**TITLE**: ​​Isolation-by-environment and its consequences for range shifts with global change: landscape genomics of the invasive plant common tansy

**RUNNING TITLE**: Isolation-by-environment in an invasive plant

**AUTHORS**: Ryan Briscoe Runquist1,2,4 and David A. Moeller2,3,5

1Corresponding author email: [rbriscoe@umn.edu](mailto:rbriscoe@umn.edu)

2email: [moeller@umn.edu](mailto:moeller@umn.edu)

3University of Minnesota, Plant and Microbial Biology Department, 140 Gortner Laboratory, 1479 Gortner Ave, St. Paul, MN 55108

4https://orcid.org/0000-0001-7160-9110

5https://orcid.org/0000-0002-6202-9912

**ABSTRACT**

Invasive species are a growing global economic and ecological problem. However, it is not well understood how environmental factors mediate invasive range expansion. In this study, we investigated the recent and rapid range expansion of common tansy across environmental gradients in Minnesota, U.S.A. We densely sampled individuals across the expanding range and performed reduced representation sequencing to generate a dataset of 3071 polymorphic loci for 176 individuals. We used non-spatial and spatially-explicit analyses to determine the relative influences of geographic distance and environmental variation on patterns of genomic variation. We found no evidence for isolation-by-distance (IBD) but strong evidence for isolation-by-environment (IBE), indicating that environmental factors may have modulated patterns of range expansion. Land use classification and soils were particularly important variables related to population structure although they operated on different spatial scales; land-use classification was related to broad-scale patterns and soils were related to fine-scale patterns. All analyses indicated a distinctive genetic cluster in the most recently invaded portion of the range. Individuals from the far northwestern range margin were separated from the remainder of the range by reduced migration, which was associated with environmental resistance. This portion of the range was invaded primarily in the last 15 years. Ecological niche models also indicated that this cluster was associated with expansion of the niche. While invasion is often assumed to be primarily influenced by dispersal limitation, our results suggest that ongoing invasion and range shifts with climate change may be strongly affected by environmental heterogeneity.

**INTRODUCTION**

Range expansion of invasive species represents a major aspect of global change facing ecosystems. Invasive species movement across landscapes is modulated by a variety of genetic and ecological factors (Lee 2002; Dlugosch *et al.* 2015). Losses of genetic variation may occur during rapid range expansion due to successive genetic bottlenecks (Nei *et al.* 1975; Dlugosch & Parker 2008; but see Roman & Darling 2007; Lake *et al.* 2023). Such losses of variation may stymie invasion because of inbreeding depression or a limited capacity to adapt to novel environments (Frankham 2005; but see Dlugosch *et al.* 2015; Oduor *et al.* 2016). Over shorter timescales, the pattern and extent of range expansion may be determined by the distribution of suitable habitat. Comparatively little is understood about how environmental heterogeneity influences pathways of movement, particularly during rapid invasions. Recent developments in landscape genomics provide an opportunity to test whether (and to identify which) environmental factors have facilitated or inhibited invasion (Clements & Ditommaso 2011; Barrett 2015; Tomiolo & Ward 2018; Clements & Jones 2021).

Invasion history may influence the geography of genetic variation in several ways. The number and geographic origin of colonists influence genetic variation in founding populations (Baker & Stebbins 1965; Dlugosch & Parker 2008). Following introduction, the nature of population dynamics can be particularly important for the amount of genetic variation at leading range edges. Successive founder events during range expansion typically result in losses of heterozygosity and rare alleles because genetic drift is stronger at greater distances from the location of introduction (Baker & Stebbins 1965; Nei *et al.* 1975; Slatkin & Excoffier 2012; Peter & Slatkin 2013). Rapid range expansion may also cause allele surfing, where new mutations arising at the invasion front rapidly increase in frequency regardless of their selection coefficient (Klopfstein *et al.* 2006; Hallatschek *et al.* 2007; Slatkin & Excoffier 2012; Peter & Slatkin 2013). In these cases, range expansion tends to affect levels of diversity and the geographic structure of variation. By contrast, when range expansion involves high rates of long-distance dispersal (as opposed to wave-like expansion), there are unlikely to be clear patterns of population structure (Ibrahim *et al.* 1996).

Over long time periods, the signature of invasion history on population genomic structure should erode due to ongoing migration, drift, and adaptation. At migration-drift equilibrium, we expect to find isolation-by-distance (IBD), whereby there is a positive correlation between genetic and geographic distance (Wright 1943; Kimura & Weiss 1964; Kimura 1964; Cox & Durrett 2002). In invasive species, IBD is not necessarily expected or observed (Slatkin 1993; Hutchison & Templeton 1999). The absence of IBD may occur because of rapid range expansion and stochastic long-distance dispersal events, which are common during invasion (Wright 1943; Slatkin 1987, 1993; Hutchison & Templeton 1999). Conversely, IBD may be observed if the invasion occurred long enough in the past that there is migration-drift equilibrium. A signature of IBD may be observed if an invasion is the result of multiple introductions into different geographic regions that have not had sufficient time for expansion and admixture (Taylor & Keller 2007; Bradburd *et al.* 2013; Sexton *et al.* 2014; van Strien *et al.* 2015). Finally, deviations from IBD may suggest that environmental factors have caused particular geographic pathways for movement such that distance is not a strong predictor of genetic distance.   
 Populations from analogous environments may exhibit greater genetic similarity than those from geographically close locations (i.e. isolation-by-environment: IBE) (Wang *et al.* 2013; Shafer & Wolf 2013; Bradburd *et al.* 2013; Sexton *et al.* 2014). Whereas IBD may be unlikely during rapid invasion, IBE could arise if invasion pathways are modulated by particular environments. Adaptive differentiation can lead to patterns of IBE when there is strong selection against migrants from different environments (Räsänen & Hendry 2008). For invasive species, adaptation may increase the set of environments available for invasion or allow for rapid responses to future climate change (Gallien *et al.* 2012; DeMarche *et al.* 2019). Indeed, local adaptation has been detected as often in non-native as native species (Oduor *et al.* 2016). However, patterns of IBE are also possible due to non-adaptive processes that decouple geographic distance from genetic similarity, such as variation in migration across environmental gradients (e.g. different vegetation types) or geographic features that function as barriers to migration (e.g. rivers, mountains, or cropland) (Shafer & Wolf 2013; Sexton *et al.* 2014; Petkova *et al.* 2016; Bradburd *et al.* 2018; Yang *et al.* 2020; Marcus *et al.* 2021). In an applied context, determining environmental drivers of movement can help prioritize eradication efforts and identify habitats at risk of future invasion (Peterman & Pope 2021).

Adaptation to new environments during invasion may change the nature or breadth of the ecological niche. Land managers often use ecological niche models (ENMs) to develop invasion forecasts and management planning (Peterson & Vieglais 2001; Guisan *et al.* 2014). These models rely on accurate characterizations of a species ecological niche, however, the models generally do not account for adaptation or dispersal limitation (Veloz *et al.* 2012; Yates *et al.* 2018; Atwater & Barney 2021). Increases in overall niche breadth via adaptation may allow for colonization of new habitats and increase the potential range size of the invader (Kawecki 2008; Tomiolo & Ward 2018; Putra *et al*. 2023). Additionally, adaptive divergence may lead to the development of distinctive genotypes that occupy only a subset of the species-level niche. For example, Johnsongrass exhibits adaptive differentiation in niche between agricultural and non-agricultural populations, which has broadened the overall species niche and increased the potential for economic damage (Atwater *et al.* 2016; Lakoba *et al.* 2021). Niche shifts such as this may enable greater invasive spread into regions previously deemed low risk. Overall, understanding niche evolution may be important for predicting range expansion and developing habitat-specific management strategies (Kawecki 2008; Mouquet *et al.* 2015).

