Landscape connectivity of a noxious invasive weed: promises and challenges of landscape genomics for knowledge-based weed management?

Diego F. Alvarado-Serrano (corresponding author)
Department of Ecology and Evolutionary Biology
University of Michigan
E-mail: dalvarad@umich.edu

Megan Van Etten
Department of Ecology and Evolutionary Biology
University of Michigan
E-mail: mvanette@umich.edu

Shu-Mei Chang
Department of Ecology and Evolutionary Biology
University of Michigan
E-mail: smchang@uga.edu

Regina S. Baucom
Department of Ecology and Evolutionary Biology
University of Michigan
E-mail: rsbaucom@umich.edu

Landscape connectivity of a noxious invasive weed: promises and challenges of landscape genomics for knowledge-based weed management?

ABSTRACT

Examining how the landscape may influence gene flow is at the forefront of understanding population differentiation and adaptation. Such understanding is crucial in light of ongoing environmental changes and the elevated risk of ecosystems alteration. In particular, knowledge of how humans may influence the structure of populations is imperative to allow for informed decisions in management and conservation. Here we characterize the population genetic structure of *Ipomoea purpurea*, a noxious invasive weed, and assess the interaction between natural and human-driven landscapes on genetic differentiation. By combining rigorous statistical analyses and different molecular markers (nuclear microsatellites and a genome-wide panel of SNPs), we detect both common and marker-specific patterns of genetic connectivity and identify human population density as an important predictor of pairwise population differentiation, suggesting that the agricultural and/or horticultural trade may be involved in maintaining some level of connectivity across distant agricultural fields. Climatic variation appears as an additional predictor. We discuss the implications of these results and the approach we followed in the context of understanding agricultural weed and invasive species' connectivity, as well as the challenges and promises of current landscape genomics research for knowledge-based weed management.

Keywords: agricultural weeds, human-aided migration, landscape genetics, morning glory, population structure

INTRODUCTION

Elucidating routes and levels of migration between subpopulations of a species is essential to understand the interplay between gene flow, adaptation, genetic drift, and selection, and hence the forces that shape its evolutionary trajectory (Barrowclough, 1980; Slatkin, 1985). Landscape features—such as rivers, mountain ranges, crop fields, and urban areas—can impact levels of gene flow between populations by determining dispersal rates and routes (Cushman, McKelvey, Hayden, & Schwartz, 2006; McRae, 2006) as well as by influencing the likelihood of successful establishment of immigrants (Nosil, Egan, & Funk, 2008; Sexton, Hangartner, & Hoffmann, 2014; Wang & Bradburd, 2014). Landscape features can also indirectly condition the effect of gene flow through its effect on local effective population sizes since the actual role that migration plays in the evolution of a species is driven by the fraction of the local population size that correspond to immigrants (Slatkin, 1985; Wright, 1949). Consequently, the landscape, loosely defined as an area with spatially variable biotic and abiotic factors (Holderegger, Buehler, Gugerli, & Manel, 2010), creates the stage for spatially variable levels of effective gene flow among populations, conditioned by species' specific physiological tolerances and behavioral preferences (Clobert, Baguette, Benton, Bullock, & Ducatez, 2012). In this way, the landscape plays a pivotal role in the evolution of species.

In contrast to species that depend almost exclusively on natural dispersal agents, species in heavily human-dominated ecosystems may exploit human activities to maintain gene flow among populations and expand their ranges (Everman & Klawinski, 2013; Fountain, Duvaux, Horsburgh, Reinhardt, & Butlin, 2014). Such species may be capable of maintaining population connectivity over vast geographic ranges (Trakhtenbrot, Nathan, Perry, & Richardson, 2005) by overcoming landscape features that would otherwise represent natural barriers and reach dispersal

distances that could be orders of magnitude greater than those attained under natural agents or do it under much smaller time frames (Mack & Lonsdale, 2001; Ricciardi, 2007). In this way, by facilitating dispersal humans have the potential to: i) condition the balance between drift and selection (Lenormand, 2002; Slatkin, 1985), ii) introduce relevant genetic variation to local populations (Kolbe et al., 2004), iii) prevent local extinction or favor recolonization (Fountain et al., 2014), and alter the overall genetic constitution of populations (Bataille, Cunningham, Cruz, Cedeño, & Goodman, 2011). Human-aided migration—intentional or unintentional—is particularly prevalent in plants (Auffret & Cousins, 2013; Hodkinson, Thompson, Journal, & Dec, 2007; Wichmann et al., 2009), where it has had major impacts on the distribution of species and stability of communities (Simberloff, 2013 and references therein). Yet, the open question remains: how might the interaction between human-dominated and more natural landscapes affect population connectivity in plant populations, especially in human-exploiter species?

A particularly amenable system to study the interaction between natural and human-aided dispersal comes from agricultural weed populations. Agricultural weeds face a highly dynamic landscape characterized by frequent spatial rearrangements (expansion of agricultural front, increased fragmentation) and a constantly changing environment (crop rotation, agricultural chemical use, climatic abnormalities) (Meehan, Werling, Landis, & Gratton, 2011; Menchari, Délye, & Le Corre, 2007) that certainly impact their opportunities for survival and local adaptation through its effect on population connectivity (Margosian, Garrett, Hutchinson, & With, 2009). At the same time, natural features such as climate, soil type, and topography are expected to play a significant role in structuring populations provided intrinsic physiological requirements and species-specific traits (Cimalová & Lososová, 2009; Navas, 2012). Under these conditions, human-aided migration is expected to be critical for weeds' success (Epperson &

Clegg, 1986), but knowledge on how or if weedy plant populations are able to maintain connectivity through a complex landscape matrix of croplands, grasslands, natural and urban areas is limited. Addressing this limitation should not only improve our understanding of the underlying processes governing weeds' population structure, but should also offer practical tools to deal with this ever-growing agricultural problem that impose severe economic costs (on the order of 33B USD per year in US agriculture alone; Pimentel, Zuniga, & Morrison, 2005).

As a first step into investigating the interplay between natural factors and human activities on structuring genetic diversity in weed populations, we estimate the intensity and extent of migration from genetic data and, under the preliminary simplifying assumption of evolutionary equilibrium (Marko & Hart, 2011), evaluate how multiple landscape features influence genetic connectivity of a noxious agricultural weed, *Ipomoea purpurea*, using two different sets of molecular markers (nuclear microsatellites and a genome-wide panel of SNPs). Specifically, we ask the following questions: 1) Which natural and/or human-influenced landscape features—soils, elevation, climate, landcover, crop types, human population density—promote or constrain genetic connectivity between populations of this agricultural weed? and 2) what additional insights can we gain from a broader representation of the genome than traditionally used in landscape genetics studies (typically microsatellites and organelle DNA)? By considering the possible interactions between natural and human effects on migration, the answers to these questions offer deeper knowledge of the interaction between human activities, landscape features, and population structure of noxious weeds and hence contribute to improve effective management and control of these damaging plants. More generally, these answers contribute to deepen our understanding of the interaction between environmental setting and population differentiation, adaptation, and persistence (Taylor, Fahrig, Henein, & Merriam, 1993).

MATERIALS AND METHODS

Study system

Ipomoea purpurea, the common morning glory, is an agricultural weed evolving under the influence of human-driven and natural landscape factors. This species is a noxious weed, of horticultural value (Defelice, 2001; Fang et al., 2013), with a widespread distribution that includes highly heterogeneous landscapes in the Eastern, South- and Mid-western regions of the United States (Culpepper, 2006; Webster & Nichols, 2012). It is a self-compatible annual bumblebee-pollinated vine, with heavy seeds, and is found primarily in agricultural fields and disturbed areas (Baucom & Mauricio, 2008; Tiffin & Rausher, 1999), as well as cultivated flower gardens and yards (Defelice, 2001). While most details on the history of *I. purpurea* remain unknown, it is hypothesized that *I. purpurea* originated in Central America, from where it was taken to Spain to be grown in monasteries as an ornamental during the XVI century (Defelice, 2001; Fang et al., 2013). From there, it is hypothesized that its cultivation expanded to other European countries, including England, and later to North America (Defelice, 2001; Fang et al., 2013). By the early XVIII century it became a popular plant in gardens in the United States, but also a known weed (Defelice, 2001). Since then, little is know about the demographic history of this species in the United States, other than the fact that gene flow has probably been maintained over time at least among populations in relatively close geographic proximity (Kuester, Chang, & Baucom, 2015). What is clear is that since its introduction its history has been tightly linked to human activities, making it suitable to assess the impact of natural and anthropogenic landscapes on structuring genetic variance.

