The relative contribution of natural landscapes and human-mediated factors on the connectivity of a noxious invasive weed

Diego F. Alvarado-Serrano, Megan Van Etten, Shu-Mei Chang, Regina S. Baucom

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 as well as to gain a better understanding of anthropogenic impacts on the interplay between gene flow, genetic drift and selection. Here we use genome-wide molecular markers to characterize the population genetic structure and connectivity of *Ipomoea purpurea*, a noxious invasive weed. We likewise assess the interaction between natural and human-driven influences on genetic differentiation among populations. Our analyses find that human population density is 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 of genetic connectivity in this species. We discuss the

implications of these results and highlight future research needed to disentangle the mechanistic processes underlying population connectivity of weeds.

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

INTRODUCTION

Elucidating routes and levels of migration between populations of a species is essential to understand 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 (McRae, 2006; Cushman *et al.*, 2006) as well as influence the likelihood of successful establishment of immigrants (Wang and Bradburd, 2014; Sexton *et al.*, 2014). Landscape features can also indirectly condition the effect of gene flow by influencing local effective population sizes (Wright, 1949; Slatkin, 1985). Consequently, the landscape, loosely defined as an area with spatially variable biotic and abiotic factors (Holderegger *et al.*, 2010), influences the levels of effective gene flow among populations (Clobert *et al.*, 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 and Klawinski, 2013; Fountain *et al.*, 2014). Such species may be capable of maintaining population connectivity over vast geographic ranges (Trakhtenbrot *et al.*, 2005) by overcoming landscape features that would

47 otherwise represent natural barriers. Such species would thus be able to attain
48 dispersal distances that could be orders of magnitude greater than those
49 dependent primarily on natural dispersal agents (Mack and Lonsdale, 2001;
50 Ricciardi, 2007). By facilitating dispersal, humans have the potential to
51 condition the balance between drift and selection (Slatkin, 1985; Lenormand,
52 2002), introduce genetic variation to local populations (Kolbe *et al.*, 2004),
53 prevent local extinction or favor recolonization (Fountain *et al.*, 2014), and alter
54 the overall genetic constitution of populations (Bataille *et al.*, 2011). Human-
55 aided migration—intentional or unintentional—is particularly prevalent in
56 plants (Hodkinson *et al.*, 2007; Auffret and Cousins, 2013), where it has had
57 major impacts on the distribution of species and stability of communities
58 (Simberloff, 2013 and references therein). Despite our knowledge of both
59 human and natural factors influencing dispersal, there remains a gap in our
60 understanding of the relative influence of each on the distribution of genetic
61 variation among populations of many (if not most) plant species.
62
63 A particularly amenable study system to fill this knowledge gap comes
64 from agricultural weed populations. Agricultural weeds experience a highly
65 dynamic landscape characterized by frequent spatial rearrangements and
66 changes in the physical environment (e.g., expansion of agricultural front,
67 increased fragmentation, crop rotation, agricultural chemical use) (Menchari *et*

al., 2007; Meehan *et al.*, 2011). At the same time, natural features such as climate, soil type, and topography likely also play a significant role in structuring populations (Cimalová and Lososová, 2009; Navas, 2012). Under these conditions, human-aided migration may be critical for weedy plant success (Epperson and Clegg, 1986). However, we have limited knowledge of how or if weedy plant populations are able to maintain connectivity through a complex landscape matrix. Addressing this limitation would improve our understanding of the underlying processes governing connectivity of weed populations and also offer practical tools to deal with the economic problems that weeds impose (on the order of 33B USD per year in US agriculture alone; Pimentel *et al.*, 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 and evaluate how multiple landscape features influence genetic connectivity of a noxious agricultural weed, *Ipomoea purpurea*. Specifically, we ask the following questions: 1) what is the overall population structure of *I. purpurea*, one of the most troublesome weeds in US agriculture (Webster and Nichols, 2012), and 2) which natural and/or human-influenced landscape features—soils, elevation, climate, landcover, crop types, human population density—may act to promote or

89 constrain genetic connectivity between populations of this weed? Answering
90 these questions offers a deeper understanding of the multiplicity of population
91 structure drivers that influence noxious weeds.
92
93 **MATERIALS AND METHODS**
94 **Study system**
95 *Ipomoea purpurea*, the common morning glory, is a noxious agricultural
96 weed (Defelice, 2001; Fang *et al.*, 2013) that has a widespread distribution
97 across highly heterogeneous landscapes in the Eastern, South- and Mid-western
98 regions of the United States (Culpepper, 2006; Webster and Nichols, 2012). It is
99 a self-compatible annual bumblebee-pollinated vine and is found primarily in
100 agricultural fields and disturbed areas (Tiffin and Rausher, 1999; Baucom and
101 Mauricio, 2008), as well as cultivated flower gardens and yards (Defelice,
102 2001). *I. purpurea* is one of the most problematic agricultural weeds of
103 southeastern agriculture (Webster and Nichols, 2012), and exhibits variable
104 levels of resistance to the commonly used herbicide glyphosate (Kuester *et al.*,
105 2015). This species is also a major concern for conservation given its
106 naturalization in multiple regions throughout the world and its aggressiveness
107 as an invasive (Chaney and Baucom, 2012; Fang *et al.*, 2013).
108
109 **Data compilation**

To capture the plausible effect of both natural and disturbed landscapes on structuring genetic diversity in *I. purpurea*, we compiled a diverse set of GIS data for the continental US from a variety of sources (Table S1). These data include human activities (human population density, landcover, planted crops, and roads) as well as natural factors such as elevation, climate (19 variables summarizing central tendencies and variability patterns in temperature and precipitation, and soil characteristics), and soil characteristics (8 variables summarizing the texture, pH, and organic and inorganic content of the top 20cm of soil). Focusing on both sets of data allowed us to assess the relative influence of natural and human effects on structuring *I. purpurea*'s populations. We first processed all these data into landscape layers at a common spatial resolution of $10km^2$ and a common spatial extent around the US states with available samples (Fig. 1). This spatial resolution was chosen to maintain a practical balance between scale and analytical manageability given available computational resources. To reduce dimensionality, we opted to perform two separate Principal Component Analyses (PCAs) on the 19 climatic and 8 soil layers, respectively. For all subsequent analyses we kept the resulting first two principal components of each of these analyses, which accounted for over 78% of the variance in each case, and primarily summarized temperature temporal gradients and precipitation seasonality, and soils' pH, sandiness, and grain size, respectively (Table S2).

131

132 We compiled data from a panel of 15 previously optimized microsatellite
133 loci (Molecular Ecology Resources Primer Development Consortium 2013) to
134 examine the genetic connectivity of populations of *I. purpurea*. These data
135 (Kuester *et al.*, 2015) encompass a total of 597 individuals from 31 localities
136 (with a minimum of 8 individuals per locality) (Fig. 1; Table S3), collected in
137 2012 from farms across the range of *I. purpurea* in the United Sates (Kuester *et*
138 *al.*, 2015). In addition, to obtain a more comprehensive representation of the
139 genome of *I. purpurea*, we generated a Next Generation Sequencing (NGS)
140 dataset from an additional set of individuals (10 individuals each from 6
141 localities represented in the SSR dataset, plus 2 additional localities in close
142 geographic proximity to localities in the SSR dataset for a total of 8
143 populations; Fig. 1).