In this study, we investigated landscape genetic variation of the invasive species, common tansy (*Tanacetum vulgare*), to determine how environmental factors shaped dispersal pathways during invasion. We also tested whether areas of recent invasion involved niche expansion. Common tansy is an economically damaging invasive species in North America that degrades rangelands, impedes restoration efforts, and outcompetes native plants. It is especially problematic across the northern tier of the United States and is particularly abundant in Minnesota, where invasive spread has been recent and rapid (Fig. 1) (MN Dept. of Agriculture: Common tansy 2023; MN Dept. of Natural Resources: Common Tansy 2023; White 2001; LeCain & Sheley 2014). First, we densely sampled individuals across the invaded range in Minnesota to test for fine-scale genetic structure. We assessed if landscape genetic structure was best explained by geographic distance (IBD) in both spatial and non-spatial frameworks. We then developed environmentally informed genetic resistance models to test which environmental factors influenced pathways of movement (IBE). Last, we constructed ENMs of individual genetic clusters to test if recovered landscape genetic structure also represented differences in the ecological niche. Together, our analyses provide insight about the factors that have shaped recent range expansion and may impede continued invasion.

**MATERIALS AND METHODS**

*Natural history:* Common tansy (*Tanacetum vulgare* L.; Asteraceae) is an herbaceous, C3 perennial native to Europe (Fig. S1). Individuals form large multi-stemmed plants with shallow root systems (Jacobs 2008). Flowers are largely self-sterile, although mixed-mating has been documented; outcrossing may occur via a diverse array of insects (Fig. 1C) (LeCain & Sheley 2014; Jacobs 2008). Recruitment occurs via seeds, rhizomes, and root fragments, which are dispersed by wind, water, and human activity (White 1997).

Common tansy is distributed across the northern tier of the United States and southern tier of Canada with very few occurrences south of 40°S latitude (Fig 1A). It was introduced to the northeastern United States from Eurasia as early as the 1600s (Fig. 1) (USDA Plants Database 2023; Mitich 1992; Mack 2003; Clasen *et al.* 2011; Lake *et al.* 2020). Minnesota has the highest density of reported occurrences in the United States (Fig. 1A) with the earliest record from 1878 in the southeastern corner of the state (Roberts 1878; Mack 2003). During the 20th century, infrequent occurrences were recorded, primarily in the northeastern and central areas of the state (Fig 1D). In the past 20 years, the distribution rapidly expanded, first into northeastern MN, followed by expansion west and south, and last into northwestern MN. In Minnesota, the southern and western range margins occur along an ecological transition from cooler, wetter forests, to warmer, drier prairie grasslands and occurrence records from the northwest are all recent (nearly all <10 years old) and are much less frequent than other areas (Fig. 1D). Within its range, common tansy typically occurs in open habitats and is a weed of pastures and rangelands. Plants are avoided by livestock and insect herbivores because they produce a diverse set of terpenes that are toxic (Wolf *et al.* 2012).

Common tansy is diploid with ca. 4300 Mbp per haploid genome (Keskitalo *et al.* 1998). Genomic resources have not been developed for the genus *Tanacetum*; past studies have involved a handful of populations and few molecular markers (e.g. microsatellites; Clasen *et al.* 2011).

*Sample collections:* In summer/fall 2019, we collected individuals from 176 populations in the invaded range of Minnesota (n = 174) and western Wisconsin (n = 2), U.S.A. These populations spanned the entire distribution of common tansy in Minnesota and represented most fine-scale environmental variation (Fig. 1B). Populations occur in 22 of the 26 ecological subregions (MN Dept. of Natural Resources Ecological Classification System 2023; Fig. S2), which describe areas with similar ecological features such as climate, topography, and soils. Throughout the paper, we will refer to populations by ecological subregion (Table S1; Fig. 1B & Fig. S3).

We collected leaves from one individual per population, dried leaf tissue in silica gel, and extracted DNA using Genesee Zymo Quick-DNA Plant/Seed Miniprep Kits. Reduced representation sequencing was performed with Illumina NextSeq at the U. of Minnesota Genomics Center (See Supplemental Materials for more detailed methods). We aligned reads and called SNPs using the Stacks 2.59 *de novo* pipeline and *populations* function (see Supplemental Methods for more details on sequence analyses) (Catchen *et al.* 2011; Rochette *et al.* 2019). We included loci that were present in at least 70% of individuals and all individuals were genotyped at >60% of loci (mean number of missing loci per individual: 19.6%). Our final dataset included 3690 loci, of which 3071 were polymorphic (Figs. S4 & S5).

*Landscape genetic analysis*

All analyses were conducted in the R v. 4.1.1 (R Core Team 2021) unless otherwise noted.

*Isolation-by-distance:* We tested for isolation-by-distance (IBD) using a Mantel test of the correlation between genomic and geographic distance. Genomic distance was calculated using Nei’s D with the ‘dist.genepop’ function in the package ‘adegenet’ (Jombart 2008; Jombart & Ahmed 2011); geographic distance was calculated using the ‘rdist.earth’ function in the package ‘fields’ (Nychka *et al.* 2021). We tested whether the observed correlation was greater than random expectation using a permutation test with 999 permutations.

We also determined the geographic distance at which the IBD correlation was maximized using the ‘dist\_max\_corr’ function from the ‘graph4lg’ package (Savary *et al.* 2020). The function uses an expanding window to calculate the correlation coefficient for pairs with geographic distances that fall within a set range for 1000 sequential range intervals from 0-659 km.

*Principal Components and Spatial Principal Components Analyses:* We conducted genetic principal components analysis using the function ‘dudi.pca’ from the package ‘ade4’ (Chessel *et al.* 2004; Dray *et al.* 2007; Dray & Dufour 2007; Bougeard & Dray 2018; Thioulouse *et al.* 2018). Principal components describe the major axes of genetic variation and cluster individuals that have similar genotypes (Novembre & Stephens 2008; Reich *et al.* 2008; Abegaz *et al.* 2019).

We also conducted spatial principal components analysis (sPCA), which is a spatially-explicit ordination. The method accounts for both genetic variation (PCA) and spatial autocorrelation (Moran’s I) (Jombart *et al.* 2008). The PCA was calculated as above using allele frequencies and spatial autocorrelation using a neighborhood weighting matrix describing the spatial proximity of samples. We calculated neighborhood weights using a Gabriel graph (Matula & Sokal 2010) (Fig. S6). Large positive eigenvalues of sPCA describe global patterns of spatial variation where there is both high genetic variation and high values of Moran’s I (i.e. physically close samples are also most similar genetically). Large negative eigenvalues describe local patterns of genetic structure (i.e. physically closer samples are more genetically dissimilar). These analyses were conducted with the ‘spca’ function in the package ‘adegenet' (Jombart 2008; Jombart & Ahmed 2011).

We tested for the overall significance of global and local genetic patterns using the ‘global.rtest’ and ‘local.rtest’ permutation tests in ‘adegenet’. These tests determine if the maximum correlation between allele frequency and global (positive) or local (negative) Moran Eigenvector Maps (MEMs) computed for the sample gabriel graph is greater than random expectation (Jombart *et al.* 2008). A significant result indicates that there was at least one sPCA axis that described significant spatial structure in genetic variation. We then used the ‘screeplot’ function to visualize the relative importance of sPCA eigenvalues. The function generates a screeplot and a histogram to determine which axes to further examine (Jombart *et al.* 2008). We visualized eigenvalues that were disjunct from the majority in the screeplot and histogram (Jombart *et al.* 2008).