139 *Ipomoea purpurea* is currently one of the most problematic agricultural weeds (Webster
140 & Nichols, 2012) and is capable of infestations leading to substantial decline in crop (closely
141 related *Ipomoea* species cause declines of up to 80% of crop yield; Rogers, Murray, Verhalen, &
142 Claypool, 1996). This species exhibits resistance to the commonly used herbicide glyphosate
143 (Baucom & Mauricio, 2004, 2008), although the exact resistance level varies widely among
144 populations of this species (Kuester et al., 2015). This species is also a major concern for
145 conservation given its naturalization in multiple regions throughout the world and its
146 aggressiveness as an invasive (Chaney & Baucom, 2012; Fang et al., 2013). Hence, unraveling
147 the population structure of this species and how it is affected by the landscape should not only
148 improve our understanding of basic evolutionary processes, but should also inform practical
149 decisions for its management and control (e.g., Is herbicide resistance better controlled by
150 avoiding the spread of resistance genotypes or by local management of moderately isolated
151 populations?).
152
153 **Data compilation**
154 To capture the plausible effect of both natural and disturbed landscapes on structuring genetic
155 diversity in *I. purpurea*, we compiled a diverse set of GIS data for the continental US from a
156 variety of sources (Table S1). These data encapsulate human activities (human population
157 density, landcover, planted crops, and roads) as well as the geographical setting of *I. purpurea*
158 (elevation, climate—19 variables summarizing central tendencies and variability patterns in
159 temperature and precipitation, soil—8 variables summarizing the texture, pH, and organic and
160 inorganic content of the top 20cm of soil). We focused on both sets of data because of the
161 possible interaction between natural and human effects, which may lead to incorrect inferences if
162 not accounted for (e.g., spurious associations due to spatial correlation between crops distribution

163 and climate; Eberhardt & Teal, 2013). We first processed all these data into landscape layers at a
164 common spatial resolution of $10km^2$ and a common spatial extent around the US states with
165 available samples (Fig. 1). These spatial resolution and extent were chosen to maintain a practical
166 balance between scale and analytical manageability given available computational resources. To
167 reduce dimensionality, we opted to perform two separate Principal Component Analyses (PCAs)
168 on the 19 climatic and 8 soil layers, respectively. For all subsequent analyses we kept the
169 resulting first two principal components of each of these analyses, which accounted for over 78%
170 of the variance in each case, and primarily summarized temperature temporal gradients and
171 precipitation seasonality, and soils' pH, sandiness, and grain size, respectively (Table S2).
172
173 With the objective of estimating the genetic connectivity of populations of *I. purpurea*, we
174 compiled genetic data on an extensive panel of 15 previously optimized microsatellite loci
175 (Molecular Ecology Resources Primer Development Consortium 2013), which quality has been
176 verified by looking for scoring errors (Kuester et al., 2015) using Micro-Checker (Van
177 Oosterhout, Hutchinson, Wills, & Shipley, 2004). These data encompass a total of 597
178 individuals from 31 localities (with a minimum of 8 individuals per locality) (Fig. 1; Table S3).
179 All individuals were collected in 2012 from farms across the range of *I. purpurea* in the United
180 Sates (Kuester et al., 2015). In addition, to obtain a more comprehensive representation of the
181 genome of *I. purpurea* and assess the robustness of results in light of coalescent and mutational
182 variance (Nielsen & Slatkin, 2013), we generated a Next Generation Sequencing (NGS) dataset
183 from an additional set of individuals (from 6 localities represented in the SSR dataset, plus 2
184 additional localities in close geographic proximity to localities in the SSR dataset; Fig. 1).
185

186 To generate the NGS dataset, we constructed genome-wide Genotype By Sequencing (GBS)
187 libraries for 80 individuals sampled across the 8 localities. The GBS library was developed using
188 7ng of genomic DNA, extracted from leaf or cotyledon tissue, using SNPsaurus' (Oregon, USA)
189 nextRAD technology. This technology uses a selective PCR primer to amplify consistent
190 genomic loci among individuals. Similarly to RAD-Seq sequences (Rowe, Renaut, &
191 Guggisberg, 2011) in which the DNA flanking a restriction enzyme cut site is selected for
192 amplification, nextRAD amplifies sequences that correspond to the DNA downstream of a short
193 selective priming site. Samples were first fragmented and then ligated to short adapter and
194 barcode sequences using a partial Nextera reaction (Illumina; California, USA) before being
195 amplified using Phusion® Hot Start Flex DNA Polymerase (New England Biolabs;
196 Massachusetts, USA). The 80 dual-barcoded PCR-amplified samples were pooled and the
197 resulting libraries were purified using AMPure XP beads (Agencourt Bioscience Corporation;
198 Massachusetts, USA) at 0.7x. The purified library was then size selected to 350-800 base pairs
199 and sequenced using two runs of an Illumina NextSeq500 sequencer (Genomics Core Facility,
200 University of Oregon).
201
202 The resulting sequences were analytically processed using the SNPsaurus nextRAD pipeline
203 (SNPsaurus, Oregon, USA; Siliceo-Cantero, García, Reynolds, Pacheco, & Lister, 2016).
204 Specifically, reads of 16 randomly selected individuals (of the 80 sequenced) were combined to
205 create a pseudo-reference genome. This was done after removing loci with read counts above
206 20,000, which presumably corresponded to repetitive genomic material, and loci with read counts
207 below 100, which presumably corresponded to off-target or read errors. The filtered reads were
208 aligned to each other using BBMap (Bushnell, 2016). All parameters were set to default values
209 with the exception of minimum alignment identity, which was set to 0.93 to identify alleles, as it

210 is a threshold found to work well for non-reference species (SNPsaurus, Oregon, USA). A single
211 read instance was chosen to represent the locus in the pseudo-reference. This resulted in a total of
212 263,658 loci. All reads from each of the 80 individuals were then aligned to the pseudo-reference
213 using BBMap (Bushnell, 2016) and converted to a vcf genotype table, using Samtools (Li et al.,
214 2009) and bcftools (Li, 2011), after filtering out nucleotides with a quality score of 10 or worse
215 (an empirically informed threshold; SNPsaurus, Oregon, USA). The resulting vcf table was
216 filtered using vcftools (Danecek et al., 2011) for SNPs with a minimum allele frequency of 0.02,
217 a minimum read depth of 5, and a maximum 15% of missing data. We chose this filtering scheme
218 as a balance between accuracy and efficiency and to avoid inadvertent errors associated with our
219 use of a pseudo-reference genome. This resulted in 9774 variable regions. Loci were further
220 filtered using vcftools to exclude loci with less than 5 high quality base-calls and with more than
221 20% missing data or an average of less than 20 high quality base calls. This resulted in a final
222 panel of 8210 Single Nucleotide Polymorphisms (SNPs) that we used in all subsequent analyses.
223

224 **Population structure analyses**

225 We first conducted a series of preliminary analyses to characterize the overall genetic
226 structure of *I. purpurea*. All analyses were run separately for the microsatellite (SSR, hereafter)
227 and SNP datasets given their intrinsic differences and distinct geographic coverage (Fig. 1; Table
228 S3). In addition, we repeated all population structure analyses, separate for the SSR and SNP
229 datasets, for the subset of 6 localities with coincident data for both markers (referred as SSRc and
230 SNPc, hereafter) to assess the robustness of results to the difference in geographic coverage.
231

232 First, we examined population differentiation by estimating F_{ST} using GenAlex v6.5 (Peakall
233 & Smouse, 2012) (because similar global F_{ST} and R_{ST} estimates were obtained for the SSR

234 dataset, we opted to report F_{ST} values only to allow direct comparisons with the SNP dataset). We
235 then estimated contemporary effective population size for each sampled locality in NeEstimator
236 v2 using the excess heterozygous method (Do et al., 2014). We performed this latter analysis to
237 assess the possibility of whether differences in local population size underlie differences in
238 genetic variation (Weckworth et al., 2013) and/or promote asymmetric effective migration rate
239 (*Nm*).
240
241 In addition, we assessed population admixture and spatial genetic clustering using TESS
242 (Chen, Durand, Forbes, & François, 2007). TESS was run using the admixture algorithm and a
243 BYM model (Durand, Jay, Gaggiotti, & François, 2009) with 10 runs per K value, and without
244 using geographic weights. The TESS model, with the lowest DIC was chosen as the optimal
245 model (Durand, Chen, & Francois, 2009). K values tested ranged from two to the maximum
246 number of sampled localities. Additionally, following Wang et al. (2009), we complemented
247 these analyses with Analyses of Molecular Variance (AMOVA; Excoffier, Smouse, & Quattro,
248 1992) run in GenAlex (Peakall & Smouse, 2012) using 9999 permutation replicates. We run
249 these AMOVAs either partitioning the variance into regions based on the spatial genetic clusters
250 previously identified by TESS—to quantify the fraction of the genetic variance explained by
251 these clusters, or leaving it ungrouped (i.e., no regions), for comparison.
252
253 Additionally, we investigated population connectivity by estimating levels of recent
254 migration between sampled localities through the identification of individuals of mixed ancestry
255 using BayesAss (Wilson & Rannala, 2003). BayesAss is a program that uses individual
256 multilocus genotypes and a Markov Chain Monte Carlo (MCMC) algorithm to probabilistically
257 distinguish between immigrants and long-term native individuals (Wilson & Rannala, 2003). We

258 ran BayesAss for 6 million generations using default parameter settings, and discarded the first
259 two million generations as burn-in (Dyer, 2009). For each marker dataset, we repeated this
260 analysis three times (for a total of 18 million generations) and combined the results from the three
261 replicates for our final inference. Then, using a posterior probability cut-off of 0.75 we assign
262 individuals' ancestry. We chose this cut-off value as a minimum credibility score to
263 simultaneously maximize sample size and reliability (stringer thresholds show similar differences
264 between marker sets; results not shown). It is important to note that because of computational
265 limits we had to randomly subsample our set of SNPs to 400 SNPs for this analysis. The same
266 subsampled set was used for the full and reduced (i.e., on the SSRc and SNPc datasets) analyses.
267