144

145 To generate the NGS dataset, we constructed genome-wide Genotype By
146 Sequencing (GBS) library. The GBS library was developed using 7ng of
147 genomic DNA, extracted from leaf or cotyledon tissue, using SNPsaurus'
148 (Oregon, USA) nextRAD technology. Samples were first fragmented and then
149 ligated to short adapter and barcode sequences using a partial Nextera reaction
150 (Illumina; California, USA) before being amplified using Phusion® Hot Start
151 Flex DNA Polymerase (New England Biolabs; Massachusetts, USA). The 80

152 dual-barcoded PCR-amplified samples were pooled and the resulting libraries
153 were purified using AMPure XP beads (Agencourt Bioscience Corporation;
154 Massachusetts, USA) at 0.7x. The purified library was then size selected to 350-
155 800 base pairs and sequenced using two runs of an Illumina NextSeq500
156 sequencer (Genomics Core Facility, University of Oregon).
157
158 The resulting sequences were analytically processed using the SNPsaurus
159 nextRAD pipeline (SNPsaurus, Oregon, USA; Siliceo-Cantero *et al.*, 2016).
160 Specifically, reads of 16 randomly selected individuals (of the 80 sequenced)
161 were combined to create a pseudo-reference genome. This was done after
162 removing loci with read counts above 20,000, which presumably corresponded
163 to repetitive genomic material, and loci with read counts below 100, which
164 presumably corresponded to off-target or read errors. The filtered reads were
165 aligned to each other using BBMap (Bushnell, 2014). All parameters were set
166 to default values with the exception of minimum alignment identity, which was
167 set to 0.93 to identify alleles, as this threshold has been found to work well for
168 non-reference species (SNPsaurus, Oregon, USA). A single read instance was
169 chosen to represent the locus in the pseudo-reference. This resulted in a total of
170 263,658 loci. All reads from each of the 80 individuals were then aligned to the
171 pseudo-reference using BBMap (Bushnell, 2014) and converted to a vcf
172 genotype table, using Samtools (Li *et al.*, 2009) and bcftools (Li, 2011), after

173 filtering out nucleotides with a quality score of 10 or worse. The resulting vcf
174 table was filtered using vcftools (Danecek *et al.*, 2011) for SNPs with a
175 minimum allele frequency of 0.02, a minimum read depth of 5, and a maximum
176 15% of missing data. This resulted in 9774 variable regions. Loci with less than
177 5 high quality base-calls and with more than 20% missing data or an average of
178 less than 20 high quality base calls were also removed using vcftools (Danecek
179 *et al.*, 2011). This resulted in a final panel of quality-vetted 8210 Single
180 Nucleotide Polymorphisms (SNPs) (Fig. S1) that we used in all subsequent
181 analyses.
182
183 **Population structure analyses**
184 We first conducted a series of analyses to characterize the overall genetic
185 structure of *I. purpurea* populations. All analyses were run separately for the
186 microsatellite (SSR, hereafter) and SNP datasets given their intrinsic differences
187 and distinct geographic coverage (Fig. 1; Table S3). In addition, we repeated all
188 population structure analyses using just the subset of 6 localities where SSR and
189 SNP datasets are both available. Running separate analyses using these two
190 different markers (referred as SSRc and SNPc, hereafter) allowed us to
191 determine if the differences between marker types was due to differences in
192 sample size (SSR = 24 localities; SNP = 8 localities) or geographic coverage
193 (Fig. 1). Similarly, to determine if the differences uncovered between marker

194 types were due to SNP sequencing or genotyping error, we repeated all
195 population genetic analyses after doing a more stringent SNP quality filtering
196 by removing loci with a genotype quality score below 20, a minimum read
197 depth of 10, or with more than 15% missing data.
198
199 First, to characterize population differentiation we estimated F_{ST} using
200 GenAlEx v6.5 (Peakall and Smouse, 2012) (because similar global F_{ST} and R_{ST}
201 estimates were obtained for the SSR dataset, we opted to report F_{ST} values only
202 to allow direct comparisons with the SNP dataset). We then estimated
203 contemporary effective population size for each sampled locality in
204 NeEstimator v2 using the excess heterozygous method (Do *et al.*, 2014). We
205 performed this latter analysis to assess the possibility that differences in local
206 population size underlie differences in genetic variability (Weckworth *et al.*,
207 2013) and/or promote asymmetric effective migration rate (*Nm*).
208
209 In addition, to further examine genetic structure we assessed population
210 admixture and spatial genetic clustering using TESS (Chen *et al.*, 2007). TESS
211 was run using the admixture algorithm and a BYM model (Durand, Jay, *et al.*,
212 2009) with 10 runs per K value, and without using geographic weights. The
213 TESS model, with the lowest DIC was chosen as the optimal model (Durand,
214 Chen, *et al.*, 2009). K values tested ranged from two to the maximum number

215 of sampled localities. Additionally, following Wang et al. (2009), we
216 complemented these analyses with Analyses of Molecular Variance (AMOVA;
217 Excoffier *et al.*, 1992) run in GenAlEx (Peakall and Smouse, 2012) using 9999
218 permutation replicates. We ran these AMOVAs either partitioning the variance
219 into regions based on the spatial genetic clusters previously identified by
220 TESS—to quantify the fraction of the genetic variance explained by these
221 clusters, or leaving it ungrouped (i.e., no regions), for comparison.
222
223 Additionally, we investigated population connectivity by estimating levels
224 of recent migration between sampled localities through the identification of
225 individuals of mixed ancestry using BayesAss (Wilson and Rannala, 2003).
226 BayesAss is a program that uses individual multilocus genotypes and a Markov
227 Chain Monte Carlo (MCMC) algorithm to probabilistically distinguish between
228 immigrants and long-term native individuals (Wilson and Rannala, 2003). We
229 ran BayesAss for 6 million generations using default parameter settings, and
230 discarded the first two million generations as burn-in (Dyer, 2009). For each
231 marker dataset, we repeated this analysis three times (for a total of 18 million
232 generations) and combined the results from the three replicates for our final
233 inference. Then, using a posterior probability cut-off of 0.75 we assign
234 individuals' ancestry. We chose this cut-off value as a minimum credibility
235 score to simultaneously maximize sample size and reliability (stringer

236 thresholds show similar differences between marker sets; results not shown). It
237 is important to note that because of computational limits we had to randomly
238 subsample our set of SNPs to 400 SNPs for this analysis. The same subsampled
239 set was used for the full and reduced (SNPc) analyses.
240
241 **Landscape genetics analyses**
242 To identify the likely landscape features underlying overall population
243 structure of *I. purpurea*, we evaluated the association between landscape
244 features and genetic differentiation based on the full datasets. First, we
245 estimated conditional genetic distances (Dyer *et al.*, 2010) using GeneticStudio
246 (Dyer, 2009). Briefly, conditional genetic distances are measures of pairwise
247 genetic distance derived from population networks, constructed based on the
248 degree of genetic similarity between sampled localities (Dyer and Nason, 2004).
249 They reflect genetic similarity between localities that better capture direct gene
250 flow (i.e., direct migration) as opposed to connectivity driven by step-wise
251 migration through intervening localities (Dyer, 2015). The complexity of the
252 associated conditional genetic network was summarized by their vertex
253 connectivity (White and Harary, 2001), whereas the congruence between
254 networks derived from different marker sets was measured by their structural
255 congruence (a measure of wether the number of congruent edges between
256 networks is greater than expected by chance) (Dyer, 2009).