*ConStruct Analysis:* We used conStruct to determine if there were areas of discrete population structure, which may indicate an environmental influence on invasive spread, while accounting for potentially continuous isolation-by-distance (Bradburd *et al.* 2018). We calculated cluster assignments for K = 1 to 6 within both a spatial framework (i.e. accounting for continuous IBD) and in a non-spatial framework (i.e. where the spatial element in the conStruct model was set to 0, which is similar to a traditional STRUCTURE analysis (Pritchard *et al.* 2000)). We used the ‘conStruct’ function from the package ‘conStruct’ and ran each model for 250,000 iterations and visualized trace plots of chains to ensure proper convergence and mixing (Bradburd 2019). We calculated layer contributions and determined the most supported number of clusters as occurring when the contribution of an additional layer was less than 5% (Bradburd *et al.* 2018). We corroborated clustering patterns using fineRADstructure (Malinsky *et al.* 2018) (See Supplement for fineRADstructure methods and results; Fig. S21).

*Effective Migration Surfaces*: We examined whether isolation-by-distance was spatially heterogeneous using Fast Estimation of Effective Migration Surfaces (FEEMS; Petkova *et al.* 2016; Marcus *et al.* 2021). The model generates expected genetic distances between populations, which approximate the resistance to gene flow between the populations (Petkova *et al.* 2016; Marcus *et al.* 2021). Weights (w) along edges in the grid are then assigned, which represent effective migration. FEEMS has the capacity to handle sparse grids with anisotropic migration between nodes, which were present in our spatial dataset.

We constructed a triangular discrete digital global grid (~100 km2 /triangle) across the state of Minnesota using the package ‘dggridR’ (Barnes & Sahr 2021). Each sample was associated with the closest node in the grid (Fig. S7). The 176 samples were associated with 146 nodes. At grid nodes assigned more than one sample, we calculated allele frequencies based on the combined genetic information. The allele frequency dataset was converted to plink format using plink 1.9 (Purcell *et al.* 2007; Purcell 2019) using plink ready files exported from Stacks 2.59 *populations* function (see above).

Effective migration surface models included a regularization term that penalized large differences in weights between adjacent nodes and was controlled by the smoothing parameter, λ. We used a ‘leave-one-out’ cross-validation procedure (i.e. 146-folds that leave a node out of estimation and used it for calculation of the loss function) to determine the most appropriate value of λ. We tested 20 models using logarithmically spaced lambda values between 1e-6 - 1e2 and chose the model with a λ value that minimized the mean L2 error (Marcus *et al.* 2021).

*Environmental Data:* We collected environmental datasets on climate, soils, and land use (Table S2; See Supplement for more detailed descriptions of environmental variables). All environmental data were downloaded in or re-projected into WGS84 with a resolution of 30 arcseconds (~1km resolution). For climate, we focused on three bioclimatic variables from the WorldClim dataset (worldclim.org) that are commonly used in ecological niche modeling (Chapman *et al.* 2019; Briscoe Runquist *et al.* 2021). These variables have been shown to be relevant to plant establishment and growth (Wolkovich *et al.* 2012; Petitpierre *et al.* 2017; Chapman *et al.* 2017; Gorton *et al.* 2019): mean temperature of the warmest quarter (Bio 10), minimum temperature of the coldest month (Bio 6), and precipitation in the warmest quarter (Bio 18) (Fig. S8).

For soils, we downloaded data from the International Soil Reference Information Center (https://files.isric.org/soilgrids/latest/data\_aggregated/1000m/) and USDA NRCS Soils gNATSGO database (https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/). Due to the large number of potential soil variables, we performed separate PCAs for soil composition and chemistry variables in order to produce composite variables that describe major axes of variation: soil composition (variables: bulk density, coarse fragment volume, volumetric water content at 33kPa, available water storage, and clay, silt, and sand composition; Fig. S9; Table S3) and soil chemistry (variables: cation exchange capacity, nitrogen content, soil organic carbon content, soil organic carbon density, and pH of soil water; Fig. S10; Table S3). We retained the first two PCs from each PCA [Soil Comp PC1 (48.4% variation), Soil Comp PC2 (20.5% variation), Soil Chem PC1 (36.6% variation), and Soil Chem PC2 (29.1% variation); (Fig. S10)]. We included soil drainage classification (7 classification categories ranging from very poorly drained to excessively drained; Fig. S11 and Table S4).

To characterize anthropogenic factors, we downloaded a land use classification map (15 classification categories; USGS National Land Cover Database for 2019; Fig. S12; Table S5) and the percent impervious surface coverage (Multi-resolution Land Characteristic Consortium,https://www.mrlc.gov/).

*Isolation-by-Environment:* We used the environmental data to generate environmental resistance surfaces with the R package ‘ResitanceGA’ (Peterman 2018). This package uses a genetic algorithm to optimize the relationship between genetic chord distances and a resistance transformation of the environmental data. More specifically, the package first generates multiple resistance models using non-linear transformations (e.g. monotonic or Ricker model) of the environmental distance data (i.e. generate putative environmental resistance functions) and then calculates their corresponding estimates of landscape resistance using CircuitScape implemented in Julia (Anantharaman *et al.* 2020). The models are allowed to compete and the best performing models, as determined by AICc, are kept and parameters are mutated for the next iteration (e.g. small perturbations to the original model parameter combinations). The algorithm runs until 25 consecutive iterations do not yield significant improvement in model fit and the optimized parameter estimates of the best model are then used for further analyses.

To test for isolation-by-environment, we optimized resistance surfaces for each environmental layer individually in ‘Resistance GA’ analyses. Each run of ‘ResistanceGA’ also estimates a null geographic distance model. To examine whether particular environmental variables generated IBE, we compared the optimized environmental resistance model to the null model (using ∆AICc). This allowed us to determine whether the inclusion of environmental resistance significantly improved model fit. We projected the resulting resistance model surfaces and compared model metrics for each environmental variable.

To determine which environmental resistance variables were the most strongly associated with genetic distance, we used Multiple Regression on Distance Matrices (MRDM) (Legendre *et al.* 1994; Balkenhol *et al.* 2009). We included resistance matrix outputs from optimized CircuitScape models (i.e. for individual environmental models) as independent variables and genetic chord distance as the dependent variable. We included geographic distance as a null independent variable (Dyer *et al.* 2010). We standardized all optimized resistance layers to mean zero and standard deviation one. Evaluating resistance layers using Circuitscape and MRDM determines which environmental factors best explain patterns of genetic distance across the landscape and potentially drive isolation-by-environment (Peterman & Pope 2021). Models were run using the ‘MRM’ function in the package ‘ecodist’ (Goslee & Urban 2007).

While useful for investigating IBE, MRDM analyses do not necessarily describe the realized pathways of movement across a landscape, which may depend on multiple and/or nonlinear interactions among environmental factors that are not included in the individual environmental ResistanceGA optimizations. To determine the overall landscape patterns of movement, we also optimized a multi-environmental factor resistance surface model that included all environmental layers in a ResistanceGA model and allowed for interactions among environmental variables in the non-linear transformations. The resulting resistance matrix is most likely to describe composite environmental IBE and how pathways of movement are influenced by the aggregate effects of environmental variables. Due to the large number of environmental variables included in this model, we executed three independent runs of ResistanceGA to check for model stability and averaged across individual results.