268 **Landscape genetics analyses**

269 After assessing overall population structure of *I. purpurea*, we evaluated the association
270 between landscape features and genetic differentiation based on the full datasets. We limited our
271 analyses to the full datasets because of the robust genetic structure recovered between the full and
272 reduced datasets (see below) and the smaller sample size of the latter datasets, which limits
273 statistical inference power. First, we estimated conditional genetic distances (Dyer, Nason, &
274 Garrick, 2010) using GeneticStudio (Dyer, 2009). Briefly, conditional distances are measures of
275 pairwise genetic distance derived from population networks, constructed based on the degree of
276 genetic similarity between sampled localities (Dyer & Nason, 2004). Because these networks are
277 pruned based on the principle of conditional independence of the total among population genetic
278 covariance (using an edge deviance principle; Magwene, 2001), conditional distances reflect
279 genetic similarity between localities that better capture direct gene flow as opposed to
280 connectivity driven by intervening localities (Dyer, 2015b). The complexity of the associated
281 conditional genetic network was summarized by their vertex connectivity (White & Harary,

2001), whereas the congruence between networks derived from different marker sets was measured by their structural congruence (a measure of wether the number of congruent edges between networks is greater than expected by chance) (Dyer, 2009).

Climate, crops, elevation, landcover, population density, roads, and soils landscape layers (Table S1) were converted into landscape resistance layers by assigning a resistance value to each landscape feature in these layers to reflect the difficulty that each feature offers to the movement of gametes or individuals. It is important to note that in contrast to previous studies that typically rely on expert opinion for resistance assignment, we utilized an unbiased statistical optimization to avoid the sensitivity of results to subjective resistance assignment (Spear, Balkenhol, Fortin, McRae, & Scribner, 2010). Specifically, resistance values were optimized through a genetic algorithm approach (Mitchell, 1996). Briefly, in this search algorithm a population of individuals with traits encoded by unique combinations of model parameters (resistance assignment proposals in our case) is allowed to compete with each other based on the fitness associated with the traits it carries (Peterman, Connette, Semlitsch, & Eggert, 2014). Specifically, in Peterman's (2014) implementation of this algorithm, which we followed here, individuals' fitness is estimated by the relative quality of a MLPE.lmm model (Maximum Likelihood Population Effects – Linear Mixture Model) that evaluates the association between pairwise genetic distance and landscape cumulative resistance between localities, estimated in Circuitscape (Shah & McRae, 2008). Individuals with parameter settings (i.e., resistance assignments) that result in better models, as measured by a Deviance Information Criterion (DIC) score, are preferentially represented in the following generation. Offspring modifications introduced by mutations (i.e., small resistance assignment perturbations) allow for exploration of the parameter space. The

algorithm was stopped once 25 generations have passed without significant improvement in fitness.

We implemented Peterman's (2014) algorithm in R (package ResistanceGA; Peterman, 2014) allowing for the independent optimization of each of our landscape layers. The optimal resistance landscapes identified in this way were then used to run a final univariate MLPE.lmm model to characterize the association between landscape features and conditional genetic distances between localities. Because the roads-association resistance was not recovered as significant for either marker dataset, we dropped this layer for all subsequent analyses. Finally, to identify the simultaneous contribution of natural and human-driven landscape features we ran Multiple Regression on Distance Matrices (MRDM; Legendre, Lapointe, & Casgrain, 1994), which has been identified as one of the best performing methods for evaluating the interplay between landscape features and genetic connectivity (Balkenhol, Waits, & Dezzani, 2009). Before running these MRDM analyses, we standardized all optimized resistance layers to mean of zero and variance of one (Dyer et al., 2010). These final regressions included geographic distance as a null model predictor as well as effective population size and were run in R (package ecodist; Goslee & Urban, 2007) using 10,000 permutations to assess significance. In none of our analyses did we implement a Bonferroni correction for multiple testing because of the overly conservative nature of this correction (Glickman, Rao, & Schultz, 2014; Nakagawa, 2004). Instead we applied a false recovery rate correction (Benjamini & Hochberg, 1995) using the function *p.adjust* in R (R Core Development Team, 2016).

RESULTS

Population structure

329 The set of preliminary genetic analyses indicated that *I. purpurea* sampled localities were
330 in no major violation of Hardy-Weinberg equilibrium, as judged by the small difference between
331 expected and observed heterozygosity (mean He = 0.294±0.014 and 0.250±0.001; mean Ho =
332 0.291±0.009 and 0.260±0.001, respectively for SSR and SNP datasets). Levels of expected and
333 observed heterozygosity for the SSR dataset were only slightly greater than those estimated for
334 the SNP dataset. Likewise, the estimated mean effective population size per sampled locality was
335 only slightly greater and more variable for the SSR dataset than for the SNP dataset (13.71±5.59,
336 9.49±0.13, respectively), but in neither case was there salient evidence of a plausible source-sink
337 dynamic, as judged by the similar effective sizes among populations. Neither were there salient
338 differences in F_{ST} estimates between datasets (0.151 and 0.140, respectively for SSR and SNP
339 datasets; Fig. S1), with F_{ST} estimates being within the range of F_{ST} values of other broadly
340 distributed agricultural weeds [F_{ST}: 0.14−0.38] (Bussell, 1999; Eschmann-Grupe, Neuffer, &
341 Hurka, 2004; Müller-Schärer & Fischer, 2001). Congruently, no major differences in genetic
342 estimates between the SSRc and SNPc estimates were found (Table S4). Further confirming the
343 limited spatial structure in this species, spatial genetic clusters identified by the best TESS model
344 (Fig. S2) explained less than 13% of the variance across datasets, and barely reduced the variance
345 explained solely by geographic location when compared to a null model with no regions assigned
346 (Tables 1, S5).
347
348 Despite these similarities between the SSR and SNP datasets, the underlying genetic
349 structure was markedly different. Estimates of recent ancestry differed between SSR and SNP
350 datasets. The analysis on the SSR dataset indicated that migration among localities is more
351 widespread and hardly geographically constrained, with only four localities being primarily
352 constituted of native individuals (Fig. 2a). Across localities, on average 73.65% of individuals

353 were inferred to be 1st or 2nd generation immigrants (it is important to note, however, that such
354 high migration rate surpasses the assumptions of the method, and hence they should be taken
355 cautiously). On the other hand, the analysis of the SNP dataset showed that most populations
356 have a much more limited number of recent immigrants, and that the relatively few inferred
357 immigrants (on average 27.42% of individuals) did not come exclusively from geographically
358 proximate localities (Fig. 2d). Accordingly, SSR and SNP pruned conditional genetic networks
359 (Dyer & Nason, 2004) indicated remarkably different underlying patterns of genetic connectivity
360 (structural congruence = 0.108; Fig. 2b,e). The SSR-based network was more interconnected
361 (vertex connectivity: 5) than the SNP-based network (vertex connectivity: 0). Furthermore,
362 strong admixture was recovered in the SSR dataset, whereas minimal admixture was identified in
363 the SNP dataset (Fig. 2c,f). As before, these differences were consistent when analyzing the
364 SSRc and SNPc datasets. The SNPc dataset was characterized by a smaller percentage of recent
365 immigrants (28.25%) than the SSRc dataset (44.93%) (Fig. S3a,d), and the corresponding genetic
366 networks were also clearly different from each other (structural congruence = 0.002)—with the
367 SSR-based network being more connected (and vertex connectivity = 2) than the SNPc-base
368 network (i.e., vertex connectivity = 0)—(Fig. S3b,e). Finally, as for the full data, a more admixed
369 genetic composition of individuals was recovered in the SSRc dataset than in the SNPc dataset
370 (Fig. S3c,f).
371
372 **Landscape genetics**
373 Unsurprisingly, given the distinct underlying genetic patterns between SSR and SNP
374 datasets (see above), the optimization of landscape resistance layers resulted in different
375 resistance optimization solutions for each dataset (Fig. S4). It is important to note, however, that
376 a formal comparison is unwarranted as the associations recovered are statistical associations

driven by the fit of the resistance parameterization to the data under the statistical model implemented (Martínez-Abraín, 2008). While these associations are expected to recapitulate real biological properties of the study system, they are constrained to the data at hand. Nonetheless, association patterns that are robust to the data used are expected to better reflect the actual impact that landscape features have on gene flow, independently of possible biases introduced by expert opinion. Therefore, we focus below on the common biological findings between marker types, while also denoting the most relevant differences. Such differences likely reflect not only the different environmental ranges covered by each dataset (Fig. 1), but most importantly, the particular population genetic structure underlying each dataset (Fig. 2).