To assess the association between landscape features and population differentiation, we first converted each landscape layer (climate, crops, elevation, landcover, population density, roads, and soils layers; Table S1) into landscape resistance layers. To do this, each landscape feature in these layers was assigned a resistance value that reflects the difficulty that each feature offers to the movement of gametes or individuals. 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 *et al.*, 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 *et al.*, 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). This model evaluates the association between pairwise genetic distance and landscape cumulative resistance between localities, estimated in Circuitscape (Shah and 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-associated 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 to population differentiation in *I. purpurea* we ran Multiple Regression on Distance Matrices (MRDM; Legendre *et al.*, 1994). Before running these MRDM models, 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

299 ecodist; Goslee and Urban, 2007) using 10,000 permutations to assess
300 significance. We accounted for multiple testing by applying a false recovery
301 rate correction (Benjamini and Hochberg, 1995) using the function *p.adjust* in R
302 (R Core Development Team, 2016).
303
304 These landscape genetic analyses, aimed at identifying the relative influence
305 of natural and human-related landscape features on *I. purpurea*'s connectivity,
306 show several differences between SSR and SNP datasets (see below).
307 Nonetheless, we expect that association patterns that are robust between
308 datasets should accurately reflect the impact of landscape features on gene flow,
309 independent of possible biases introduced by marker idiosyncrasies. Therefore,
310 we focus below on the common biological findings between marker types,
311 while also denoting the most relevant differences.
312
313 **RESULTS**
314 **Population structure**
315 The initial genetic analyses indicated that *I. purpurea* sampled localities
316 were in no violation of Hardy-Weinberg equilibrium (Fig. S1c, d), as evidenced
317 by the small difference between expected and observed heterozygosity (mean
318 He = 0.294±0.014 and 0.250±0.001; mean Ho = 0.291±0.009 and 0.260±0.001,
319 respectively for SSR and SNP datasets). Levels of expected and observed

320 heterozygosity for the SSR dataset were only slightly greater than those
321 estimated for the SNP dataset. Likewise, the estimated mean effective
322 population size per sampled locality was only slightly greater and more variable
323 for the SSR dataset than for the SNP dataset (13.71±5.59, 9.49±0.13,
324 respectively), but in neither case was there salient evidence of a plausible
325 source-sink dynamic, as judged by the similar effective sizes among
326 populations. Average F_{ST} estimates between datasets were also similar (0.151
327 and 0.140, respectively for SSR and SNP datasets; Fig. S2).
328
329 Interestingly, we found that estimates of recent ancestry differed
330 between SSR and SNP datasets. The analysis of the SSR dataset indicated that
331 recent migration among localities seems to be more widespread, with only four
332 localities being primarily constituted of native individuals (Fig. 2a). Across
333 localities, on average 73.65% of individuals were inferred to be 1st or 2nd
334 generation immigrants. In comparison, analysis of the SNP dataset showed that
335 most populations seem to have a more limited number of recent immigrants,
336 and that the relatively few inferred immigrants (on average 27.42% of
337 individuals) did not come exclusively from geographically proximate localities
338 (Fig. 2d). Accordingly, SSR and SNP pruned conditional genetic networks
339 (Dyer and Nason, 2004) indicated different underlying patterns of genetic
340 connectivity (structural congruence = 0.108; Fig. 2b,e). While both were fully

341 closed, the SSR-based network was more interconnected (vertex connectivity:
342 5) than the SNP-based network (vertex connectivity: 0). Further, based on the
343 best TESS models (Fig. S3), widespread admixture was recovered in the SSR
344 dataset (median individual maximum Q-score = 0.51), whereas minimal
345 admixture was identified in the SNP dataset (median individual maximum Q-
346 score = 0.73) (Fig. 2c,f). Finally, grouping individuals according to the
347 corresponding TESS-identified spatial genetic clusters in AMOVA analyses
348 only slightly reduced the variance explained solely by geographic location in
349 both datasets (Table 1).
350
351 Similar to our results from the entire datasets, when we subset the SSR
352 and SNP dataset to the 6 localities in common (SSRc and SNPc datasets), we
353 found no major differences in genetic estimates between the SSRc and SNPc
354 datasets (Table S4), and we again identified differences in the underlying
355 population structure (Fig. S4). Specifically, the SNPc dataset was characterized
356 by a smaller percentage of recent immigrants (28.25%) than the SSRc dataset
357 (44.93%) (Fig. S4a,d), and the corresponding genetic networks were also
358 different from each other (structural congruence = 0.002)—with the SSR-based
359 network being more connected (vertex connectivity = 2) than the SNPc-based
360 network (vertex connectivity = 0) (Fig. S4b,e). Finally, as for the full data, a
361 more admixed genetic composition of individuals was recovered in the SSRc

362 dataset (median individual maximum Q-score = 0.71) than in the SNPc dataset
363 (median individual maximum Q-score = 0.85) (Fig. S4c,f). Also, further
364 confirming the limited spatial structure in this species, using TESS-identified
365 spatial genetic clusters as regions in AMOVA analyses barely reduced the
366 variance explained solely by geographic location when compared to a null
367 model with no regions assigned (Tables S5).
368
369 Similarly, when using a more stringently filtered SNP dataset, which
370 comprised 5811 SNPs, we found that this reduced dataset produced highly
371 similar results to the original SNP dataset (percentage of recent immigrants =
372 25.94%, genetic network vertex connectivity = 0, average individual maximum
373 Q-score = 0.71, and percentage explained by TESS-groupings = 7.31%). Hence,
374 these results using a more rigorous SNP dataset further support the differences
375 in population structure inferences between SSR and SNP data.
376
377 **Landscape genetics**
378 Both the SNP and SSR datasets provide evidence that human-impacted
379 landscapes play an important role in shaping genetic connectivity in *I.*
380 *purpurea*. In both sets of MLPE.lmm models, null (geographic distance),
381 natural (climate, elevation, and soils), and human-related landscapes (landcover
382 and human population density) were identified as significant ($p<=0.05$) or

383 marginally significant (0.05<p<=0.1) predictors of genetic differentiation
384 between localities. Interestingly, the variables with the greatest association
385 coefficient and lowest AICc value in both the SSR and SNP models were
386 human-related variables (landcover and human population density, respectively;
387 Table 2). However, when considering all variables together in a multivariate
388 manner—while accounting for geographic distance—human population density,
389 local effective population size, and different aspects of climate were the only
390 variables that remained as significant or marginally significant predictors of
391 genetic differentiation across both SSR and SNP datasets (Table 2). In contrast,
392 elevation and soil were identified as significant or marginally significant
393 predictors only in the SNP dataset. Both multivariate regressions also differed
394 in the proportion of the variance explained (MRDM R^2 for SSR and SNP
395 dataset were 0.109 ($F_{1,29}$ = 3.654, p-val. = 0.063) and 0.532 ($F_{1,6}$ = 1.932, p-val.
396 = 0.113), respectively).
397
398 In summary, across datasets, results indicated that human-population-
399 density resistance was robustly associated with differentiation among *I.*
400 *purpurea*'s populations, with sparsely to moderately populated areas identified
401 as more conducive areas for migration and potential corridors available between
402 all regions (Fig. S5b). In contrast, climatic variables produced potential barriers
403 to gene flow, with temperature temporal gradients isolating the northernmost

