1   **Landscape connectivity of a noxious invasive weed:**

2   **human-aided or natural dispersal?**

3

4   Diego F. Alvarado-Serrano (corresponding author)
5   Department of Ecology and Evolutionary Biology
6   University of Michigan
7   E-mail: dalvarad@umich.edu
8
9   Megan Van Etten
10   Department of Ecology and Evolutionary Biology
11   University of Michigan
12   E-mail: mvanette@umich.edu
13
14   Shu-Mei Chang
15   Department of Plant Biology
16   University of Georgia
17   E-mail: smchang@uga.edu
18
19   Regina S. Baucom
20   Department of Ecology and Evolutionary Biology
21   University of Michigan
22   E-mail: rsbaucom@umich.edu
23

24   **Landscape connectivity of a noxious invasive weed: human-aided or natural dispersal?**

25

26   **ABSTRACT**

27   Examining how the landscape may influence gene flow is at the forefront of

28   understanding population differentiation and adaptation. Such understanding is crucial in light of

29   ongoing environmental changes and the elevated risk of ecosystems alteration. In particular,

30   knowledge of how humans may influence the structure of populations is imperative to elucidate

31   their role in shaping the evolution of other species, and specifically how humans may alter the

32   balance between genetic drift and selection. Here we characterize the population genetic structure

33   of *Ipomoea purpurea*, a noxious invasive weed, and assess the relative roles of natural and

34   human-driven landscapes on genetic differentiation. By combining rigorous statistical analyses

35   and a combination of different molecular markers, we detect both common and marker-specific

36   patterns of genetic connectivity and identify human-aided migration as an important component

37   shaping the evolutionary history of this species. In particular, we identified human population

38   density as an important predictor of pairwise population differentiation, suggesting that the

39   agricultural and/or horticultural trade may be involved in maintaining some level of connectivity

40   across agricultural fields. Climatic variation, primarily temperature, appears as an additional

41   predictor when considering agricultural fields in the northern United States. We discuss the

42   implications of these results and the approach we followed in the context of understanding

43   agricultural weed and invasive species' expansions, as well as the challenges and promises of

44   current landscape genetics research.

45

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

47 **INTRODUCTION**

48   Elucidating routes and levels of migration between subpopulations of a species is

49   essential to understand the interplay between gene flow, adaptation, genetic drift, and selection,

50   and hence the forces that shape its evolutionary trajectory (Barrowclough 1980; Slatkin 1985).

51   Landscape features—such as rivers, mountain ranges, crop fields, and urban areas—can impact

52   levels of gene flow between populations by determining dispersal rates and routes (McRae 2006;

53   Cushman *et al.* 2006) as well as by influencing the likelihood of successful establishment of

54   immigrants (Nosil *et al.* 2008; Sexton *et al.* 2014; Wang & Bradburd 2014). Landscape features

55   can also indirectly condition the effect of gene flow through its effect on local effective

56   population sizes since the actual role that migration plays in the evolution of a species is driven

57   by the fraction of the local population size that correspond to immigrants (Wright 1949; Slatkin

58   1985). Consequently, the landscape, loosely defined as an area with spatially variable biotic and

59   abiotic factors (Holderegger *et al.* 2010), creates the stage for spatially heterogeneous functional

60   population connectivity, conditioned by species' specific physiological tolerances and behavioral

61   preferences (Clobert *et al.* 2012). In this way, the landscape plays a pivotal role in the evolution

62   of species.

63

64   Previous studies have identified common responses of species to their surrounding

65   landscape across a range of very different systems, including evidence of frequent long distance

66   dispersal (Berthouly-Salazar *et al.* 2013), dispersal over large unsuitable areas (Manel &

67   Holderegger 2013), and a prevalent mixed effect of local and intervening landscapes on

68   structuring populations (Sexton *et al.* 2014). Such findings have led to novel emergent

69   evolutionary hypotheses that better represent the complex interaction of processes determining

3

70   the evolutionary fate of natural populations (Dyer 2015a). However, our current knowledge of the

71   interplay between landscape features and the genetic structure of populations comes mostly from

72   human-avoider species (Blair 2001) facing indirect human impact (e.g., Culley *et al.* 2007;