In spite of the distinct underlying genetic structure, there were some landscape features that showed consistent inferred conductance to migration between datasets (Fig. S5). For example, a tendency towards greater landscape conductivity in relatively warm and precipitation-seasonal areas was observed in both datasets along with a steep decline in connectivity towards areas with the greatest temperatures and precipitation-seasonality values in the study area (Fig. S5). Likewise, cotton-dominated areas were recovered as substantially more permeable landscape features than areas dominated by soybean fields, evergreen forests, open shrublands, and grasslands (Fig. S5). Furthermore, both datasets pointed towards human-impacted landscapes playing an important role in shaping genetic connectivity in this species. While in both sets of MLPE.lmm models, null (geographic distance), natural (climate, elevation, and soils), and human-related landscapes (landcover and human population density) were identified as significant (hereafter, $0.01<p<=0.05$ after correction for multiple testing) or marginally significant predictors (hereafter, $p<=0.01$ after correction for multiple testing) of genetic similarity between localities, the variable with the greatest association coefficient and lowest

401 AICc value in these models was in both cases a variable closely linked to human presence
402 (landcover in the SSR dataset, and human population density in the SNP dataset; Table 2) (for
403 comparison, Table S6 shows comparable analyses based on distance-based Redundancy
404 Analysis—another commonly used algorithm in landscape genetics). However, when considering
405 all variables together in a multivariate manner while accounting for geographic distance, human
406 population density, local effective population size, and different aspects of climate were the only
407 variables that remained as significant or marginally significant predictors of genetic
408 differentiation across both SSR and SNP datasets (Table 2). Elevation and soil were identified as
409 significant or marginally significant predictors only in the SNP dataset.
410
411 In summary, across datasets results indicated that human-population-density resistance
412 was robustly associated with population differentiation, with highly populated areas identified as
413 less conducive areas for migration (although the exact association varied between datasets; Fig.
414 S5). Local effective population size was also a significant predictor when considering all other
415 variables. It is important to note, however, that these multivariate regressions explained a variable
416 proportion of the variance (MRDM R^2 for SSR and SNP dataset were 0.109 ($F_{1,29}$ = 3.654, p-val.
417 = 0.063) and 0.532 ($F_{1,6}$ = 1.932, p-val. = 0.113), respectively). In addition to population density
418 and effective population size, temperature was also recovered as a significant predictor of genetic
419 dissimilarity across most analyses for the SSR dataset but not for the SNP dataset (Table 2).
420
421 **DISCUSSION**
422 The results suggest that broadly distributed populations of *I. purpurea* are partially
423 genetically distinct (more so when analyzing the SNP dataset), although there is some indication
424 of long-distance and putatively human-mediated migration between localities—as suggested by

425 the recovered association between human population density and genetic similarity. The levels of
426 differentiation observed and inferred long-distance migration strongly contrast with this species'
427 patchy distribution, which is tightly linked to isolated agricultural patches that are surrounded by
428 a complex matrix of natural and urbanized areas. Contrary to what has been seen in other
429 agricultural weeds (Menchari et al., 2007; Ye, Mu, Cao, & Ge, 2004), populations of *I. purpurea*
430 in relatively close geographic proximity do not form clusters of genetically similar individuals.
431 This finding, supported by our spatial population structure and landscape genetics analyses,
432 suggests that the local agricultural matrix does not seem to have an overarching impact on the
433 connectivity in this species at the landscape level—albeit it likely influences connectivity at the
434 small spatial scales. Instead, climate and human population density were robustly recovered as
435 predictors of genetic connectivity in this species across datasets and analyses. Of these landscape
436 variables, climate has a stronger effect, as judged by its greater MRDM coefficient. Of note,
437 temperature (summarized by climate PC1) was recovered as marginally significant only when
438 considering the SSR dataset, which is the only dataset that covers the northern portion of the
439 range, whereas precipitation seasonality (summarized by climate PC2) was recovered as
440 marginally significant only in the SNP dataset. Otherwise, population density was the only
441 variable across datasets with a marginally significant effect—even after accounting for multiple
442 tests. In addition, local effective population size was found to be a significant predictor only after
443 accounting for all landscape variables, suggesting a plausible superseding effect of genetic drift
444 driving differentiation across localities (Weckworth et al., 2013). Taken together, these results
445 highlight the significant interplay between human-driven and natural landscapes in structuring
446 populations of this species. The role that humans play in this system is likely mediated by their
447 impact on migration patterns themselves as well as by the reduction of population size of this
448 weed through pest control (Barker, Thompson, & Godley, 1984; Baucom & Mauricio, 2008).

449

450 The results also highlight how inferences about population structure and patterns of
451 connectivity may be dataset-dependent, with marked differences becoming apparent only after
452 careful dissection of roughly similar F_{ST} and heterozygosity estimates across molecular markers.
453 Such differences cannot be attributed to the more widespread samples of our SSR dataset, as our
454 findings were robust to subsampling this dataset to match the available SNP samples (see
455 Supporting Information). This represents a rather unexpected finding as both the overall
456 population structure and the influence of landscape features on population connectivity should be
457 inherent species properties and no marker specific realizations of common underlying biological
458 processes (but see below). Next, we detail each of these novel findings and place them in the
459 context of agricultural weed movement across the landscape, invasive species, and landscape
460 genetics practice.

461

462 **Human impact**

463 Given that *I. purpurea* is a naturalized species in the United States that is found primarily
464 associated with cultivated crops and horticultural gardens (Baucom & Mauricio, 2004; Defelice,
465 2001; Fang et al., 2013), the finding that human population density is a predictor of genetic
466 similarity in this species is at first glance unsurprising. Yet, because habitat requirements for
467 establishment and migration are not always the same, especially for organisms with distinct
468 migration stages (e.g., pollen or seeds in plants) and dormant stages (Murphy & Lovett-Doust,
469 2004), this finding is not as straightforward as it seems. In particular, the fact that human
470 population density is recovered as an informative predictor throughout the entire sampled
471 distribution—even after accounting for climate and landcover variation, highlights the direct
472 influence that humans most likely have on structuring the populations of this species and helps to

473 discern the factors involved in the spread of this noxious weed. In this sense, the results point
474 towards humans not only as likely responsible for the introduction of this weed into the United
475 States (Fang et al., 2013), but also as likely responsible for facilitating its current spatial
476 connectivity and genetic structure, and hence its opportunities for thriving in the complex
477 landscapes it inhabits. While further testing is required to formally test this hypothesis, especially
478 considering the limitations of current landscape genetic approaches (see below), our findings
479 suggest a multifaceted effect of human activities.
480
481 For one, it is theoretically possible that human population density primarily facilitates
482 connectivity at local to intermediate scales, which encompasses agricultural fields in relatively
483 close geographic proximity, suggesting that factors such as regional sharing of contaminated
484 agricultural machines, regional trade between farmers, or regional distribution of contaminated
485 crop seeds were at play (Benvenuti, 2007; Boyd & White, 2009; Dastgheib, 1989; Thill &
486 Mallory-Smith, 1997). Yet, the limited spatial clustering at the regional scale and the lack of a
487 significant effect of roads (results not shown) make this possibility unlikely. Instead, considering
488 that i) the horticultural trade has been recognized as the main source of invasive introductions and
489 spread in the United States (Lehan, Murphy, Thorburn, & Bradley, 2013), ii) that *I. purpurea* is
490 an appreciated horticultural species (Fang et al., 2013), and that, given current agricultural
491 practices, crop seed contamination is unlikely to be a major factor (Economic Research Service,
492 1998), it is probable that ornamentals' trade between population centers may help explain both
493 the long distance dispersal events recovered in both datasets and the overall population structure.
494 Alternatively, the impact of human populations on the distribution and abundance of bumblebees
495 (Jha, 2015; Martins, Goncalves, & Melo, 2013), which are *I. purpurea*'s predominant pollinators
496 (Baucom & Mauricio, 2008; Ennos, 1981), could also be partially responsible for the landscape

497 genetic patterns recovered as changes in the pollinators community would have strong effects on
498 gene flow (Jha & Kremen, 2013).
499
500 In reality a combination of all these factors may be involved. While further analyses are
501 needed to elucidate the ultimate causes behind the recovered association between human
502 population density and genetic dissimilarity in *I. purpurea*, our findings bring much needed
503 insight to limit the spread of this noxious weed. Our findings are not only relevant to *I. purpurea*
504 and to the evolution of herbicide resistance in this species (i.e., is herbicide resistance evolving
505 independently across populations or is it being disseminated through human-aided migration?),
506 but also have important implications for other weeds of agricultural concern as well as other
507 human-exploiter species (Blair, 2001), such as other invasives. Specifically, in line with previous
508 work (Auffret, Berg, & Cousins, 2014; Banks, Paini, Bayliss, & Hodda, 2015; Bataille et al.,
509 2011), the results here point towards the need of better strategies to minimize the impact that
510 humans have on the spread of species. In particular, our results further support that humans may
511 not only facilitate the introduction of invasive species into non-colonized areas, but also
512 contribute to the maintenance of gene flow among naturalized populations (Medley, Jenkins, &
513 Hoffman, 2015), which may be critical in providing relevant genetic variants to respond to novel
514 selective regimes as well as prevent inbreeding depression in these newly colonized areas
515 (Edelaar & Bolnick, 2012; Kolbe et al., 2004).
516
517 **SSR- vs. SNP-based inferences**
518 The unique patterns observed for each marker offer the opportunity to explore the underlying
519 causes for such differences and hence a more in-depth understanding of the plausible landscape
520 influences on species' genetic structure. For instance, an important consideration in any