*Ecological Niche Analysis:*

We tested for niche differentiation between clusters identified by the spatially-explicit conStruct results using the ‘ecospat’ package (Di Cola *et al.* 2017). We defined the available environmental background space using 1000 randomly-generated points drawn from a spatial polygon of Minnesota. We investigated niche differences for the full set of environmental variables (climate, soil, and anthropogenic, *heretofore*: aggregate environmental variables) and for climate variables, alone. For the aggregate environmental variables, we conducted an Hill-Smith ordination (ordination of tables mixing quantitative variables and factors) (Hill & Smith 1976) for background and sample points using the function ‘dudi.hillsmith’ from the ‘ade4’ package. We retained the first three axes of variation for niche analyses. For the climate variables, we conducted a principal component analysis using the ‘dudi.pca’ function from the package ‘ade4’ and retained the first two axes.

We quantified four measures of niche differentiation. We quantified niche overlap using Schoener’s D, which accounts for both breadth and density and varies between 0 and 1 (zero and complete niche overlap, respectively). We also calculated what proportion of the combined niche space represented niche stability, expansion, and unfilling. Niche stability is the proportion of niche that overlaps, niche expansion is the proportion of the cluster 2 niche that does not overlap cluster 1, and niche unfilling is the proportion of cluster 1 that does not overlap cluster 2. We tested for niche equivalency using a permutation test (n = 999) (i.e. within the niche space of common tansy, do randomly assigned clusters overlap less than would be expected by chance?). All tests were one-tailed. We tested the hypothesis that overlap, stability, and unfilling were less than random expectation and that expansion was greater than random expectation.

**RESULTS**

*Landscape Genetic Analyses*:

*Isolation-by-distance:* We did not detect evidence of isolation-by-distance (IBD) using a parametric test (r = 0.06, P = 0.1) or when we used loess to explore the relationship at varying distances (Fig. 2A). Isolation-by-distance was highest at 108 km (r = 0.14, Fig. 2B); however, the IBD correlation coefficient was similar across the range of distances (r = 0.08 to 0.13 for distances of 150 to 660 km).

*Principal Components Analysis:* The first three axes of the PCA accounted for 7.6% of genetic variation (PC1: 3.6%, PC2: 2.1%, PC3: 1.9%). Axis 1 and Axis 3 had high spatial autocorrelation (Axis 1: Moran’s I = 0.48, *P* < 0.001; Moran’s I = 0.27, *P* < 0.001); whereas axis 2 had moderate spatial autocorrelation (Moran’s I = 0.10, *P* = 0.034).

The far northwestern and several northeastern samples separated from the rest along PC1 (Fig. 2C, 2E, & S13). Two samples from the northwest and two from the southwest separated from the remainder along PC2 (Fig. 2C, 2F, & S13). Northwestern and northeastern samples occupied largely overlapping but offset areas along PC3 (Fig. S14). Otherwise, we did not detect population structure based on subsequent axes.

*Spatial Principal Components Analysis:* We found significant global spatial structure: geographically proximal populations had more similar allele frequencies than random expectation (permuted statistic = 0.009, *P* = 0.01; Fig. S15). By contrast, we found no evidence for local spatial structure: neighboring populations did not have more dissimilar allele frequencies than random expectation (simulation statistic = 0.007, *P* = 0.85; Fig. S15).

The first three global sPCA axes (i.e. positive) were distinct from the rest in diagnostic plots (Fig. S16), represented substantial genetic variance, and had high spatial autocorrelation (Tables 1 & S6). Axis 1 separated most far northwestern samples (Red River Prairie/Aspen Parklands/Agassiz Lowlands) from the remaining samples (Fig. 2D & S13). Axis 2 separated most northeastern samples from the remainder, particularly those from the north shore of Lake Superior (Fig. 2D & S13). Axis 3 further supported the differentiation of far northwestern and northeastern samples from the remainder (Fig. S17).

We found a strong correlation between PC1 and sPC1 (r = -0.70; *P* < 0.001) and moderate correlations between PC2 and sPC3 (r = -0.35; *P* < 0.001) and PC3 and sPC2 (r = -0.46; *P* < 0.001). Correlations between all other combinations were weak (r < 0.12; P < 0.05; Table S7).

*ConStruct Analysis:* The conStruct analysis supported K=4 and K=2 as the most probable number of clusters for the non-spatial and spatial analyses, respectively (Figs. 3 & S18-S20). Both analyses revealed similar patterns of population structure as PCA and sPCA analyses (Fig. 3).

In the non-spatial analysis, far northwestern samples (Red River Prairie/Aspen Parklands/Agassiz Lowlands) were differentiated from the remaining samples; far northwestern samples were assigned primarily to one cluster and had low proportions of the other clusters (lilac cluster; Figs. 3A & 3B). The remainder of samples had more variable cluster proportions. The remaining northwestern and southwestern samples had higher proportions of a second cluster (burgundy cluster; Figs. 3A & 3B) compared to northeastern individuals, which had higher proportions of a third cluster (teal cluster; Figs. 3A & 3B). The remaining cluster was found in low proportion across the regions with some notable exceptions in individuals from the south (yellow cluster; Figs. 3A & 3B).

Spatial autocorrelation and continuous IBD in genetic samples may inadvertently suggest discrete population structure that is not a true representation of demographic processes. The spatial conStruct analysis can account for these continuous relationships while generating cluster assignments. In the spatial analyses, cluster assignments were similar to the non-spatial analysis but more clearly demarcated individuals from far northwestern MN. These samples were predominantly assigned to one cluster and had very low proportions of the other clusters (light blue cluster; Fig. 3C & 3D). There was also a small region of northeastern MN (North Shore Highlands/Laurentian Uplands) where individuals had higher proportions of this cluster. The remaining individuals were primarily assigned to the alternate cluster (dark berry cluster; Fig. 3C & 3D).

*Fast Estimated Effective Migration Surfaces:* FEEMS identified two areas with reduced effective migration (Fig. 4). There was a particularly pronounced area of reduced effective migration surrounding far northwestern populations. The area of lowest effective migration separated northwest populations from north-central populations (Aspen Parklands/Red River Prairie/Agassiz Lake from Chippawa Plains/Hardwood Hills; Fig. 4). This boundary aligned geographically with groupings identified by PCA, sPCA, and conStruct. In northeastern MN, there was reduced effective migration among a group of populations in the middle of the region (North Shore Highlands/Laurentian Uplands). Southern MN had greater variation compared to other areas of the state in effective migration rates. Conversely, effective migration was relatively high among populations in north-central and northeastern MN (Fig. 4).

The model with lambda = 0.1128 had the lowest L2 cross-validation error (Fig. S22) and was used to calculate and visualize the effective migration surface (Figs. 4 & S23). Additionally, we found models using a range of lambda values identified similar geographic patterns of reduced migration (Fig. S23).

*CircuitScape:* We calculated resistance surfaces for environmental factors individually and in combination. When analyzed individually, the inclusion of any environmental variable improved model performance compared to a model that included only distance. However, land classification, minimum temperature of the coldest month, precipitation of the warmest month, and soil chemistry PC1 resulted in the most substantial gains to model performance (∆AIC < -100 & marginal R2 > 0.20; Table S8). Including percent impervious surface in a resistance model resulted in moderate performance gains (∆AIC = -45.4 & marginal R2 = 0.07; Table S8). Geographic distance, alone, explained only 2% of the variation (marginal R2).

The following environmental factors had the greatest influence on genetic distance. Croplands, grasslands, and managed developed areas (e.g. parks) had the highest resistance values compared to pastures, low-to-moderately developed areas, and mixed forests, which had the lowest resistance values (Table S9, Fig. S24). For climate variables, areas with the coldest temperatures had higher resistance and areas with the lowest and highest precipitation had higher resistance (Table S9, Fig. S24). Areas of very high and low values of soil chemistry PC1 (high pH or high Nitrogen) had higher resistance (Table S9, Fig. S24). Last, areas with high percentages of impervious surface had less resistance to gene flow than areas with less impervious surface (Table S9, Fig. S24).