404 localities from the rest in the SSR dataset, and precipitation seasonality
405 isolating the eastern and western localities in the SNP dataset (Fig. S5a).
406 Finally, local effective population size was also a significant predictor in both
407 datasets, with population size inversely associated with genetic differentiation
408 (Table 2).
409

410 **DISCUSSION**

411 Our results reveal that broadly distributed populations of *I. purpurea* are
412 not genetically isolated from each other. They also suggest the existence of
413 long-distance and putatively human-mediated migration between localities. At
414 the scale of our analyses, the regional agricultural matrix does not seem to have
415 an overarching impact on population connectivity in this species, despite *I.*
416 *purpurea*'s tight link to agricultural fields. Instead, genetic connectivity in this
417 species seems to be primarily influenced by climate and human population
418 density. The effective population size (Ne) also appears to influence population
419 connectivity in this species, which suggests a plausible additional effect of
420 genetic drift on the effectiveness of gene flow (Weckworth *et al.*, 2013). Taken
421 together, these results highlight the significant interplay between human-driven
422 and natural landscapes in structuring *I. purpurea* populations.
423

424 **Population connectivity patterns**

425 Despite *I. purpurea*'s expected low natural seed dispersal, due to heavy,
426 gravity dispersed seeds, and large and patchy distribution, we found evidence of
427 limited genetic differentiation through the species' range and overall weak
428 geographic structure. In fact, population differentiation was uneven (pairwise
429 F_{ST} values ranged from 0.02 to 0.24), and only partially dependent on the
430 geographic distance between populations. Further, genetic networks for both
431 datasets were fully closed, suggesting the existence of direct or indirect gene
432 flow among all sampled populations. In line with this finding, admixed
433 individuals were present in all sampled localities (although levels of admixture
434 vary for SSR and SNP datasets), and several instances of recent short- and long-
435 distance migration were recovered. Nevertheless, the evidence of
436 interconnectedness we uncovered is likely also influenced by the shared
437 evolutionary history of populations and, thus, shared ancestral genetic variation
438 likely confounds our estimates of genetic differentiation (Marko and Hart,
439 2011). Given the relatively recent invasion of the US by *I. purpurea* and its
440 inclusion in horticultural trade (Fang *et al.*, 2013), it is possible that historical
441 connectivity between populations due to human transport maintained gene flow
442 between populations after its introduction into the US (Mack, 1991).
443 Alternatively, recurrent colonization from few genetically similar sources might
444 instead have resulted in partial homogenization of otherwise isolated
445 populations (Dlugosch and Parker, 2008). To disentangle the relative

446 contribution of recent gene flow from that of historical patterns of population
447 connectivity, empirical estimates of inter-population migration (e.g., through
448 the use of pollen traps) or direct measurements of gene flow (e.g., paternity
449 analysis) would be needed.
450
451 Our estimates of differentiation as well as heterozygosity and allelic
452 richness differ from other agricultural weedy species. Specifically, our
453 estimates suggest that *I. purpurea* has lower genetic diversity (as measured by
454 allelic richness and/or heterozygosity) than primarily outcrossing annual weeds
455 (e.g., ragweed (Martin *et al.*, 2016) and black-grass (Menchari *et al.*, 2007)),
456 including many broadly distributed invasives (Dlugosch and Parker, 2008). This
457 finding is potentially explained by both differences in reproduction system, as
458 outcrossing plants often show greater genetic diversity than plants with mixed
459 or primarily selfing reproduction (Hamrick and Godt, 1996), and the recent
460 population bottlenecks likely experienced by *I. purpurea* (Kuester *et al.*, 2016).
461 Genetic differentiation found in this study (as measured by F_{ST}), on the other
462 hand, was relatively low for a broadly distributed gravity-dispersed plant
463 (Hamrick and Godt, 1996). This contrasts with many broadly distributed weeds,
464 which show moderate to high F_{ST} values (0.15 - 0.48) likely due to dispersal
465 limitations over distances above those commonly allowed by natural dispersal
466 agents (Schmidt *et al.*, 2009; Treier and Müller-Schärer, 2011). This is the case,

467 for example, of invasive weeds such as weedy *Silene* (Barluenga *et al.*, 2011)
468 that rely on specialist pollinators or seed dispersers that might not be present in
469 the invaded range. Considering the natural history of *I. purpurea*—heavy seeds,
470 bumblebee pollination (Osborne *et al.*, 1999; Schulke and Waser, 2001), and
471 strong agricultural and horticultural ties (Defelice, 2001), this finding suggests
472 that human-aided dispersal presumably contributes to maintain connectivity in
473 this species.
474
475 **Landscape features influencing population connectivity**
476 Our assessment of the association between population differentiation
477 and landscape resistance identified several predictors of population
478 connectivity. Of the landscape features examined, climate and human
479 population density are the only robust predictors, suggesting a role for both
480 human activities and climatic barriers in shaping population connectivity in this
481 species. Specifically, our results suggest that while climatic variables provide
482 some resistance to migration, scarcely to moderately populated areas (i.e., those
483 corresponding with rural areas) offer potential corridors between all sampled
484 regions. The existence of climatic dispersal barriers presumably arises in
485 response to physiological preferences or local adaptations of the species
486 (Cimalová and Lososová, 2009); humans seem to oppose these natural
487 limitations and help *I. purpurea* overcome climatic barriers (e.g., by allowing

488 dispersal between northern and southern populations that are separated by areas
489 of high temperature-related resistance to gene flow). Indeed, human-mediated
490 migration is expected to be particularly prevalent among wild populations of
491 species with commercial value such as *I. purpurea* (Mack, 1991; Defelice,
492 2001).
493
494 Population connectivity of species in human-dominated environments is
495 potentially also affected by human-induced changes in population size (Méndez
496 *et al.*, 2014). Specifically, by influencing population size, anthropogenic
497 activities condition the effectiveness of migration as it depends on migration
498 relative rate (*Nm*) (Wright, 1931). Furthermore, population size differences
499 influence overall dispersal spatial dynamics because the number and direction
500 of migrants depend on local population sizes (e.g., proportionately greater
501 number of migrants move from densely to sparsely populated areas than vice
502 versa; Lenormand, 2002). In line with these expectations, we found that
503 effective population size is also a significant predictor of population
504 differentiation in this species. Considering the prevalence of weed management
505 practices (e.g., tillage, herbicide application) and their effect on weed
506 population sizes (Kuester *et al.*, 2016), it is likely that these anthropogenic
507 activities play a significant role in controlling the rate of population