73   McRae & Beier 2007; Diniz-Filho *et al.* 2009; but see Harris *et al.* 2016). In contrast to species

74   that almost exclusively depend on natural dispersal agents, species in heavily human-dominated

75   ecosystems may exploit human activities to maintain gene flow among populations and expand

76   their ranges (Everman & Klawinski 2013; Fountain *et al.* 2014). Such species may be capable of

77   maintaining population connectivity over vast geographic ranges (Trakhtenbrot *et al.* 2005) by

78   overcoming landscape features that would otherwise represent natural barriers and reach dispersal

79   distances that could be orders of magnitude greater than those attain under natural agents or do it

80   under much smaller time frames (Mack & Lonsdale 2001; Ricciardi 2007). In this way, by

81   facilitating dispersal humans have the potential to: i) condition the balance between drift and

82   selection (Slatkin 1985; Lenormand 2002), ii) introduce relevant genetic variation to local

83   populations (Kolbe *et al.* 2004), iii) prevent local extinction or favor recolonization (Fountain *et*

84   *al.* 2014), and alter the overall genetic constitution of populations (Bataille *et al.* 2011).

85

86   Human-aided migration—intentional or unintentional— is particularly prevalent in plants

87   (Hodkinson *et al.* 2007; Wichmann *et al.* 2009; Auffret & Cousins 2013), where it has had major

88   impacts on the distribution of species and stability of communities (Simberloff 2013 and

89   references therein). Yet, the open question remains: how important is human-aided migration

90   relative to migration through other more natural means in maintaining population connectivity in

91   plant populations? This question is further complicated in plants because in this group dispersal is

92   a two-step process that involves the movement of both pollen and/or seeds (Holderegger *et al.*

93   2010), and these separate vectors of gene flow entail the movement of different genetic states

94   (haploid vs. diploid) as well as different dispersal agents and mechanisms (i.e., wind, water,

95   animal-mediated) (Levin & Kerster 1974).

96

97   A particularly amenable system to study the interaction between natural and human-aided

98   dispersal comes from agricultural weed populations. Agricultural weeds face a highly dynamic

99   landscape characterized by frequent spatial rearrangements (expansion of agricultural front,

100   increased fragmentation) and a constantly changing environment (crop rotation, agricultural

101   chemical use, climatic abnormalities) (Menchari *et al.* 2007; Meehan *et al.* 2011) that certainly

102   impact their opportunities for survival and local adaptation through its effect on population

103   connectivity (Margosian *et al.* 2009). Under these conditions, human-aided migration is expected

104   to be critical for weeds' success, but knowledge on how or if weedy plant populations are able to

105   maintain connectivity through such complex landscape matrix of croplands, grasslands, natural

106   and urban areas is limited. This is a striking omission given that agricultural weeds impose severe

107   economic costs (on the order of 33B USD per year in US agriculture alone, Pimentel *et al.* 2005).

108

109   As a first step into investigating the impact of human activities on structuring genetic

110   diversity, we estimate the intensity and extent of migration from genetic data and evaluate how

111   multiple landscape features influence genetic connectivity of a noxious agricultural weed, *I.*

112   *purpurea*, using two different sets of molecular markers (nuclear microsatellites and a genome-

113   wide panel of SNPs). Specifically, we ask the following questions: 1) Which natural or human-

114   influenced landscape features—soils, elevation, climate, landcover, crop types, human population

115   density—promote or constrain genetic connectivity between populations of this agricultural

5

116 weed? 2) Do human-associated landscapes disproportionally influence genetic connectivity in

117 this species, suggesting that this weed's evolutionary fate is strongly linked to human-aided

118 dispersal?, and 3) what additional insights can we gain from a broader representation of the

119 genome than traditionally used in landscape genetics studies (typically microsatellites and

120 organelle DNA)? Taken together, the answers to these questions offer deeper knowledge of the

121 interaction between human activities, landscape features, and population structure of noxious

122 weeds and hence contribute to improve effective management and control of these damaging

123 plants. More generally, these answers contribute to deepen our understanding of the interaction

124 between geographic setting and population differentiation, adaptation, and persistence (Taylor *et*

125 *al.* 1993).