521 landscape genetics study, including this one, is the spatial distribution of samples and spatial
522 scale of environmental data (Wang & Bradburd, 2014), as it can strongly impact the associations
523 recovered. It is thus theoretically possible that the particular geographic sampling of each dataset
524 exclusively drives the differences in genetic structure recovered by the two markers. Yet, the
525 robust differences that we report between the localities and regions common to both datasets
526 (SSRc and SNPc subsampled datasets; see Supporting Information) renders this possibility highly
527 unlikely and suggest that our results are at least moderately robust to sample distribution. Still, it
528 is important to recognize that all our datasets contain sets of spatially clustered samples, partially
529 in reflection of the also spatially cluster distribution of agricultural fields (Ramankutty, Evan,
530 Monfreda, & Foley, 2008). Hence, it is in principle possible that our inferences on all 4 datasets
531 might be strongly impacted by the lack of samples from intervening areas. However, our analyses
532 do not show the pattern of genetic separation between geographic sample clusters that is expected
533 under clustered sampling (Schwartz & McKelvey, 2009), suggesting that the patterns recovered
534 are not simply a sampling artifact.
535
536 Instead, differences between SSR- and SNP-based patterns might be related to the different
537 mutation rates underlying the two type of markers (Wang, 2010, but see Bohonak and Vandergast
538 2011). SSR mutation rates per generation per site (μ) are typically estimated to be between 10^{-3}
539 and 10^{-4} mutations per genome site per generation (Garza & Freimer, 1996), whereas SNP
540 mutation rates are typically estimated to be on average orders of magnitudes slower, around 10^{-8}-
541 10^{-10} (Morin, Luikart, & Wayne, 2004). All else being equal, the inferential power of population
542 structure is tightly linked to the number of mutations (Hubisz, Falush, Stephens, & Pritchard,
543 2009; Turakulov & Easteal, 2003). As a consequence, unless there is widespread homoplasy it is
544 expected to be more likely to recover signatures of population differentiation using the faster

evolving SSR loci. This is true even considering the total number of loci on each dataset (15 SSR loci vs. 8210 SNP loci) (Selkoe & Toonen, 2006). Yet, our results are in contrast with this theoretical expectation as we recovered weaker population structure using the faster evolving SSR loci than using SNP loci. It is still possible that the greater number of expected mutations for SSR loci, which increase the opportunities for homoplasy (Garza & Freimer, 1996), may explain the lower degree of population differentiation in this dataset. Nonetheless, the likelihood that widespread loci homoplasy in SSR allele size across populations has been maintained over the temporal window since the introduction of *I. purpurea* to the US seems small. Hence, this mutation-differential hypothesis is unlikely to be solely responsible for the differences observed. In fact, a large proportion of genetic variation in current populations might, depending on effective population size and generation time, precede the temporal window of many landscape genetics studies. In this regard, the greater number of SNP loci translates into a better genomic representation. Since different genomic regions reflect different coalescent histories (Nielsen & Slatkin, 2013), increased genomic coverage should better capture the range of processes conditioning genetic patterns of populations and thus the combined effect of historical demographic processes and current landscapes. It is then important to consider the relative contribution of both processes: i) input from new mutations and ii) sorting of standing genetic variation, rather than exclusively focus on mutation rates differences. Such sorting is expected to be specific to different genomic regions, which most likely contribute to the differences observed between our SSR and SNP datasets.

Advancing landscape genetics practice

Signals of population structure may arise from a wide range of evolutionary processes, including historical demographic events (He, Edwards, & Knowles, 2013), local adaptation

(Orsini, Vanoverbeke, Swillen, Mergeay, & De Meester, 2013), and reproductive strategies (e.g., selfing in mixed mating species such as *I. purpurea* can lead to a spurious identification of structure; Gao, Williamson, & Bustamante, 2007). Yet, landscape genetics approaches traditionally overlook these plausible confounding processes by working under the assumptions of an equilibrium between migration and genetic drift and an implicit predominance of recent landscape configurations over alternative explanations for the observed population structure (Dyer, 2015b; He et al., 2013; Marko & Hart, 2011). Thus, traditional landscape genetics analyses presumably present an incomplete picture of the evolutionary processes driving current patterns of genetic diversity (Wang & Bradburd, 2014). Nonetheless, these approaches undoubtedly offer a valuable hypothesis-generation framework about the possible role that environmental setting plays in structuring genetic diversity against which the effect of other demographic processes can be evaluated. Arguably, the integration of landscape genetics with historical demographic reconstruction is key for robust population genetics inference since disregarding the plausible effects of either current landscape processes or historical demographic changes would impair the ability to understand species' complex responses to spatio-temporal environmental variation.

In this context, considering the likely complex demographic dynamics of this introduced agricultural weed, our results should be taken as a working hypothesis of the possible role of the interaction between natural and anthropogenic landscapes in structuring *I. purpurea* populations. Nonetheless, our analyses represent a step towards integrating traditional landscape genetics with modern population genetics inference by taking advantage of recent analytical developments and richer molecular datasets. On one hand, our analyses use novel methodological tools that i) surpass the need of arbitrary landscape resistance assignment that make inferences sensitive to

subjectivity of expert opinion (Dyer, 2015a), ii) account for the indirect genetic similarity of populations (Dyer & Nason, 2004), iii) use rigorous statistical inferences (Balkenhol et al., 2009; Peterman et al., 2014), and account for plausible confounding processes (i.e., local effective population size; Weckworth et al., 2013)—although more work on accounting for additional processes such as historical demographic changes is needed. On the other hand, in contrast to the common practice in the field of using a single analysis (commonly Mantel test; Guillot & Rousset, 2013) and one or a few loci (although a few notable exceptions exist; e.g. Perry et al., 2013), which prevents an assessment of common patterns across the genome (Bohonak & Vandergast, 2011), our inferences are derived from common findings among two rather different sets of molecular markers. In doing so, we provide not only statistically robust inferences, but also a better representation of the genome. Hence, our inferences are not only less sensitive to ascertainment bias (Brandström & Ellegren, 2008) and coalescent and mutational variance (Buschiazzo & Gemmell, 2006; Nielsen & Slatkin, 2013; Steiner, Putnam, Hoeck, & Ryder, 2013), but have also the ability to uncover differences in the underlying population dynamics. Such differences have strikingly important implications. For example, when evaluating plausible approaches to the threat of an invasive species such as *I. purpurea*, recommendations would be quite different depending on whether gene flow is believed to be relatively widespread (as inferred by the SSR dataset) or whether it is believed to be minimal (as inferred by the SNP dataset). In this example, it is clear that knowledge-based management would clearly benefit from recognizing the current uncertainty in regards to the exact population connectivity as opposed to automatically relying on a single-marker story.

Given recent advances in next generation sequencing, it seems straightforward to focus on landscape genomics instead of few loci. Hence, development of methods for explicitly integrating

inferences from multiple genome regions and marker types, as it is customary in population genetics, would be of great value. By incorporating multiple loci and coupling traditional landscape genetic tools with coalescent-based simulations to explicitly model landscape effects on genetic population structure, a robust hypothesis framework could be develop to simultaneously account for both current landscape processes and demographic history of species (Balkenhol & Landguth, 2011; Hoban, Bertorelle, & Gaggiotti, 2012). Advances in this area are already being developed with promising perspectives (Alvarado-Serrano & Knowles, 2014; Harris et al., 2016; He et al., 2013).

Final remarks

By offering a working hypothesis of the effect of current landscapes on genetic differentiation, traditional landscape genetics results serve the purpose of identifying relevant models for further testing (Baguette, Blanchet, Legrand, Stevens, & Turlure, 2013; Dyer, 2015a). Under this framework, our results pave the way for rigorous simulation-based assessments of the role of landscape features in promoting or deterring population differentiation in a noxious agricultural weed, and hence for successful knowledge-based invasive management. Specifically, we identify a probably major role of human-driven gene flow and long distance dispersal events in the demographic history of this species. If this weed was, as hypothesized, singly introduced through horticulture from a European bottlenecked population during the European colonization of North America (Fang et al., 2013), distinct ancestral structure (pre-dating US introduction) would be unlikely. Under this scenario, the rather distinct clustering of individual subpopulations recovered would likely reflect the connectivity driven by agricultural and horticultural activities and the complex natural/human landscape *I. purpurea* has experienced post-introduction to the US. Alternatively, human trade might have allowed for recurrent introduction events, which

effects would have probably been amplified by local agricultural and horticultural activities. Regardless, these findings call for future model-based inference that explicitly considers the impact of human population density in conjunction with climate to further investigate the evolutionary drivers of population structure in this noxious weed.

ACKNOWLEDGEMENTS

We thank Adam Kuester for seed collecting and for contributing valuable data for this study. We also thank Ariana Wilson, Eva Fall, and Dan York for tissue collection. This research was funded by USDA NIFA grants 04180 and 07191to R.S.B.

DATA ARCHIVING

All data generated is in the process of being archived in Dryad.