508 differentiation in weeds. Thus, all evidence suggests a predominant role of
509 human activities in shaping *I. purpurea*'s current genetic structure.
510
511 Relatively few studies have explored how landscape features impact
512 population connectivity in weeds at large spatial scales, making it hard to
513 evaluate how distinctive *I. purpurea*'s response to climatic and anthropogenic
514 factors may be from that of other weeds. Yet, simulations modeling short-
515 distance dispersal based on spatial distances and landscape configuration have
516 identified dispersal capabilities and landscape use (including availability of
517 disturbed habitats and distribution of crop types) to be the most prevalent
518 determinants of local level connectivity in several weed systems (Woolcock and
519 Cousens, 2000; Fénart *et al.*, 2007; Will and Tackenberg, 2008). For instance,
520 using a ~10km^2 aerial photograph to inform a spatial mechanistic model of
521 Canadian horseweed's interfield dispersal, Dauer et al. (Dauer *et al.*, 2009)
522 showed that distribution of suitable habitat primarily determined the rate and
523 extent of this weed's dispersal at this spatial scale. In agreement with these
524 predictions, empirical data show that local dispersal in mountain pasture weed
525 is heavily influenced by the spatial distribution of human-dominated landscapes
526 and the opportunities for interfield contamination (Treier and Müller-Schärer,
527 2011). Yet, while neither landcover nor crop types distribution were identified
528 as significant predictors in our multivariate analyses, we cannot rule out the

529 possibility of high local gene flow at the scale of contiguous agricultural fields
530 mediated by these landscape features given the scale of our analyses.
531
532 **Marker-specific inferences**
533 Our results identified interesting differences between marker types in
534 terms of the inferred population structure of *I. purpurea*. The differences
535 uncovered between datasets are at first glance unexpected as all loci in a
536 species' genome evolve under a common evolutionary history (Payseur and
537 Cutter, 2006). Nevertheless, differences between marker types have been
538 similarly observed in other studies (Dixon *et al.*, 2011; Martin *et al.*, 2016). For
539 example, several studies have recovered F_{ST} estimate differences when using
540 SNP or SSR loci on the same set of samples (Coates *et al.*, 2009; Gärke *et al.*,
541 2012), presumably due to mutation rates and genomic representation differences
542 of SSR and SNP loci (Payseur and Cutter, 2006; Coates *et al.*, 2009). Here, the
543 population structure differences we uncovered may likewise be related to
544 marker-specific rates of mutation, drift and/or marker-specific biases—such as
545 greater ascertainment bias on SSR data (Väli *et al.*, 2008; Defaveri *et al.*,
546 2013)—and not likely caused by the different SSR and SNP dataset sampling.
547 Because of these intrinsic marker differences, both markers could provide
548 complementary information (Payseur and Cutter, 2006). While our
549 identification of robust environmental predictors of genetic differentiation for

550 SSR and SNP datasets—despite differences in underlying population genetic
551 patterns—is encouraging, further work is needed to reconcile traditional
552 landscape genetics studies based on a few highly variable markers with
553 increasing landscape genomics studies based on thousands of SNPs.
554
555 **Conclusions**
556 While our results suggest that *I. purpurea* experiences moderate inter-
557 population connectedness and potential long-distance dispersal, we cannot rule
558 out the influence of historical relatedness on the observed genetic patterns. To
559 determine the amount of current-day gene flow that occurs between
560 populations, direct estimates of gene flow need to be examined. Regardless, the
561 limited population structure recovered as well as the identification of human
562 population density as a significant predictor of population differentiation calls
563 attention to the need for investigating the possible impact of human-mediated
564 gene flow on the evolutionary path of this species—including its response to
565 selection and the likelihood of further naturalization. In particular, it remains to
566 be investigated whether the pattern of maintained connectivity we identify here
567 could facilitate the success of this weed (e.g., by introducing relevant genetic
568 variants; Kolbe *et al.*, 2004) or reduce the fitness of local populations (Keller *et*
569 *al.*, 2000). What seems clear from this study is that human-aided migration
570 presumably is an important component of gene flow between populations,

which may counter the isolating effects of natural environmental barriers and genetic drift.

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.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA ARCHIVING

All data generated is in the process of being archived in Dryad. The corresponding doi would be made available upon acceptance.

586 **REFERENCES**

587 Auffret AG, Cousins SAO (2013). Humans as long-distance dispersers of rural
588 plant communities. *PLoS One* **8**: e62763.
589 Barluenga M, Austerlitz F, Elzinga JA, Teixeira S, Goudet J, Bernasconi G
590 (2011). Fine-scale spatial genetic structure and gene dispersal in *Silene*
591 *latifolia*. *Heredity* **106**: 13–24.
592 Barrowclough GF (1980). Gene flow, effective population sizes and genetic
593 variance components in birds. *Evolution* **34**: 789–798.
594 Bataille A, Cunningham AA, Cruz M, Cedeño V, Goodman SJ (2011).
595 Adaptation, isolation by distance and human-mediated transport determine
596 patterns of gene flow among populations of the disease vector *Aedes*
597 *taeniorhynchus* in the Galapagos Islands. *Infect Genet Evol* **11**: 1996–
598 2003.
599 Baucom, Mauricio R (2008). The evolution of novel herbicide tolerance in a
600 noxious weed: The geographic mosaic of selection. *Evol Ecol* **22**: 85–101.
601 Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a
602 practical and powerful approach to multiple testing. *J R Stat Soc B* **57**:
603 289–300.
604 Bushnell B (2014). BBMap short read aligner.
605 *http://sourceforge.net/projects/bbmap/*.
606 Chaney L, Baucom RS (2012). The evolutionary potential of Baker's weediness

607 traits in the common morning glory, *Ipomoea purpurea* (Convolvulaceae).
608 *Am J Bot* **99**: 1524–1530.
609 Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms
610 ascertaining spatial population structure: A new computer program and a
611 comparison study. *Mol Ecol Notes* **7**: 747–756.
612 Cimalová Š, Lososová Z (2009). Arable weed vegetation of the northeastern
613 part of the Czech Republic: Effects of environmental factors on species
614 composition. *Plant Ecol* **203**: 45–57.
615 Clobert J, Baguette M, Benton TG, Bullock JM, Ducatez S (2012). *Dispersal*
616 *Ecology and Evolution*, 1st ed. Oxford University Press: Oxford.
617 Coates BS, Sumerford D V., Miller NJ, Kim KS, Sappington TW, Siegfried
618 BD, *et al.* (2009). Comparative performance of single nucleotide
619 polymorphism and microsatellite markers for population genetic analysis.
620 *J Hered* **100**: 556–564.
621 Culpepper AS (2006). Glyphosate-induced weed shifts. *Weed Technol* **20**: 277–
622 281.
623 Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006). Gene flow in
624 complex landscapes: testing multiple hypotheses with causal modeling. *Am*
625 *Nat* **168**: 486–499.
626 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.*
627 (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–

2158.

Dauer JT, Luschei EC, Mortensen DA (2009). Effects of landscape composition on spread of an herbicide-resistant weed. *Landsc Ecol* **24**: 735–747.

Defaveri J, Viitaniemi H, Leder E, Merilä J (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Mol Ecol Resour* **13**: 377–392.

Defelice MS (2001). Tall morningglory, *Ipomoea purpurea* (L .) roth-flower or foe? *Weed Technol* **15**: 601–606.