126

127 **MATERIALS AND METHODS**

128

129 **Study system**

130 *Ipomoea purpurea*, the common morning glory, is an agricultural weed well suited for the

131 study of human-aided migration on population connectivity because of its relatively well-

132 characterized biology, its recent introduction into the United States, and its close association with

133 agricultural crops. Its tight association with humans since its introduction allows for assessing the

134 impact of recent landscapes on structuring genetic variance under the initial simplifying

135 assumption of evolutionary equilibrium (Marko & Hart 2011). *Ipomoea purpurea* is a noxious

136 weed with a widespread distribution that includes highly heterogeneous landscapes in the

137 Eastern, South- and Mid-western regions of the United States (Culpepper 2006; Webster &

138 Nichols 2012). It is a self-compatible annual bumblebee-pollinated vine, with heavy seeds, and is

139   found primarily in agricultural fields and disturbed areas (Tiffin & Rausher 1999; Baucom &

140   Mauricio 2008). It is currently one of the most problematic weeds of agriculture (Webster &

141   Nichols 2012) and is capable of infestations leading to substantial decline in crop (although not

142   quantified for *I. purpurea*, related *Ipomoea* species may cause declines up to 80% of crop yield,

143   Rogers *et al.* 1996). This species exhibits in general low-level resistance to the commonly used

144   herbicide glyphosate (Baucom & Mauricio 2004, 2008), although the exact resistance level varies

145   widely among populations of this species (Kuester *et al.* 2015). This species is also a major

146   concern for conservation given its naturalization in multiple regions throughout the world and its

147   aggressiveness as an invasive (Chaney & Baucom 2012; Fang *et al.* 2013). Previous analyses of

148   its genetic diversity and differentiation using a panel of 15 nuclear microsatellite markers identify

149   limited population structure and a less-than definitive isolation-by-distance pattern (Kuester *et al.*

150   2015). It remains to be seen, however, the role that landscape features, both natural and human-

151   driven, play in structuring genetic variation within this species.

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, and planted crops) 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 first processed all these data into landscape layers

161   at a common spatial resolution of 10km^2 and a common spatial extent around the US states with

162   available samples (Fig. 1). These spatial resolution and extent were chosen to maintain a practical

163   balance between scale and analytical manageability given available computational resources. To

164   reduce dimensionality, we opted to perform two separate Principal Component Analyses (PCAs)

165   on the 19 climatic and 8 soil layers, respectively. For all subsequent analyses we kept the

166   resulting first principal component of each of these analyses, which accounted for over 60% of

167   the variance and primarily summarized temperature temporal gradients and soils' pH and

168   sandiness, respectively (Table S2).

169

170   With the objective of estimating the genetic connectivity of populations of *I. purpurea*, we

171   compiled genetic data on an extensive panel of 15 previously optimized microsatellite loci

172   (Molecular Ecology Resources Primer Development Consortium; *et al.* 2013). These data

173   encompass a total of 597 individuals from 31 localities (with a minimum of 8 individuals per

174   locality) (Fig. 1; Table S3). All individuals were collected in 2012 from farms across the range of

175   *I. purpurea* in the United Sates (for further details and basic genetic variability analyses see

176   Kuester *et al.* 2015). In addition, to obtain a more comprehensive representation of the genome of

177   *I. purpurea* and assess the robustness of results in light of coalescent and mutational variance

178   (Nielsen & Slatkin 2013), we generated a Next Generation Sequencing (NGS) dataset from a

179   subsample of individuals. In conjunction with the microsatellite data, this complementary dataset,

180   which offers a more widespread representation of this species' genome, is expected to lead to

181   more comprehensive population genetic inferences (Bohonak & Vandergast 2011; Epps &

182   Keyghobadi 2015).

183

184   To generate the NGS dataset, we constructed genome-wide Genotype By Sequencing (GBS)

185   libraries for 80 individuals sampled across 8 of localities (10 individuals per locality; Fig 1). The

186   GBS library was developed using 7ng of genomic DNA extracted from leaf or cotyledon tissue

187   using SNPsaurus' (Oregon, USA) nextRAD technology, which uses a selective PCR primer to

188   amplify consistent genomic loci among individuals. Similarly to RAD-Seq sequences (Rowe *et*

189   *al.* 2011) in which the DNA flanking a restriction enzyme cut site is selected for amplification,

190   nextRAD amplifies sequences that correspond to the DNA downstream of a short selective

191   priming site. Samples were first fragmented and then ligated to short adapter and barcode

192   sequences using a partial Nextera reaction (Illumina; California, USA) before being amplified

193   using Phusion® Hot Start Flex DNA Polymerase (New England Biolabs; Massachusetts, USA).