AUTHOR CONTRIBUTIONS

D.F.A.-S. and R.S.B conceived the study. D.F.A.-S. and M.V.E. generated and compiled the molecular and GIS data. D.F.A.-S. analyzed the data. D.F.A.-S., M.V.E., S.M.C., and R.S.B. wrote the manuscript. All authors read and approved the final submission.

LITERATURE CITED

Alvarado-Serrano, D. F., & Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources*, *14*, 233–248.

Auffret, A. G., Berg, J., & Cousins, S. A. O. (2014). The geography of human-mediated dispersal. *Diversity and Distributions*, *20*, 1450–1456.

Auffret, A. G., & Cousins, S. A. O. (2013). Humans as long-distance dispersers of rural plant communities. *PLoS ONE*, *8*, e62763.

Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., & Turlure, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, *88*, 310–326.

Balkenhol, N., & Landguth, E. L. (2011). Simulation modelling in landscape genetics: on the need to go further. *Molecular Ecology*, *20*, 667–670.

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

Banks, N. C., Paini, D. R., Bayliss, K. L., & Hodda, M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology Letters*, *18*, 188–199.

Barker, M. A., Thompson, L., & Godley, F. M. (1984). Control of annual morningglories (*Ipomoea* spp.) in soybeans (*Glycine max*). *Weed Science*, *32*, 813–818.

Barrowclough, G. F. (1980). Gene flow, effective population sizes and genetic variance components in birds. *Evolution*, *34*, 789–798.

Bataille, A., Cunningham, A. A., Cruz, M., Cedeño, V., & Goodman, S. J. (2011). Adaptation, isolation by distance and human-mediated transport determine patterns of gene flow among populations of the disease vector *Aedes taeniorhynchus* in the Galapagos Islands. *Infection,*

Genetics and Evolution, *11*, 1996–2003.

Baucom, R. S., & Mauricio, R. (2004). Fitness costs and benefits of novel herbicide tolerance in a noxious weed. *Proceedings of the National Academy of Sciences*, *101*, 13386–13390.

Baucom, R. S., & Mauricio, R. (2008). The evolution of novel herbicide tolerance in a noxious weed: the geographic mosaic of selection. *Evolutionary Ecology*, *22*, 85–101.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*, 289–300.

Benvenuti, S. (2007). Weed seed movement and dispersal strategies in the agricultural environment. *Weed Biology and Management*, *7*, 141–157.

Blair, R. B. (2001). Birds and butterflies along urban gradients in two ecoregions of the US. In J. L. Lockwood & M. L. McKinney (Eds.), *Biotic homogenization* (pp. 33–56). New York: Kluwer Academic.

Bohonak, A. J., & Vandergast, A. G. (2011). The value of DNA sequence data for studying landscape genetics. *Molecular Ecology*, *20*, 2477–2479.

Boyd, N. S., & White, S. (2009). Impact of wild blueberry harvesters on weed seed dispersal within and between fields. *Weed Science*, *57*, 541–546.

Brandström, M., & Ellegren, H. (2008). Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Research*, *18*, 881–887.

Buschiazzo, E., & Gemmell, N. J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*, *28*, 1040–1050.

Bushnell, B. (2016). BBMap short read aligner. Retrieved from http://sourceforge.net/projects/bbmap/

Bussell, J. D. (1999). The distribution of random amplified polymorphic DNA (RAPD) diversity

amongst populations of *Isotoma petraea* (Lobeliaceae). *Molecular Ecology*, *8*, 775–789.

Chaney, L., & Baucom, R. S. (2012). The evolutionary potential of Baker's weediness traits in the common morning glory, *Ipomoea purpurea* (Convolvulaceae). *American Journal of Botany*, *99*, 1524–1530.

Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, *7*, 747–756.

Cimalová, Š.˘, & Lososová, Z. (2009). Arable weed vegetation of the northeastern part of the Czech Republic: effects of environmental factors on species composition. *Plant Ecology*, *203*, 45–57.

Clobert, J., Baguette, M., Benton, T. G., Bullock, J. M., & Ducatez, S. (2012). *Dispersal Ecology and Evolution* (1st ed.). Oxford: Oxford University Press.

Culpepper, A. S. (2006). Glyphosate-induced weed shifts. *Weed Technology*, *20*, 277–281.

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*, 486–499.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Dastgheib, F. (1989). Relative importance of crop seed, manure and irrigation water as sources of weed infestation. *Weed Research*, *29*, 113–116.

Defelice, M. S. (2001). Tall morningglory, *Ipomoea purpurea* (L .) roth-flower or foe? *Weed Technology*, *15*, 601–606.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective

population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*, 209–214.

Durand, E., Chen, C., & Francois, O. (2009). TESS version 2.3 - reference manual August 2009. *Available at: memberstimc.imag.fr/Olivier.Francois/tess.html*, 1–30.

Durand, E., Jay, F., Gaggiotti, O. E., & François, O. (2009). Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, *26*, 1963–1973.

Dyer, R. J. (2009). GeneticStudio: a suite of programs for spatial analysis of genetic-marker data. *Molecular Ecology Resources*, *9*, 110–3.

Dyer, R. J. (2015a). Is there such a thing as landscape genetics? *Molecular Ecology*, *10*, 139–143.

Dyer, R. J. (2015b). Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 327–342.

Dyer, R. J., & Nason, J. D. (2004). Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, *13*, 1713–1727.

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

Eberhardt, M., & Teal, F. (2013). No mangoes in the tundra: spatial heterogeneity in agricultural productivity analysis. *Oxford Bulletin of Economics and Statistics*, *75*, 914–939.

Economic Research Service, U. S. D. A. (1998). Seed industry structure is characterized by growth and consolidation. *AIB*, *786*, 25–29.

Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology and Evolution*, *27*, 659–665.

Ennos, R. A. (1981). Quantitative studies of the mating system in two sympatric species of *Ipomoea* (Convolvulaceae). *Genetica*, *57*, 93–98.

Epperson, B. K., & Clegg, M. T. (1986). Spatial-autocorrelation analysis of flower color

polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). *The American Naturalist*, *128*, 840–858.

Eschmann-Grupe, G., Neuffer, B., & Hurka, H. (2004). Extent and structure of genetic variation in two colonising *Diplotaxis* species (Brassicaceae) with contrasting breeding systems. *Plant Systematics and Evolution*, *244*, 31–43.

Everman, E., & Klawinski, P. (2013). Human-facilitated jump dispersal of a non-native frog species on Hawai'i Island. *Journal of Biogeography*, *40*, 1961–1970.

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

Fang, Z., Gonzales, A. M., Durbin, M. L., Meyer, K. K. T., Miller, B. H., Volz, K. M., … Morrell, P. L. (2013). Tracing the geographic origins of weedy *Ipomoea purpurea* in the Southeastern United States. *Journal of Heredity*, *104*, 666–677.

Fountain, T., Duvaux, L., Horsburgh, G., Reinhardt, K., & Butlin, R. K. (2014). Human-facilitated metapopulation dynamics in an emerging pest species, *Cimex lectularius*. *Molecular Ecology*, *23*, 1071–1084.

Gao, H., Williamson, S., & Bustamante, C. D. (2007). A markov chain monte carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, *176*, 1635–1651.

Garza, J. C., & Freimer, N. B. (1996). Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research*, *6*, 211–217.

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*, *67*, 850–857.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19.

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, *4*, 336–344.

Harris, S. E., Xue, A. T., Alvarado-Serrano, D. F., Boehm, T., Joseph, T., Hickerson, M. J., & Munshi-South, J. (2016). Urbanization shapes the demographic history of a native rodent (the white- footed mouse, *Peromyscus leucopus*) in New York City. *Biology Letters*, *12*, 1–6.

He, Q., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, *67*, 3386–3402.

Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, *13*, 110–122.

Hodkinson, D. J., Thompson, K., Journal, T., & Dec, N. (2007). Plant dispersal: the role of man. *Journal of Applied Ecology*, *34*, 1484–1496.

Holderegger, R., Buehler, D., Gugerli, F., & Manel, S. (2010). Landscape genetics of plants. *Trends in Plant Science*, *15*, 675–683.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*, 1322–1332.

Jha, S. (2015). Contemporary human-altered landscapes and oceanic barriers reduce bumble bee gene flow. *Molecular Ecology*, *24*, 993–1006.

Jha, S., & Kremen, C. (2013). Urban land use limits regional bumble bee gene flow. *Molecular Ecology*, *22*, 2483–2495.

Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2004).

Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, *431*, 177–81.

Kuester, A., Chang, S.-M., & Baucom, R. S. (2015). The geographic mosaic of herbicide resistance evolution in the common morning glory, *Ipomoea purpurea*: evidence for resistance hotspots and low genetic differentiation across the landscape. *Evolutionary Applications*, *8*, 821–833.

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

Lehan, N. E., Murphy, J. R., Thorburn, L. P., & Bradley, B. A. (2013). Accidental introductions are an important source of invasive plants in the continental United States. *American Journal of Botany*, *100*, 1287–1293.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*, 183–189.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Mack, R. N., & Lonsdale, W. M. (2001). Humans as global plant dispersers: getting more than we bargained for. *BioScience*, *51*, 95–102.