In the Multiple Regression on Distance Matrices analysis, genetic distance had a significant or nearly significant relationship with geographic distance, soil variables, and land use (Table S10). Higher resistance was associated with both soils that are poorly and excessively drained and soils that have higher organic matter. Highly cultivated areas also had very high resistance values.

In the multi-surface CircuitScape models, the composite resistance surfaces indicated less resistance to gene flow in northeastern and north-central Minnesota compared to the northwest (Figs. 4B & S25). Areas of low gene flow identified by CircuitScape were consistent with the FEEMS migration surface (Figs. 4A & 4B). Replicate CircuitScape model optimizations were largely consistent across the three independent runs. The three models explained between 12.0% - 26.4% of the variation in genetic distance (marginal R2). Across all three models, the environmental factors that consistently had the strongest explanatory relationships with gene flow were land use classification (24.7% - 46.4% contribution), soil drainage class (10.7% - 24.3%), and precipitation of the warmest quarter (7.8% - 27.5%) (Table S11).

*Ecological niche analysis:*

In the ordination of aggregate environmental variables (climate, soil, and anthropogenic), the first three axes accounted for approximately 30% of variation (Axis 1: 15.0 %; Axis 2: 7.9%; Axis 3: 5.8%). The first axis described the environmental gradation that distinguishes northern and southern regions of the state and was defined by temperature, soils drainage, and forest cover. Axis 2 primarily separated eastern and western regions of the state and was defined by differences in precipitation and forest cover. Axis 3 was primarily associated with the degree of development in an area (Fig. S26; See Supp. Mat. for more details). In the principal components analysis of climate variables, the first two principal component axes accounted for 96.5% of the variation (PC1: 61.6%; PC2: 35.0%). PC1 described mainly temperature; PC2 described mainly precipitation (Fig. S27).

The two main clusters identified by the spatially-explicit conStruct analysis occupy overlapping ecological niches for both the aggregate environmental variables and climate variables (Fig. 5). Based on aggregate environments, cluster 1 included environments that were colder and drier than those occupied by cluster 2 with moderately to poorly drained soils (Fig. 5A-C). Based on climate alone, cluster 1 occupied cooler and drier environments (Fig. 5D). The non-overlapping sections of the aggregate and climate niche for cluster 1 represented significant niche expansion when compared to cluster 2; permutation tests for Schoener’s D, niche stability and niche expansion all indicated differences from random expectation (Table 2). The niche differences we found are consistent with a niche shift concurrent with the timeline of range expansion into northwestern MN.

**DISCUSSION**

Limited knowledge exists regarding how environmental variation impacts invasion dynamics and population genetic diversity, particularly of plants and across landscapes (Cruzan & Hendrickson 2020). In this study, we took advantage of a recent and rapid invasive range expansion across stark environmental gradients to dissect the relative influences of distance and environmental factors on landscape genetic patterns. We used a dense sampling scheme to capture potential fine-scale genetic structure. We found that genomic variation did not follow a pattern of isolation-by-distance but instead was influenced more strongly by environmental variation that likely shaped dispersal during range expansion. Soil and land use were particularly influential on landscape genomic structure. Our results also suggested that the recent invasion has been accompanied by an expansion in the set of occupied environments. These results indicate that interacting environmental factors influenced fine-scale invasive spread. Further, these results suggests that invasive range shifts with climate change may be strongly affected by environmental gradients and that the management of future invasion will need to better account for complex environmentally dependent invasion dynamics.

There was little evidence of isolation-by-distance for our common tansy samples. The correlation between genomic and geographic distance was weak (r = 0.06). This pattern suggested that migration was not consistent with a stepping-stone model and that invasion did not always occur via geographically nearest neighbors. IBD is often reduced or absent in population genetic studies of invasive species (Slatkin 1993; Hutchison & Templeton 1999; Moyle 2006). For example, in a global analysis of the genus *Silene*, Moyle (2006) found that invasive populations showed no evidence for IBD but populations in their native range often showed evidence of IBD. A lack of IBD may occur due to historical introduction effects, rapid expansion, stochastic long-distance dispersal, or environmentally mediated migration (Duforet-Frebourg & Blum 2014; Bradburd & Ralph 2019) and all have been found in invasive species (Leblois *et al.* 2000; Dlugosch & Parker 2008; Schlaepfer *et al.* 2008; Lu *et al.* 2022; Lake *et al.* 2023). Due to the density and scale of our sampling, we were able to investigate IBD patterns at very fine spatial scales. We found elevated correlations for geographically close pairs (rkm=108 = 0.14; Fig. 2) and moderate variance across the region. These IBD patterns are consistent with limited localized dispersal among smaller geographically isolated patches but dominance of drift or environmentally mediated factors at larger scales (Hutchison & Templeton 1999). The PCA, sPCA, and clustering analyses were also consistent with these findings. In PCA and sPCA, geographically close genotypes were more similar but there was broad geographic overlap in ordination space and in conStruct analysis, cluster assignment proportions followed the same general pattern of geographic similarity with broad overlap (Figs. 2 & 3). Additionally, FEEMS identified fine-scale geographic pockets of high effective migration separated by distinct areas of low effective migration (Fig. 4). Together, these results indicate that the recent invasion was likely multifactorial and scale dependent.

Our results indicated the environmental factors helped shape the recent invasion of common tansy in Minnesota. Although research involving IBE of invasive plants has received less study to date (Etherington 2015; Cruzan & Hendrickson 2020; but see Alvarado-Serrano *et al.* 2019), our findings are consistent with a number of recent studies of IBE in invasive species (Ray & Ray 2014; Alvarado-Serrano *et al.* 2019; Yang *et al.* 2020; Hofmeister *et al.* 2021). In these studies, climate and land use most often influenced spatial patterns of population genetic structure. However, many of the studies assessed vagile species or populations spaced over larger geographical areas. Additionally, many of the previous studies focused on how patterns of IBE informed potential adaptive evolution across their range and were less focused on how environmental factors modulated range expansion. Our study, using fine-scale sampling on a recent range expansion, provides insight on how environmental factors directly impacted dispersal and invasion dynamics (Sexton *et al.* 2014; Wang & Bradburd 2014; Bradburd *et al.* 2018). Our study indicated that IBE was likely a potent factor during the phase of range expansion as all environmental variables assessed had stronger relationships with genetic distance than geographic distance alone (Table S8). Identification of the environmental factors and the scales over which they operate will allow for better identification of regions that are likely to promote invasive spread and which areas or habitats are at highest risk of future invasion.

Soil variables and land use classification had the greatest influence on landscape genetic patterns in common tansy, but their influence differed in scale across the state. In a MLRM analysis that included all CircuitScape transformed environmental variables land use classification, soil variables, and geographic distance best explained genetic distances. Additionally, all CircuitScape optimizations that included all environmental variables also always identified land use classification and soils as important for pathways of dispersal. The patterns of resistance for land use classification echoed large-scale patterns of recent range expansion (Fig. S24). The lowest resistance to gene flow was found in disturbed/barren, moderately developed, and pastures of northeastern and central MN. Conversely, the highest areas of resistance were found in heavily cultivated and grassland ecosystems found in southern and western regions of the state. From this, we discern that invasion was likely facilitated by disturbed or barren areas and pastures along corridors between the outlying areas of towns and cities. Soils were also an important determinant of IBE especially at smaller geographic scales. This result may be unsurprising given that soil characteristics play an important part in the microenvironment and may affect local colonization by plants. Soil chemistry, composition, and drainage contributed to lower resistance to gene flow in northeastern MN, particularly along the northern border where there were continuous corridors of lower resistance. The same set of variables also contributed to localized areas of higher resistance to gene flow in northwestern and southern areas of the state. Investigating the mechanisms underlying the differences in resistance, particularly among different agricultural land uses, may help management efforts to stem further range expansion. Furthermore, land use changes concomitant with globalization may aid invasion and should be accounted for in predictions of invasion risk. Additionally, ecological studies of the effect of edaphic conditions on colonization and establishment will aid in identifying the habitats most at risk within already invaded regions.