194   The 80 dual-barcoded PCR-amplified samples were pooled and the resulting libraries were

195   purified using AMPure XP beads (Agencourt Bioscience Corporation; Massachusetts, USA) at

196   0.7x. The purified library was then size selected to 350-800 base pairs and sequenced using two

197   runs of an Illumina NextSeq500 sequencer (Genomics Core Facility, University of Oregon).

198

199   The resulting sequences were analytically processed by combining the reads of 16 randomly

200   selected individuals (of the 80 sequenced) to create a pseudo-reference genome. This was done

201   after removing loci with read counts above 20,000, which presumably corresponded to repetitive

202   genomic material, and loci with read counts below 100, which presumably corresponded to off-

203   target or read errors. The filtered reads were aligned to each other using BBMap

204   (http://sourceforge.net/projects/bbmap/). All parameters were set to default values with the

205   exception of minimum sequence identity, which was set to 0.93 to identify alleles. A single read

206   instance was chosen to represent the locus in the pseudo-reference. This resulted in a total of

9

207 263,658 loci. All reads from each of the 80 individuals were then aligned to the pseudo-reference

208 using BBMap (http://sourceforge.net/projects/bbmap/) and converted to a vcf genotype table,

209 using Samtools (Li *et al.* 2009) and bcftools (Li 2011), after filtering out nucleotides with a

210 quality score of 10 or worst. The resulting vcf table was filtered using vcftools (Danecek *et al.*

211 2011) for SNPs with a minimum allele frequency of 0.02, a minimum read depth of 5, and a

212 maximum 15% of missing data. This resulted in 9774 variable regions. Loci were further filtered

213 using vcftools to exclude loci with less than 5 high quality base-calls and with more than 20%

214 missing data or an average of less than 20 high quality base calls. This resulted in a final panel of

215 8210 Single Nucleotide Polymorphisms (SNPs) that we used in all subsequent analyses.

216

217 **Population structure analyses**

218   We first conducted a series of preliminary analyses to characterize the overall genetic

219 structure of *I. purpurea*. All analyses were run separately for the microsatellite (SSR, hereafter)

220 and SNP datasets given their different sampling and geographic coverage (Fig. 1). First, we

221 examined population differentiation by estimating F_{ST} using GenAlex v6.5 (Peakall & Smouse

222 2012) (because similar global F_{ST} and R_{ST} estimates were obtained for the SSR dataset, we opted

223 to report F_{ST} values only to allow direct comparisons with the SNP dataset). We then estimated

224 contemporary effective population size for each sampled locality in NeEstimator v2 using the

225 excess heterozygous method (Do *et al.* 2014). We performed this latter analysis to assess the

226 possibility of whether effective rate of migration (*Nm*) may be asymmetric in response to

227 differences in population size (e.g., greater migration into satellite populations).

228

10

229   In addition, we assessed population admixture and spatial genetic clustering using 2 spatially

230   explicit Bayesian clustering algorithms: BAPS (Corander *et al.* 2004) and TESS (Chen *et al.*

231   2007). BAPS was run using its spatial clustering algorithm (Cheng *et al.* 2013) with 10 runs for

232   each K value tested, whereas TESS was run using the admixture algorithm and a BYM model

233   (i.e., using a spatial prior of admixture proportions; Durand *et al.* 2009b) with 10 runs per K

234   value, and without using geographic weights. The TESS model, with the lowest DIC was chosen

235   as the optimal model (Durand *et al.* 2009a). For both analyses K values ranging from 2 to the

236   maximum number of sampled locations (31 and 10 for the SSR and SNP datasets, respectively)

237   were tested. Results were generally consistent between BAPS and TESS analyses; thus, here we

238   report exclusively the TESS results. Additionally, following Wang et al. (2009), we

239   complemented these analyses with an Analysis of Molecular Variance (AMOVA; Excoffier *et al.*

240   1992) run in GenAlex (Peakall & Smouse 2012) using 9999 permutation replicates. This analysis

241   partitioned the variance into regions based on the spatial genetic clusters previously identified to

242   quantify the fraction of the genetic variance explained by spatial genetic clusters, and hence the

243   relative importance of spatial genetic structure in *I. purpurea*.