Dixon CJ, Kapralov M V., Filatov DA (2011). Gene flow and species cohesion following the spread of *Schiedea globosa* (Caryophyllaceae) across the Hawaiian Islands. *J Evol Biol* **24**: 1–11.

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

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour* **14**: 209–214.

Durand E, Chen C, Francois O (2009). Tess version 2.3 - Reference Manual August 2009 ∗. : 1–30 [http://membres–imag.fr/Olivier.Francois.

Durand E, Jay F, Gaggiotti OE, François O (2009). Spatial inference of

649 admixture proportions and secondary contact zones. *Mol Biol Evol* **26**:
650 1963–1973.
651 Dyer RJ (2009). GeneticStudio: A suite of programs for spatial analysis of
652 genetic-marker data. *Mol Ecol Resour* **9**: 110–113.
653 Dyer RJ (2015). Population graphs and landscape genetics. *Annu Rev Ecol Evol*
654 *Syst* **46**: 327–342.
655 Dyer RJ, Nason JD (2004). Population graphs: The graph theoretic shape of
656 genetic structure. *Mol Ecol* **13**: 1713–1727.
657 Dyer RJ, Nason JD, Garrick RC (2010). Landscape modelling of gene flow:
658 improved power using conditional genetic distance derived from the
659 topology of population networks. *Mol Ecol* **19**: 3746–3759.
660 Epperson BK, Clegg MT (1986). Spatial-autocorrelation analysis of flower
661 color polymorphisms within substructured populations of morning glory
662 (*Ipomoea purpurea*). *Am Nat* **128**: 840–858.
663 Everman E, Klawinski P (2013). Human-facilitated jump dispersal of a non-
664 native frog species on Hawai'i Island. *J Biogeogr* **40**: 1961–1970.
665 Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance
666 inferred from metric distances among DNA haplotypes: Application to
667 human mitochondrial DNA restriction data. *Genetics* **491**: 479–491.
668 Fang Z, Gonzales AM, Durbin ML, Meyer KKT, Miller BH, Volz KM, *et al.*
669 (2013). Tracing the geographic origins of weedy *Ipomoea purpurea* in the

670 Southeastern United States. *J Hered* **104**: 666–677.
671 Fénart S, Austerlitz F, Cuguen J, Arnaud JF (2007). Long distance pollen-
672 mediated gene flow at a landscape level: The weed beet as a case study.
673 *Mol Ecol* **16**: 3801–3813.
674 Fountain T, Duvaux L, Horsburgh G, Reinhardt K, Butlin RK (2014). Human-
675 facilitated metapopulation dynamics in an emerging pest species, *Cimex*
676 *lectularius*. *Mol Ecol* **23**: 1071–1084.
677 Gärke C, Ytournel F, Bed'Hom B, Gut I, Lathrop M, Weigend S, *et al.* (2012).
678 Comparison of SNPs and microsatellites for assessing the genetic structure
679 of chicken populations. *Anim Genet* **43**: 419–428.
680 Goslee SC, Urban DL (2007). The ecodist package for dissimilarity-based
681 analysis of ecological data. *J Stat Softw* **22**: 1–19.
682 Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity
683 in plant species. *Philos Trans R Soc B Biol Sci* **351**: 1291–1298.
684 Hodkinson DJ, Thompson K, Journal T, Dec N (2007). Plant dispersal: The role
685 of man. *J Appl Ecol* **34**: 1484–1496.
686 Holderegger R, Buehler D, Gugerli F, Manel S (2010). Landscape genetics of
687 plants. *Trends Plant Sci* **15**: 675–683.
688 Keller M, Kollmann J, Edwards PJ (2000). Genetic introgression from distant
689 provenances reduces fitness in local weed populations. *J Appl Ecol* **37**:
690 647–659.

691 Kolbe JJ, Glor RE, Rodríguez Schettino L, Lara AC, Larson A, Losos JB
692 (2004). Genetic variation increases during biological invasion by a Cuban
693 lizard. *Nature* **431**: 177–81.
694 Kuester A, Chang S-M, Baucom RS (2015). The geographic mosaic of
695 herbicide resistance evolution in the common morning glory, *Ipomoea*
696 *purpurea*: Evidence for resistance hotspots and low genetic differentiation
697 across the landscape. *Evol Appl* **8**: 821–833.
698 Kuester A, Wilson A, Chang S-M, Baucom RS (2016). A resurrection
699 experiment finds evidence of both reduced genetic diversity and potential
700 adaptive evolution in the agricultural weed *Ipomoea purpurea*. *Mol Ecol*
701 **25**: 4508–4520.
702 Legendre P, Lapointe FJ, Casgrain P (1994). Modeling brain evolution from
703 behavior: A permutational regression approach. *Evolution* **48**: 1487–1499.
704 Lenormand T (2002). Gene flow and the limits to natural selection. *Trends Ecol*
705 *Evol* **17**: 183–189.
706 Li H (2011). A statistical framework for SNP calling, mutation discovery,
707 association mapping and population genetical parameter estimation from
708 sequencing data. *Bioinformatics* **27**: 2987–2993.
709 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The
710 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:
711 2078–2079.

712 Mack RN (1991). The commercial seed trade: An early disperser of weeds in
713 the United States. *Econ Bot* **45**: 257–273.
714 Mack RN, Lonsdale WM (2001). Humans as global plant dispersers: Getting
715 more than we bargained for. *Bioscience* **51**: 95.
716 Marko PB, Hart MW (2011). The complex analytical landscape of gene flow
717 inference. *Trends Ecol Evol* **26**: 448–456.
718 Martin MD, Olsen MT, Samaniego JA, Zimmer EA, Gilbert MTP (2016). The
719 population genomic basis of geographic differentiation in North American
720 common ragweed (*Ambrosia artemisiifolia* L.). *Ecol Evol* **6**: 3760–3771.
721 McRae B (2006). Isolation by resistance. *Evolution* **60**: 1551–1561.
722 Meehan TD, Werling BP, Landis DA, Gratton C (2011). Agricultural landscape
723 simplification and insecticide use in the Midwestern United States. *Proc*
724 *Natl Acad Sci U S A* **108**: 11500–11505.
725 Menchari Y, Délye C, Le Corre V (2007). Genetic variation and population
726 structure in black-grass (*Alopecurus myosuroides* Huds.), a successful,
727 herbicide-resistant, annual grass weed of winter cereal fields. *Mol Ecol* **16**:
728 3161–3172.
729 Méndez M, Vögeli M, Tella JL, Godoy JA (2014). Joint effects of population
730 size and isolation on genetic erosion in fragmented populations: Finding
731 fragmentation thresholds for management. *Evol Appl* **7**: 506–518.
732 Mitchell M (1996). *An introduction to genetic algorithms*. MIT Press:

733 Cambridge Mass.
734 Molecular Ecology Resources Primer Development Consortium, Aksoy S,
735 Almeida-Val VMF, Azevedo VCR, Baucom R, Bazaga P, *et al.* (2013).
736 Permanent genetic resources added to Molecular Ecology Resources
737 database 1 October 2012–30 November 2012. *Mol Ecol Resour* **13**: 341–
738 343.
739 Navas ML (2012). Trait-based approaches to unravelling the assembly of weed
740 communities and their impact on agro-ecosystem functioning. *Weed Res*
741 **52**: 479–488.
742 Osborne JL, Clark SJ, Morris RJ, Williams IH, Riley JR, Smith AD, *et al.*
743 (1999). A landscape-scale study of bumble bee foraging range and
744 constancy, using harmonic radar. *J Appl Ecol* **36**: 519–533.
745 Payseur BA, Cutter AD (2006). Integrating patterns of polymorphism at SNPs
746 and STRs. *Trends Genet* **22**: 424–429.
747 Peakall R, Smouse PE (2012). GenAlEx 6.5: Genetic analysis in Excel.
748 Population genetic software for teaching and research—an update.
749 *Bioinformatics* **28**: 2537–2539.
750 Peterman WE (2014). ResistanceGA: An R package for the optimization of
751 resistance surfaces using genetic algortihms. *bioRxiv*:
752 https://doi.org/10.1101/007575.
753 Peterman WE, Connette GM, Semlitsch RD, Eggert LS (2014). Ecological

resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Mol Ecol* **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. *Ecol Econ* **52**: 273–288.

R Core Development Team (2016). R: a language and environment for statistical computing, v3.2.4. : [http://www.r-project.org].

Ricciardi A (2007). Are modern biological invasions an unprecedented form of global change? *Conserv Biol* **21**: 329–336.

Schmidt T, Arens P, Smulders MJM, Billeter R, Liira J, Augenstein I, *et al.* (2009). Effects of landscape structure on genetic diversity of *Geum urbanum* L. populations in agricultural landscapes. *Flora Morphol Distrib Funct Ecol Plants* **204**: 549–559.

Schulke B, Waser NM (2001). Long-distance pollinator flights and pollen dispersal between populations of *Delphinium nuttallianum*. *Oecologia* **127**: 239–245.

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.

Shah VB, McRae BH (2008). Circuitscape: A tool for landscape ecology. *Proc 7th Python Sci Conf*: 62–65.

775 Siliceo-Cantero HH, García A, Reynolds RG, Pacheco G, Lister BC (2016).
776 Dimorphism and divergence in island and mainland *Anoles*. *Biol J Linn*
777 *Soc* **118**: 852–872.
778 Simberloff D (2013). *Invasive species: what everyone needs to know*. Oxford
779 University Press: Oxford.
780 Slatkin M (1985). Gene flow in natural populations. *Annu Rev Ecol Syst* **16**:
781 393–430.
782 Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010). Use of
783 resistance surfaces for landscape genetic studies: Considerations for
784 parameterization and analysis. *Mol Ecol* **19**: 3576–3591.
785 Tiffin P, Rausher M (1999). Genetic constraints and selection acting on
786 tolerance to herbivory in the common morning glory *Ipomoea purpurea*.
787 *Am Nat* **154**: 700–716.
788 Trakhtenbrot A, Nathan R, Perry G, Richardson DM (2005). The importance of
789 long-distance dispersal in biodiversity conservation. *Divers Distrib* **11**:
790 173–181.
791 Treier UA, Müller-Schärer H (2011). Differential effects of historical migration,
792 glaciations and human impact on the genetic structure and diversity of the
793 mountain pasture weed *Veratrum album* L. *J Biogeogr* **38**: 1776–1791.
794 Väli Ü, Einarsson A, Waits L, Ellegren H (2008). To what extent do
795 microsatellite markers reflect genome-wide genetic diversity in natural

796 populations? *Mol Ecol* **17**: 3808–3817.
797 Wang IJ, Bradburd GS (2014). Isolation by environment. *Mol Ecol* **23**: 5649–
798 5662.
799 Wang IJ, Savage WK, Shaffer HB (2009). Landscape genetics and least-cost
800 path analysis reveal unexpected dispersal routes in the California tiger
801 salamander (*Ambystoma californiense*). *Mol Ecol* **18**: 1365–1374.
802 Webster TM, Nichols RL (2012). Changes in the prevalence of weed species in
803 the major agronomic crops of the southern United States: 1994/1995 to
804 2008/2009. *Weed Sci* **60**: 145–157.
805 Weckworth B V, Musiani M, Decesare NJ, Mcdevitt AD, Hebblewhite M,
806 Mariani S (2013). Preferred habitat and effective population size drive
807 landscape genetic patterns in an endangered species. *Proc R Soc B* **280**: 1–
808 9.
809 White DR, Harary F (2001). The cohesiveness of blocks in social networks:
810 node connectivity and conditional density. *Sociol Methodol* **31**: 305–359.
811 Will H, Tackenberg O (2008). A mechanistic simulation model of seed
812 dispersal by animals. *J Ecol* **96**: 1011–1022.
813 Wilson GA, Rannala B (2003). Bayesian inference of recent migration rates
814 using multilocus genotypes. *Genetics* **163**: 1177–1191.
815 Woolcock JL, Cousens R (2000). A mathematical analysis of factors affecting
816 the rate of spread of patches of annual weeds in an arable field. *Weed Sci*

48: 27–34.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Wright S (1949). The genetical structure of populations. *Ann Eugen* **15**: 323–354.

822 **TITLES AND LEGENDS OF FIGURES**

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

827

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

833

834 **Figure 1**. Distribution of *Ipomoea purpurea*'s sampled localities. Sample
835 sizes for both SSR (black bars) and SNP (white bars) datasets are indicated
836 (locality numbers are given in squares). Elevation is provided as
837 background.

838

839 **Figure 2**. Inferred population connectivity. The estimated origin of
840 individuals for each sampled locality (sink) is depicted according to the
841 locality they were inferred to have originated from (source) (a, d). The

842 color of each cell in these plots depicts the relative proportion of
843 individuals in the sink population that were estimated to be recent
844 immigrants from each locality along the x-axis. Cells on the minor diagonal
845 correspond to the proportion of native individuals. Pruned conditional
846 genetic networks (b, e) and posterior estimates of admixture proportion
847 identified by TESS analysis (c, f) are also displayed. The top row shows
848 SSR-based results, the bottom shows the SNP-based results. Locality
849 numbers follow Fig. 1. Localities shared between SSR and SNP datasets are
850 denoted by colored arrows (for a similar figure based exclusively on
851 these shared localities, see Fig. S4).

