Phytol 100:457–472.

Kluth S, A Kruess, T Tscharntke 2001 Interactions between the rust fungus *Puccinia punctiformis* and ectophagous and endophagous insects on creeping thistle. J Appl Ecol 38:548–556.

Kopelman NM, J Mayzel, M Jakobsson, NA Rosenberg, I Mayrose 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191.

Lalonde RG, BD Roitberg 1994 Mating system, life-history, and reproduction in Canada thistle (*Cirsium arvense*; Asteraceae). Am J Bot 81:21–28.

Lawson Handley LJ, A Estoup, DM Evans, CE Thomas, E Lombaert, B Facon, A Aebi, HE Roy 2011 Ecological genetics of invasive alien species. Biol Control 56:409–428.

Louda SM, MA Potvin 1995 Effect of inflorescence-feeding insects on the demography and lifetime fitness of a native plant. Ecology 76:229–245.

Manel S, MK Schwartz, G Luikart, P Taberlet 2003 Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197.

Mantel N 1967 The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

McClay AS 2002 Canada thistle. Page 413 *in* R Van Driesche, B Blossey, M Hoddle, S Lyon, R Reardon. Biological control of invasive plants in the eastern United States. USDA Forest Service Publication FHTET-2002-04. USDA, Morgantown, WV.

Moore RJ 1975 The biology of Canadian weeds. 13. *Cirsium arvense* (L.) Scop. Can J Plant Sci 55:1033–1048.

Mosena A, T Steinlein, W Beyschlag 2018 Reconstructing the historical spread of non-native plants in the North American West from herbarium specimens. Flora 242:45–52.

Nadeau LB, WH Vanden Born 1989 The root system of Canada thistle. Can J Plant Sci 69:1199–1206.

Nuzzo V 1997 Element stewardship abstract for *Cirsium arvense*. Nature Conservancy, Arlington, VA.

Peakall R, PE Smouse 2012 GenAlEx 6.5: genetic analysis in Excel: population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539.

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

QGIS Development Team 2018 QGIS geographic information system. Open Source Geospatial Foundation Project, Beaverton, OR. http://qgis.osgeo.org.

Raymond M, F Rousset 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

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

——— 2008 GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106.

RStudio Team 2012 RStudio: integrated development environment for R (ver. 1.1.456). RStudio, Boston. http://www.rstudio.com/.

Sakai AK, FW Allendorf, JS Holt, DM Lodge, J Molofsky, KA With, S Baughman, et al 2001 The population biology of invasive species. Annu Rev Ecol Evol Syst 32:305–332.

Sheldon JC, FM Burrows 1973 The dispersal effectiveness of the achene-pappus units of selected Compositae in steady winds with convection. New Phytol 72:665–675.

Sinnott RW 1984 Virtues of the haversine. Sky Telesc 68:158.

Skuhrovec J, S Koprdova, J Mikulka 2008 How can seed feeders regulate seed dispersion of thistles, *Cirsium arvense* and *C. heterophyllum*? J Plant Dis Prot 21:281–284.

Solé M, W Durka, S Eber, R Brand 2004 Genotypic and genetic diversity of the common weed *Cirsium arvense* (Asteraceae). Int J Plant Sci 165:437–444.

Stevens A 1847 Extirpation of Canada thistles. Trans NY State Agric Soc 6:405–428.

Storfer A, MA Murphy, JS Evans, CS Goldberg, S Robinson, SF Spear, R Dezzani, E Delmelle, L Vierling, LP Waits 2007 Putting the "landscape" in landscape genetics. Heredity 98:128–142.

Theis N 2006 Fragrance of Canada thistle (*Cirsium arvense*) attracts both floral herbivores and pollinators. J Chem Ecol 32:917–927.

Tipping PW 1993 Field studies with *Cassida rubiginosa* (Coleoptera, Chrysomelidae) in Canada thistle. Environ Entomol 22:1402–1407.

van Oosterhout C, WF Hutchinson, DPM Wills, P Shipley 2004 MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538.

Walther-Hellwig K, R Frankl 2000 Foraging habitats and foraging distances of bumblebees, *Bombus* spp. (Hymenoptera, Apidae), in an agricultural landscape. J Appl Entomol 124:299–306.

Ward SM, JF Gaskin, LM Wilson 2008 Ecological genetics of plant invasion: what do we know? Invasive Plant Sci Manag 1:98–109.

Weir BS, CC Cockerham 1984 Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wu CA, AD Hakkenberg, VB Beauchamp 2018 Characterization of polymorphic microsatellite loci for invasive wavyleaf basketgrass, *Oplismenus undulatifolius* (Poaceae). Appl Plant Sci 6:e1016.

Table S1. Landscape factors values

Pop	Slope	Distance from Hwy 93 (meter)	Elevation (meter)	Longitude	Latitude	Precipitation (mm)
1	6.25095	120	1125	-113.965	45.36589	24.1
2	12.5718	1.76	1122	-113.943	45.3639	24.1
3	2.97197	8.86	1131	-113.926	45.35323	23.9
4	4.26254	7.7	1152	-113.904	45.32159	23.5
5	4.83862	7.5	1150	-113.895	45.2984	23
6	6.7408	7080	1409	-113.803	45.27956	22.4
7	8.53384	1333	1210	-113.875	45.24178	22.3
8	9.8913	6.2	1223	-113.893	45.12032	20.6
9	8.96419	10.5	1309	-113.947	44.96081	19.8
10	6.03471	169	1464	-113.987	44.82944	17.3
11	8.31609	2646	1457	-114.04	44.66603	16.9
12	8.77569	23.5	1633	-114.336	44.25793	21.5