Several analyses identified a genetically distinct group of samples from far northwestern MN (Aspen Parklands/Red River Prairie and a portion of the Agassiz Lowlands). These samples separated from other regions along both PC1 and sPCA axis 1, had high admixture proportions of a single genetic cluster assignment, and were separated geographically by an area of reduced effective migration in the FEEMS analysis (Figs. 2-4). These samples are from an expanding range margin that is still sparsely populated, and the vast majority of occurrence records are dated from 2010 to present (EDDMapS 2023). Land managers from this region have also noticed a very recent increase in westward expansion. Genetically, the samples from this area are most like a small group of samples from the center of the northeastern region and along the northern border (i.e. ice blue cluster in spatial conStruct analysis; Fig. 3). The clustering and IBE results suggest that colonization did not occur via the most direct geographic path and likely occurred though dispersal along more favorable environments. In addition, the realized niche of the conStruct cluster that included the far northwestern populations occupied a slightly different portion of aggregate and climate environmental space that is outside of the historical environmental niche of common tansy. This result indicates that invasion risk may be underpredicted for many habitats, particularly if they were developed based on ENMs that do not account for adaptation. Establishment and persistence of these populations in novel environments may also indicate adaptation at the range margin that could further drive range expansion. Experimental work is needed to determine the nature of potential adaptive evolution and its implications for continued invasion risk in MN and across the US.

Isolation-by-environment was more likely to explain landscape genomic differentiation of common tansy indicating that environmental factors helped shape recent and rapid range expansion. Our results add to a growing body of research showing that invasion often does not occur strictly via wave-like spread but is modulated by climate, land use, and soils. Using fine-scale regional sampling, we were able to demonstrate that environmentally mediated gene flow was also scale-dependent with different environmental factors influencing movement at different scales. Long-term management will need to account for these complex interactions of environmental factors to halt further range expansion especially in the face of continued global change. Moreover, the results indicate that adaptation to leading edge environments may facilitate range expansion and that predictive models of future habitat suitability should account for such adaptive differentiation.

**ACKNOWLEDGEMENTS**

We thank T. Lake for help with sample collection and B. Greene and D. Schoenecker for help with sample organization and preservation. We thank T. Lake, R. Venette, Y. Brandvain, and L. Flagel for thoughtful conversations about analyses. We thank the natural resources community for helpful conversations about common tansy invasion including G. Foltz (Clay Co. SWCD), M. Watland (Becker Co. SWCD), MN Dept. of Natural Resources, in particular, A. Danielson and S. Pennington and L. Van Riper, and MN Dept. of Agriculture, in particular T. Cortilet. Funding for this project was provided by the Minnesota Invasive Terrestrial Plants and Pests Center through the Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR).

**REFERENCES**

Abegaz F, Chaichoompu K, Génin E *et al.* (2019) Principals about principal components in statistical genetics. *Briefings in Bioinformatics*, **20**, 2200–2216.

Alvarado-Serrano DF, Van Etten ML, Chang S-M, Baucom RS (2019) The relative contribution of natural landscapes and human-mediated factors on the connectivity of a noxious invasive weed. *Heredity*, **122**, 29–40.

Anantharaman R, Hall K, Shah VB, Edelman A (2020) Circuitscape in Julia: High performance connectivity modelling to support conservation decisions. *JuliaCon Proceedings*, **1**, 58.

Atwater DZ, Barney JN (2021) Climatic niche shifts in 815 introduced plant species affect their predicted distributions. *Global Ecology and Biogeography*, **30**, 1671–1684.

Atwater DZ, Sezen UU, Goff V *et al.* (2016) Reconstructing changes in the genotype, phenotype, and climatic niche of an introduced species. *Ecography*, **39**, 894–903.

Baker HG, Stebbins GL (eds) (1965) *The genetics of colonizing species*. Academic Press, New York.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818–830.

Barnes R, Sahr K (2021) *dggridR: Discrete Global Grids*. https://github.com/r-barnes/dggridR/

Barrett SCH (2015) Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology*, **24**, 1927–1941.

Bougeard S, Dray S (2018) Supervised Multiblock Analysis in R with the ade4 Package. *Journal of Statistical Software*, **86**, 1–17.

Bradburd G (2019) *conStruct: Models Spatially Continuous and Discrete Population Genetic Structure. R package version 1.0. 4*.

Bradburd GS, Coop GM, Ralph PL (2018) Inferring Continuous and Discrete Population Genetic Structure Across Space. *Genetics*, **210**, 33–52.

Bradburd GS, Ralph PL (2019) Spatial Population Genetics: It’s About Time. *Annual Review of Ecology, Evolution, and Systematics*, **50**, 427–449.

Bradburd G, Ralph P, Coop G (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, **67**, 3258–3273.

Briscoe Runquist RD, Lake TA, Moeller DA (2021) Improving predictions of range expansion for invasive species using joint species distribution models and surrogate co‐occurring species. *Journal of Biogeography*, **48**, 1693–1705.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3* , **1**, 171–182.

Chapman D, Pescott OL, Roy HE, Tanner R (2019) Improving species distribution models for invasive non‐native species with biologically informed pseudo‐absence selection. *Journal of Biogeography*, **46**, 1029–1040.

Chapman DS, Scalone R, Štefanić E, Bullock JM (2017) Mechanistic species distribution modeling reveals a niche shift during invasion. *Ecology*, **98**, 1671–1680.

Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I : One-table methods. *R News*, **4**, 5-10.

Clasen B, Moss N, Chandler M, Smith A (2011) A preliminary genetic structure study of the non-native weed, common tansy (Tanacetum vulgare). *Canadian Journal of Plant Science*, **91**, 717–723.

Clements DR, Ditommaso A (2011) Climate change and weed adaptation: can evolution of invasive plants lead to greater range expansion than forecasted? *Weed Research*, **51**, 227–240.

Clements DR, Jones VL (2021) Rapid Evolution of Invasive Weeds Under Climate Change: Present Evidence and Future Research Needs. *Frontiers in Agronomy*, **3**, https://doi.org/10.3389/fagro.2021.664034.

Cox TJ, Durrett R (2002) The stepping stone model: New formulas expose old myths. *The Annals of Applied Probability*, **12**, 1348–1377.

Cruzan MB, Hendrickson EC (2020) Landscape Genetics of Plants: Challenges and Opportunities. *Plant Communications*, **1**, 100100.

DeMarche ML, Doak DF, Morris WF (2019) Incorporating local adaptation into forecasts of species’ distribution and abundance under climate change. *Global Change Biology*, **25**, 775–793.

Di Cola V, Broennimann O, Petitpierre B *et al.* (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, **40**, 774–787.

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

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.

Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.

Dray S, Dufour A, Chessel D (2007) The ade4 package-II: Two-table and K-table methods. *R News*, **7**, 47-52.