Magwene, P. M. (2001). New tools for studying integration and modularity. *Evolution*, *55*, 1734–1745.

Margosian, M. L., Garrett, K. A., Hutchinson, J. M. S., & With, K. A. (2009). Connectivity of the American agricultural landscape: assessing the national risk of crop pest and disease spread.

BioScience, *59*, 141–151.

Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology and Evolution*, *26*, 448–456.

Martínez-Abraín, A. (2008). Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology. *Acta Oecologica*, *34*, 9–11.

Martins, A. C., Goncalves, R. B., & Melo, G. A. R. (2013). Changes in wild bee fauna of a grassland in Brazil reveal negative effects associated with growing urbanization during the last 40 years. *Zoologia*, *30*, 157–176.

McRae, B. (2006). Isolation by resistance. *Evolution*, *60*, 1551–1561.

Medley, K. A., Jenkins, D. G., & Hoffman, E. A. (2015). Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito. *Molecular Ecology*, *24*, 284–295.

Meehan, T. D., Werling, B. P., Landis, D. A., & Gratton, C. (2011). Agricultural landscape simplification and insecticide use in the Midwestern United States. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 11500–11505.

Menchari, Y., Délye, C., & Le Corre, V. (2007). Genetic variation and population structure in black-grass (*Alopecurus myosuroides* Huds.), a successful, herbicide-resistant, annual grass weed of winter cereal fields. *Molecular Ecology*, *16*, 3161–3172.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge Mass.: MIT Press.

Molecular Ecology Resources Primer Development Consortium, -, Aksoy, S., Almeida-Val, V. M. F., Azevedo, V. C. R., Baucom, R., Bazaga, P., … Cian, M. J. (2013). Permanent genetic resources added to molecular ecology resources database 1 October 2012–30 November 2012. *Molecular Ecology Resources*, *13*, 341–343.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, *19*, 208–216.

Müller-Schärer, H., & Fischer, M. (2001). Genetic structure of the annual weed *Senecio vulgaris* in relation to habitat type and population size. *Molecular Ecology*, *10*, 17–28.

Murphy, H. T., & Lovett-Doust, J. (2004). Context and connectivity in plant metapopulations and landscape mosaics: does the matrix matter? *Oikos*, *105*, 3–14.

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology*, *15*, 1044–1045.

Navas, M. L. (2012). Trait-based approaches to unravelling the assembly of weed communities and their impact on agro-ecosystem functioning. *Weed Research*, *52*, 479–488.

Nielsen, R., & Slatkin, M. (2013). *An introduction to population genetics: theory and applications*. Sunderland Mass.: Sinauer Associates.

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, *62*, 316–36.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Helene, W. (2015). vegan: community ecology package. Available at: http://cran.r-project.org/package=vegan.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, *22*, 5983–5999.

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

Perry, G. H., Louis, E. E., Ratan, A., Bedoya-Reina, O. C., Burhans, R. C., Lei, R., … Miller, W. (2013). Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar. *Proceedings of the National Academy of Sciences of the United States*

of America, *110*, 5823–5828.

Peterman, W. E. (2014). ResistanceGA: an R package for the optimization of resistance surfaces using genetic algortihms. *bioRxiv*, *7575*.

Peterman, W. E., Connette, G. M., Semlitsch, R. D., & Eggert, L. S. (2014). Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, *23*, 2402–2413.

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

R Core Development Team. (2016). R: a language and environment for statistical computing, 3.2.4.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, *22*, 1–19.

Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global change? *Conservation Biology*, *21*, 329–336.

Rogers, J. B., Murray, D. S., Verhalen, L. M., & Claypool, P. L. (1996). Ivyleaf morningglory (*Ipomoea hederacea*) interference with cotton (*Gossypium hirsutum*). *Weed Technology*, *10*, 107–114.

Rowe, H. C., Renaut, S., & Guggisberg, A. (2011). RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, *20*, 3499–34502.

Schwartz, M. K., & McKelvey, K. S. (2009). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, *10*, 441–452.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using

and evaluating microsatellite markers. *Ecology Letters*, *9*, 615–629.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, *68*, 1–15.

Shah, V. B., & McRae, B. H. (2008). Circuitscape: a tool for landscape ecology. *Proceedings of the 7th Python in Science Conference*, 62–65.

Siliceo-Cantero, H. H., García, A., Reynolds, R. G., Pacheco, G., & Lister, B. C. (2016). Dimorphism and divergence in island and mainland Anoles. *Biological Journal of the Linnean Society*, *118*, 852–872.

Simberloff, D. (2013). *Invasive species: what everyone needs to know*. Oxford: Oxford University Press.

Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, *16*, 393–430.

Spear, S. F., Balkenhol, N., Fortin, M. J., McRae, B. H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, *19*, 3576–3591.

Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., & Ryder, O. A. (2013). Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences*, *1*, 261–281.

Taylor, P. D., Fahrig, L., Henein, K., & Merriam, G. (1993). Connectivity is a vital element of landscape structure. *Oikos*, *68*, 571–573.

Thill, D. C., & Mallory-Smith, C. A. (1997). The nature and consequence of weed spread in cropping systems. *Weed Science*, *45*, 337–342.

Tiffin, P., & Rausher, M. (1999). Genetic constraints and selection acting on tolerance to herbivory in the common morning glory *Ipomoea purpurea*. *The American Naturalist*, *154*, 700–716.

Trakhtenbrot, A., Nathan, R., Perry, G., & Richardson, D. M. (2005). The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions*, *11*, 173–181.

Turakulov, R., & Easteal, S. (2003). Number of SNPs loci needed to detect population structure. *Human Heredity*, *55*, 37–45.

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

Wang, I. J. (2010). Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, *19*, 2605–2608.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*, 5649–5662.

Wang, I. J., Savage, W. K., & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Molecular Ecology*, *18*, 1365–1374.

Webster, T. M., & Nichols, R. L. (2012). Changes in the prevalence of weed species in the major agronomic crops of the southern United States: 1994/1995 to 2008/2009. *Weed Science*, *60*, 145–157.

Weckworth, B. V, Musiani, M., Decesare, N. J., Mcdevitt, A. D., Hebblewhite, M., & Mariani, S. (2013). Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society B*, *280*, 1–9.

White, D. R., & Harary, F. (2001). The cohesiveness of blocks in social networks: node connectivity and conditional density. *Sociological Methodology*, *31*, 305–359.

Wichmann, M. C., Alexander, M. J., Soons, M. B., Galsworthy, S., Dunne, L., Gould, R., … Bullock, J. M. (2009). Human-mediated dispersal of seeds over long distances. *Proceedings*

of the Royal Society B, *276*, 523–532.

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, *163*, 1177–1191.

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.

Ye, W. H., Mu, H. P., Cao, H. L., & Ge, X. J. (2004). Genetic structure of the invasive *Chromolaena odorata* in China. *Weed Research*, *44*, 129–135.

954 **TABLES**

955 **Table 1.** Analysis of Molecular Variance (AMOVA) of SSR and SNP data. The contribution of
956 spatial clusters (regions), localities, and individuals is shown. For comparison, results from
957 an AMOVA analysis with no region category defined are presented in parentheses
958 underneath.

959

Effect	F-statistic	Variance explained		F-value		P-value	
		SSR	SNP	SSR	SNP	SSR	SNP
Regions	F_{RT}	3.94%	8.51%	0.006	0.085	0.001	0.001
Localities	F_{SR}	9.05% (11.06%)	6.10% (13.02%)	0.106	0.067	0.001	0.001
Individuals (among)	F_{ST}	0.67% (38.33%)	24.85% (25.31%)	0.112 (0.111)	0.146 (0.130)	0.001 (0.001)	0.001 (0.001)
Individuals (within)	F_{IS}	86.33% (50.61%)	60.54% (61.67%)	0.431 (0.431)	0.291 (0.291)	0.001 (0.001)	0.001 (0.001)
Total	F_{IT}	100% (100%)	100% (100%)	0.494 (0.494)	0.395 (0.383)	0.001 (0.001)	0.001 (0.001)