852 **TABLE 1.**

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)

853

854 **TABLE 2.**

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** (<0.001)	**1.358** (0.003)	0.340 (0.218)	3.887 (0.184)

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

855

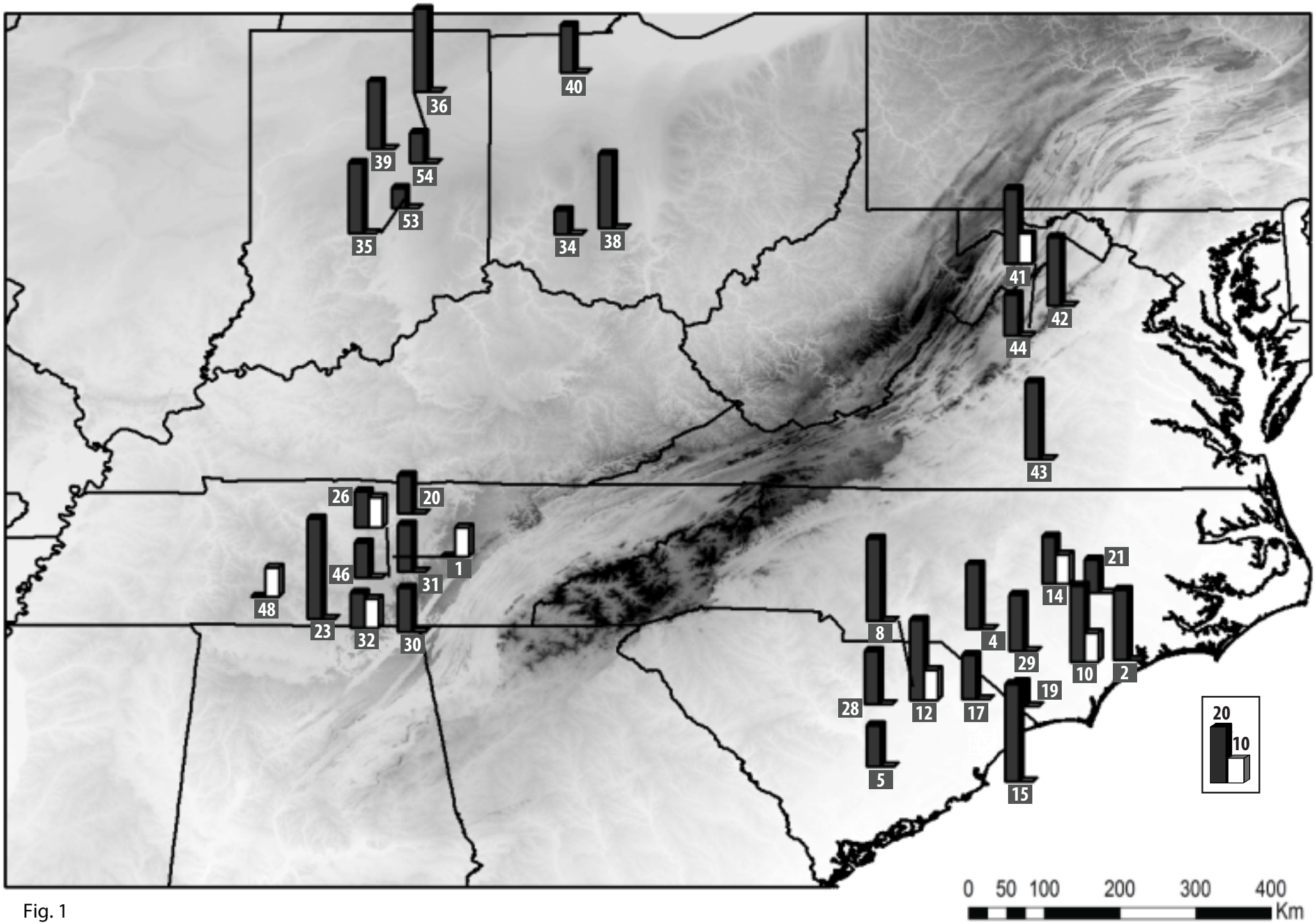
36
40
39
54
53
35
34
38
41
42
44
43
26
20
46
31
1
48
23
32
30
8
28
5
12
17
4
29
19
15
14
21
10
2
20
10
0 50 100 200 300 400
Km

Fig. 1

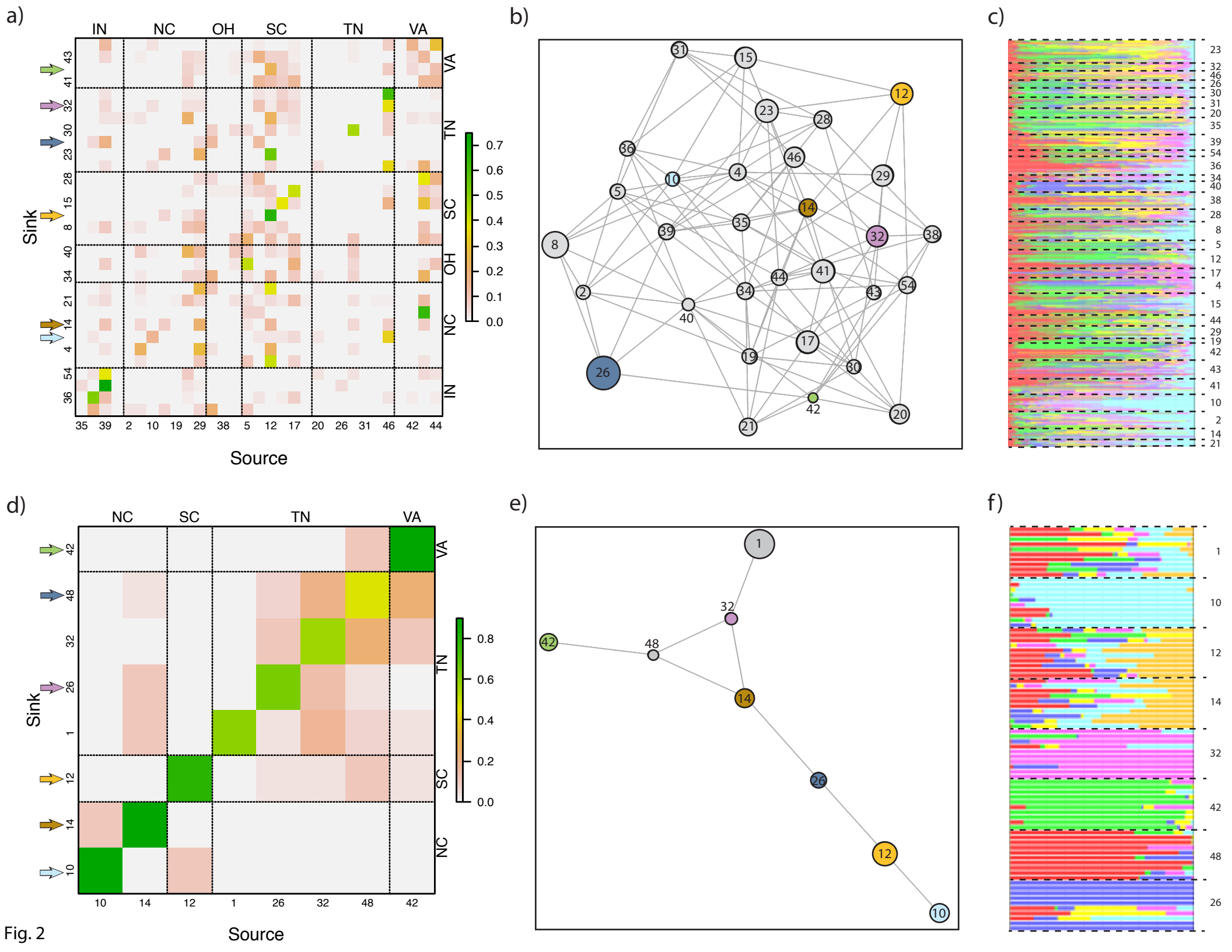

Fig. 2