Duforet-Frebourg N, Blum MGB (2014) Nonstationary patterns of isolation-by-distance: inferring measures of local genetic differentiation with Bayesian kriging. *Evolution*, **68**, 1110–1123.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.

EDDMapS. 2023. Early Detection & Distribution Mapping System. *The University of Georgia - Center for Invasive Species and Ecosystem Health*. Available online at http://www.eddmaps.org/; last accessed November 10, 2023.

Etherington TR (2015) Geographical isolation and invasion ecology. *Progress in Physical Geography: Earth and Environment*, **39**, 697–710.

Frankham R (2005) Resolving the genetic paradox in invasive species. *Heredity*, **94**, 385.

Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W (2012) Invasive species distribution models - how violating the equilibrium assumption can create new insights: Beyond the equilibrium assumption of SDMs. *Global Ecology and Biogeography*, **21**, 1126–1136.

Gorton AJ, Tiffin P, Moeller DA (2019) Does adaptation to historical climate shape plant responses to future rainfall patterns? A rainfall manipulation experiment with common ragweed. *Oecologia*, **190**, 941–953.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Guisan A, Petitpierre B, Broennimann O, Daehler C, Kueffer C (2014) Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution*, **29**, 260–269.

Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, **104**, 19926–19930.

Hill MO, Smith AJE (1976) Principal Component Analysis of Taxonomic Data with Multi-State Discrete Characters. *Taxon*, **25**, 249–255.

Hofmeister NR, Werner SJ, Lovette IJ (2021) Environmental correlates of genetic variation in the invasive European starling in North America. *Molecular Ecology*, **30**, 1251–1263.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.

Jacobs J (2008) *Ecology and Management of Common Tansy (Tanacetum vulgare L.)*. Invasive Species Technical Note No. MT-18. USDA Natural Resources Conservation Service.

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

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , **27**, 3070–3071.

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kawecki TJ (2008) Adaptation to Marginal Habitats. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 321–342.

Keskitalo M, Linden A, Valkonen JPT (1998) Genetic and morphological diversity of Finnish tansy. *TAG. Theoretical and applied genetics*, **9**, 6–1141.

Kimura M (1964) Diffusion models in population genetics. *Journal of Applied Probability*, **1**, 177–232.

Kimura M, Weiss GH (1964) The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, **49**, 561–576.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

Lake TA, Briscoe Runquist RD, Flagel LE, Moeller DA (2023) Chronosequence of invasion reveals minimal losses of population genomic diversity, niche expansion, and trait divergence in the polyploid, leafy spurge. *Evolutionary Applications*, **16**, 1680-1696.

Lake TA, Briscoe Runquist RD, Moeller DA (2020) Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. *Diversity and Distributions*, **26**, 1767–1779.

Lakoba, V. T., Atwater, D. Z., Thomas, V. E., Strahm, B. D., & Barney, J. N. (2021). A global invader’s niche dynamics with intercontinental introduction, novel habitats, and climate change. *Global Ecology and Conservation*, **31**, e01848. https://doi.org/10.1016/j.gecco.2021.e01848

Leblois R, Rousset F, Tikel D, Moritz C, Estoup A (2000) Absence of evidence for isolation by distance in an expanding cane toad (Bufo marinus) population: an individual-based analysis of microsatellite genotypes. *Molecular Ecology*, **9**, 1905–1909.

LeCain R, Sheley R (2014) Common tansy. *Tanacetum vulgare*, Montana State University Extension: MontGuide MT199911AG.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.

Legendre P, Lapointe F-J, Casgrain P (1994) Modeling brain evolution from behavior: A permutational regression approach. *Evolution*, **48**, 1487–1499.

Lu S, Luo X, Han L *et al.* (2022) Genetic patterns reveal differences between the invasion processes of common ragweed in urban and non-urban ecosystems. *Global Ecology and Conservation*, **38**, e02214.

Mack RN (2003) Plant Naturalizations and Invasions in the Eastern United States: 1634-1860. *Annals of the Missouri Botanical Garden*, **90**, 77–90.

Malinsky M, Trucchi E, Lawson DJ, Falush D (2018) RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, **35**, 1284–1290.

Marcus J, Ha W, Barber RF, Novembre J (2021) Fast and flexible estimation of effective migration surfaces. *eLife*, **10**, e61927.

Matula DW, Sokal RR (2010) Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. *Geographical Analysis*, **12**, 205–222.

Mitich LW (1992) Tansy. *Weed Technology*, **6**, 242–244.

Mouquet N, Lagadeuc Y, Devictor V *et al.* (2015) Predictive ecology in a changing world. *The Journal of Applied Ecology*, **52**, 1293–1310.

Moyle LC (2006) Correlates of genetic differentiation and isolation by distance in 17 congeneric Silene species. *Molecular Ecology*, **15**, 1067–1081.

MN Dept. of Agriculture: Common Tansy (2023) Common Tansy. Minnesota Department of Agriculture. https://www.mda.state.mn.us/plants/pestmanagement/weedcontrol/noxiouslist/commontansy

MN Dept. of Natural Resources: Common Tansy (2023) Common Tansy. Minnesota Department of Natural Resources. https://www.dnr.state.mn.us/invasives/terrestrialplants/herbaceous/commontansy.html

MN Dept. of Natural Resources Ecological Classification System (2023) Ecological Classification System. https://www.dnr.state.mn.us/ecs/index.html

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, **40**, 646–649.

Nychka D, Furrer R, Paige J, Sain S (2021) *fields: Tools for spatial data*. University Corporation for Atmospheric Research, Boulder, CO, USA.

Oduor AMO, Leimu R, van Kleunen M (2016) Invasive plant species are locally adapted just as frequently and at least as strongly as native plant species. *The Journal of Ecology*, **104**, 957–968.

Peter BM, Slatkin M (2013) Detecting range expansions from genetic data. *Evolution*, **67**, 3274–3289.

Peterman WE (2018) ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, **9**, 1638–1647.

Peterman WE, Pope NS (2021) The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology*, **30**, 37–47.

Peterson AT, Vieglais DA (2001) Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem. *Bioscience*, **51**, 363–371.

Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, **26**, 275–287.

Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, **48**, 94–100.

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

Purcell S (2019) *PLINK v. 1.9*. http://pngu.mgh.harvard.edu/purcell/plink/

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.

Putra AR, Hodgins KA, Fournier-Level A (2023) Assessing the invasive potential of different source populations of ragweed (*Ambrosia artemisiifolia* L.) through genomically informed species distribution modelling. *Evolutionary Applications*, **17**, e13632. https://doi.org/10.1111/eva.13632

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.

Ray A, Ray R (2014) Rapid divergence of ecotypes of an invasive plant. *AoB Plants*, **6**.

Reich D, Price AL, Patterson N (2008) Principal component analysis of genetic data. *Nature Genetics*, **40**, 491–492.

Roberts T (1878) Tanacetum vulgare. Bell Atlas Detailed Collection Record Information. University of Minnesota Bell Museum Herbarium. https://bellatlas.umn.edu/collections/individual/index.php?occid=158375

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, **28**, 4737–4754.

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

Savary P, Foltête J-C, Moal H, Vuidel G, Garnier S (2020) graph4lg: a package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, **12**, 538-547.

Schlaepfer DR, Edwards PJ, Widmer A, Billeter R (2008) Phylogeography of native ploidy levels and invasive tetraploids of Solidago gigantea. *Molecular Ecology*, **17**, 5245–5256.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, **191**, 171–181.

van Strien MJ, Holderegger R, Van Heck HJ (2015) Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, **114**, 27–37.

Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, **61**, 334–345.