960

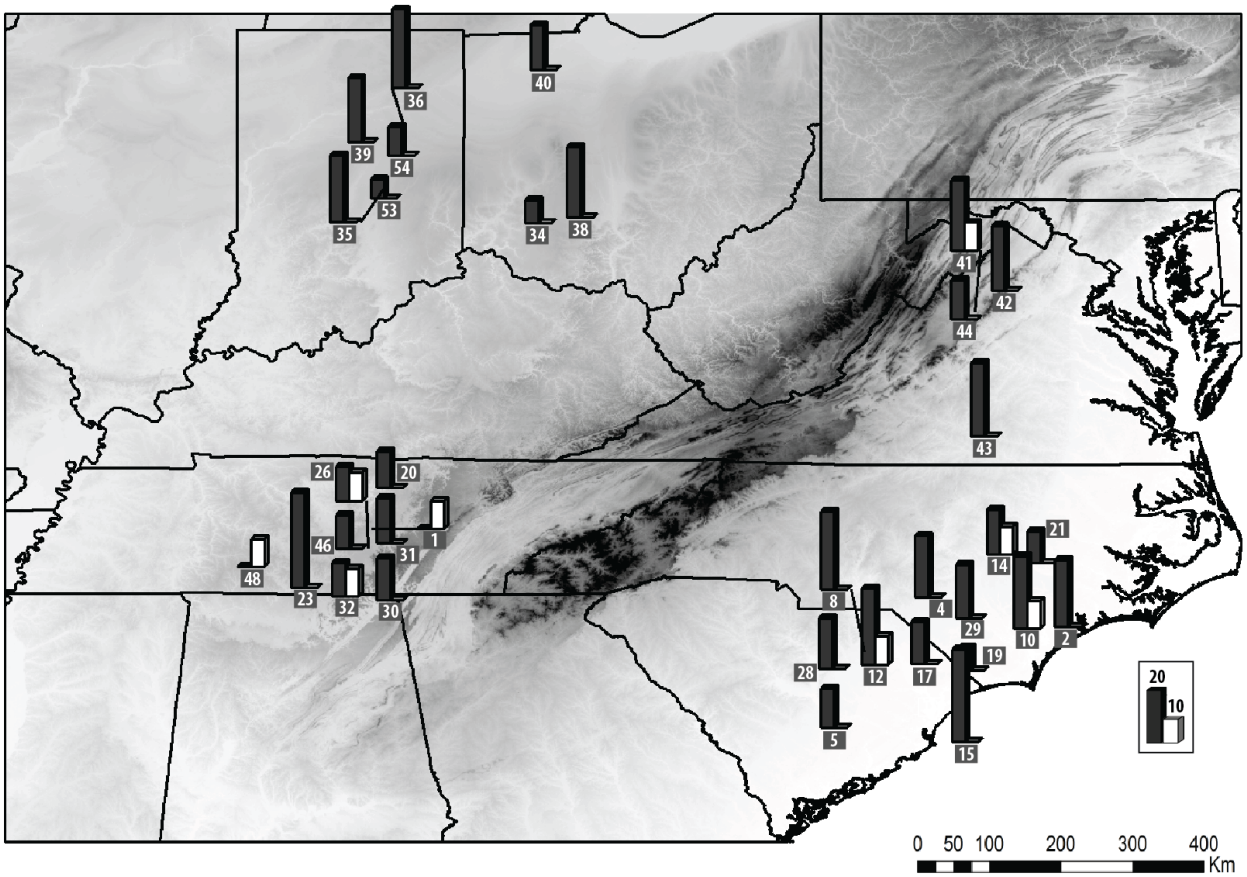
Table 2. Summary of landscape genetics models. Model coefficients are reported followed by associated p-value (in parenthesis) and, for MLPE.lmm models, followed by AICc difference and ranking (in square brackets). Significant coefficients are in bold, marginally significant coefficients are marked with an asterisk.

Feature	MLPE.lmm		MRDM	
	SSR	SNP	SSR	SNP
Intrinsic variables				
Geographic distance	0.187* (0.052) +12.303 [8]	0.780* (0.055) +0.396 [4]	0.051 (0.121)	0.255* (0.056)
Population size (Ne)	—	—	**-0.550** (0.001)	-2.650* (0.051)
Natural environment variables				
Climate PC1	**0.243** (0.034) +10.356 [3]	**0.967** (0.045) +0.032 [2]	0.533* (0.064)	1.782 (0.517)
Climate PC2	**0.205** (0.048) +12.030 [7]	0.814* (0.055) +1.560 [7]	-0.029 (0.978)	21.443* (0.075)
Elevation	**0.244** (0.034) +10.742 [4]	0.840* (0.054) +1.423 [6]	-0.607 (0.480)	-31.789* (0.095)
Soil PC1	**0.208** (0.039) +11.378 [6]	**1.044** (0.045) +0.471 [5]	0.305 (0.510)	**15.789** (0.038)
Soil PC2	**0.320** (0.018) +8.139 [2]	**0.941** (0.045) + 0.284 [3]	0.187 (0.703)	-6.838 (0.476)
Human-impact variables				
Crops	-0.226 (0.134) +14.491 [9]	0.858* (0.054) +15.371 [8]	0.154 (0.526)	0.360 (0.840)
Landcover	**0.582**	**1.358**	0.340	3.887

	(<0.001) – [1]	(0.003) +39.775 [9]	(0.218)	(0.184)
Population density	**0.227** (0.034) +10.821 [5]	**0.912** (0.045) – [1]	-0.519* (0.095)	**-3.271** (0.037)

966

967 **FIGURES**



968

969 **Figure 1**. Distribution of *Ipomoea purpurea*'s sampled localities. Sample sizes for both SSR

970 (black bars) and SNP (white bars) datasets are indicated (locality numbers are given in

971 squares). Elevation is provided as background.

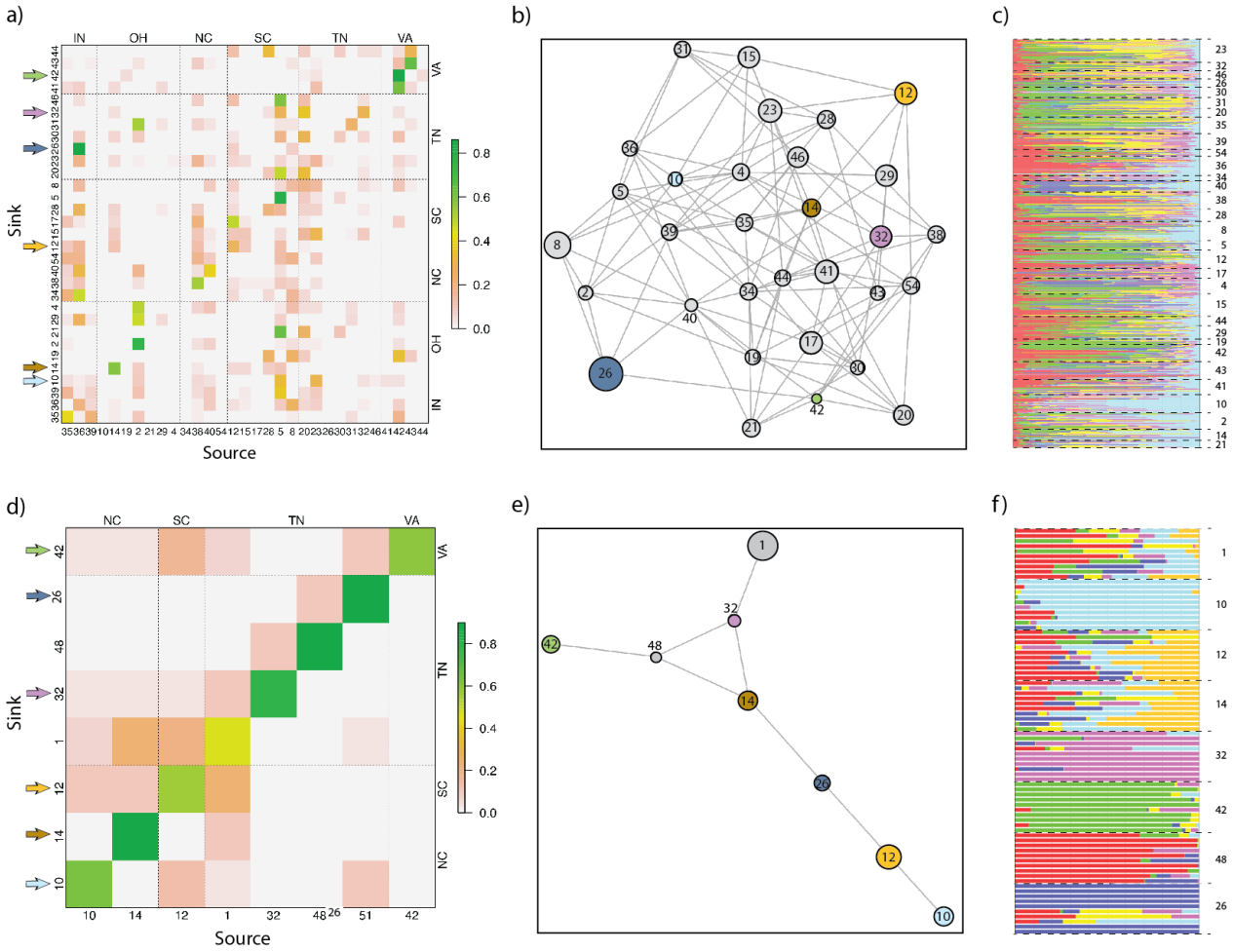


Figure 2. Inferred population connectivity. The estimated origin of individuals for each sampled locality (i.e., sink) is depicted according to the locality they were inferred to have originated from (source) (a, d). The color of each cell in these plots depicts the proportion of individuals in the sink population that were estimated to be recent immigrants from each locality along the x-axis. Cells on the minor diagonal correspond to the proportion of native individuals. Pruned conditional genetic networks (b, e) and posterior estimates of admixture proportion identified by TESS analysis (c, f) are also displayed. The top row shows SSR-based results, the bottom shows the SNP-based results. Locality numbers follow Fig. 1. Localities shared between SSR and SNP datasets are denoted by unique colors in all panels (for a corresponding figure based exclusively on these shared localities, see Fig. S3).