Thioulouse J, Dray S, Dufour A-B *et al.* (2018) *Multivariate Analysis of Ecological Data with ade4*. Springer.

Tomiolo S, Ward D (2018) Species migrations and range shifts: A synthesis of causes and consequences. *Perspectives in Plant Ecology, Evolution and Systematics*, **33**, 62–77.

USDA Plants Database (2023) Tanacetum vulgare L. https://plants.usda.gov/home/plantProfile?symbol=TAVU

Veloz SD, Williams JW, Blois JL *et al.* (2012) No‐analog climates and shifting realized niches during the late quaternary: implications for 21st‐century predictions by species distribution models. *Global Change Biology*, **18**, 1698–1713.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649–5662.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

White DJ (1997) Tanacetum vulgare L: weed potential, biology, response to herbivory, and prospects for classical biological control in Alberta. University of Alberta, Department of Entomology Ph.D. Thesis.

White DJ (2001) Tanacetum vulgare L., common tansy (Asteraceae). *Biological Control Programmes in Canada, 1981-2000*, 425–426.

Wolf VC, Gassmann A, Clasen BM, Smith AG, Müller C (2012) Genetic and chemical variation of Tanacetum vulgare in plants of native and invasive origin. *Biological control*, **61**, 240–245.

Wolkovich EM, Cook BI, Allen JM *et al.* (2012) Warming experiments underpredict plant phenological responses to climate change. *Nature*, **485**, 494–497.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–138.

Yang F, Liu N, Crossley MS *et al.* (2020) Cropland connectivity affects genetic divergence of Colorado potato beetle along an invasion front. *Evolutionary Applications*, **39**, 920.

Yates KL, Bouchet PJ, Caley MJ *et al.* (2018) Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, **33**, 790–802.

**DATA ACCESSIBILITY AND BENEFIT-SHARING**

**Data Accessibility Statement**

Raw sequence reads are deposited in the SRA (BioProject XXX – will be uploaded at publication acceptance)

Individual genotype data are archived and available through DOI: XXX (UMN DRUM – will be uploaded at publication acceptance)

All additional data, metadata and scripts are available at github/rdbrunquist/common-tansy-landscape-genomics

**Benefit-Sharing Statement**

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

**AUTHOR CONTRIBUTIONS**

RBR and DM conceived of and designed the research study. RBR collected samples and performed the research. RBR performed the analyses and drafted the paper. RBR and DM both contributed substantially to the editing and final version of the paper.

**CONFLICT OF INTEREST DISCLOSURE**: The authors declare no conflicts of interest

**TABLES**

Table 1. Eigenvalues of the first three global sPCA axes and their decomposition into genetic variance and Moran’s I values.

|  |  |  |  |
| --- | --- | --- | --- |
| Axis (λ) number | Eigenvalue | Genetic Variance | Moran’s I |
| λ1 | 4.0 | 24.3 | 0.66 |
| λ2 | 2.0 | 10.4 | 0.76 |
| λ3 | 1.8 | 10.6 | 0.67 |

Table 2. Environmental niche equivalency tests for aggregate environmental variables and climatic variables

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Aggregate environmental niche | | | | | | Climatic niche | |
| Metric | statistic  (Axes 1 & 2) | *P* | statistic  (Axes 1 & 3) | *P* | statistic  (Axis 2 & 3) | *P* | statistic (Axes 1 & 2) | *P* |
| Schoener’s D | **0.47** | **0.008** | 0.69 | 0.63 | 0.40 | 0.08 | **0.39** | **0.004** |
| Stability | **0.82** | **0.002** | 0.99 | **0.81** | **0.81** | **0.002** | **0.85** | **0.03** |
| Expansion | **0.18** | **0.002** | 0.00 | **0.81** | **0.19** | **0.002** | **0.14** | **0.03** |
| Unfilling | 0.25 | 0.902 | 0.12 | 0.51 | 0.24 | 0.876 | 0.34 | 0.98 |

**FIGURE LEGENDS**

**Figure 1.** A ) Distribution of common tansy in North America based on occurrence records from EDDMaps and GBIF. Grid cells with darker red coloration have more occurrences whereas grid cells with yellow coloration have very few occurrences. The dashed box outline indicates the study region centered on Minnesota. B) Distribution of common tansy in Minnesota and Wisconsin. Sampled populations are colored by MN ecological subsection. The subsections are divided roughly into four geographic quadrants: northwest, northeast, southwest, and southeast. The remaining occurrence records from EDDMaps and GBIF are shown as light gray dots. C) Photo of flowering common tansy plant with an inset of a cluster of inflorescences. D) Timeline of common tansy records in Minnesota and Wisconsin (See Supplementary Materials for more detailed methods). Grid cell coloration indicates the cumulative number of records up through the year indicated. Yellow indicates very few records (<10) ; grid cells with red colors indicate 100’s to 1000’s of records.

**Figure 2.** A) Relationship of genetic distance (Nei’s D) and geographic distance for pairs of populations to test for isolation-by-distance. Fit is shown by loess. B) The Mantel correlation statistic between population pairs for genetic distance and geographic distance when (y-axis) calculated using an expanding window of distances to calculate correlation. C) Principal Components Analysis biplot of the first two2 PCs. Samples are colored by the quadrant of MN from which they originated: northeast (dark berry-purple), northwest (light blue), southeast (light berry-pink), and southwest (dark blue). D) Spatial Principal Components Analysis biplot of the first two positive sPCA axes. Colors are the same as for PCA. E-H) Geographic pattern of PCs and sPCA axes. Dark blue indicates lower values and light yellow indicates higher values. E) PC1. F) PC2. G) sPC1. H) sPC2.

**Figure 3**. Non-spatial and spatial conStruct analyses. A-B) K = 4 from the non-spatial conStruct analysis. Clusters 1, 2, 3, and 4 are shown in lilac, burgundy, teal, and yellow, respectively. C-D) K = 2 from the spatially-explicit conStruct analysis. Clusters 1 and 2 are shown in light blue and dark berry-purple, respectively A & C) Spatial view of cluster assignment where each pie chart shows the fraction of each cluster. B & D) STRUCTURE barplot view of conStruct assignments. Each bar represents a sample and the bar proportion represents the cluster assignment fraction. Bars are clustered by hierarchical geography, first within their MN quadrant and then within the ecological subsection.

**Figure 4.** Effective migration and resistance to gene flow of common tansy across Minnesota. A) Fast Estimation of Effective Migration Surface (FEEMS) of common tansy in Minnesota. White points represent nodes with assigned samples and edges denote potential paths of travel. Edges colored blue have higher effective migration and edges colored brown have less effective migration. B) Multivariable Environmental Resistance surface estimated using CircuitScape. The projection is the average of three model runs. Darker purple colors indicate greater resistance values; lighter cream and orange colors indicate lower resistance. Small black circles represent sampling sites.

**Figure 5.** Ecological niche of common tansy genetic clusters. A-C) Niche space based on all environmental variables used in the paper. Dark Berry color represents the exclusive niche space of cluster 1; ice blue represents the exclusive niche space of cluster 2; yellow represents the niche space overlap for cluster 1 and cluster 2. The gray outline surrounds the total available niche space in Minnesota. A) Biplot of niche space for mixed ordination axes 1 and B) Biplot of niche space for mixed ordination axes 1 and 3. C) Biplot of niche space for mixed ordination axes 2 and 3. D) Principal components climatic niche space. Burgundy represents the exclusive niche space of cluster 1; teal represents the exclusive niche space of cluster 2; lilac represents the niche space overlap for cluster 1 and cluster 2. The gray outline surrounds the total available niche space in Minnesota.