244

245   Finally, we investigated population connectivity by estimating levels of recent migration

246   between sampled localities through the identification of individuals of mixed ancestry using

247   BayesAss (Wilson & Rannala 2003). BayesAss is a program that uses individual multilocus

248   genotypes and a Markov Chain Monte Carlo (MCMC) algorithm to probabilistically distinguish

249   between immigrants and long-term native individuals (Wilson & Rannala 2003). It provides

250   posterior probability distributions of individual immigrant ancestries (i.e., the probability that an

251   individual is a first or second generation migrant from each of the populations in the sample). We

11

252   ran BayesAss for 6 million generations using default parameter settings, and discarded the first 2

253   million generations as burn-in (Dyer 2009). For each marker dataset, we repeated this analysis 3

254   times (for a total of 18 million generations) and combined the results from the 3 replicates for our

255   final inference. Then, using a posterior probability cut-off of 0.75 we assign individuals'

256   ancestry. It is important to note that because of computational limits we had to randomly

257   subsample our set of SNPs to 400 SNPs for this analysis. We complemented this analysis with a

258   spatial autocorrelation analysis run in GenAlex (Peakall & Smouse 2012) to evaluate the overall

259   spatial scale of genetic turnover.

260

261   **Landscape genetics analyses**

262   After assessing overall population structure of *I. purpurea*, we evaluated the association

263   between landscape features and genetic differentiation. To do this, we first converted our

264   pairwise F_{ST} estimates into conditional genetic distances (Dyer *et al.* 2010) using GeneticStudio

265   (Dyer 2009). Briefly, conditional distances are measures of pairwise genetic distance derived

266   from population networks, constructed based on the degree of genetic similarity between sampled

267   localities (Dyer & Nason 2004). Because these networks are pruned based on the principle of

268   conditional independence of the total among population genetic covariance (the specific method

269   of pruning used is edge deviance; Magwene 2001), conditional distances reflect genetic similarity

270   between localities that better capture direct gene flow as opposed to connectivity driven by

271   intervening localities (Dyer 2015b).

272

273   The climate, crops, elevation, landcover, population density, and soils landscape layers (Table

274   S1) were converted into landscape resistance layers by assigning a resistance value to each

12

275   landscape feature in these layers to reflect the difficulty that each feature offers to the movement

276   of gametes or individuals (i.e., pollen or seeds). It is important to note that in contrast to previous

277   studies that typically rely on expert opinion for resistance assignment, we utilized an unbiased

278   statistical optimization to avoid the sensitivity of results to subjective resistance assignment

279   (Spear *et al.* 2010). Specifically, resistance values were optimized through a genetic algorithm

280   approach, which is a heuristic stochastic optimization algorithm that emulates the process of

281   inheritance, mutation, and natural selection (Mitchell 1996). Briefly, in this search algorithm a

282   population of individuals with traits encoded by unique combinations of model parameters

283   (resistance assignment proposals in our case) is allowed to compete with each other based on the

284   fitness associated with the traits it carries (Peterman *et al.* 2014). Specifically, in Peterman's

285   (2014) implementation of this algorithm, which we followed here, individuals' fitness is

286   estimated by the relative quality of a MLPE.lmm model (Maximum Likelihood Population

287   Effects – Linear Mixture Model) that evaluates the association between pairwise genetic distance

288   and landscape cumulative resistance between localities, estimated in Circuitscape (Shah &

289   McRae 2008). Individuals with parameter settings (resistance assignments in our case) that result

290   in better models, as measured by a Deviance Information Criterion (DIC) score, are preferentially

291   represented in the following generation. Offspring modifications introduced by mutations (i.e.,

292   small resistance assignment perturbations) allow for exploration of the parameter space. The

293   algorithm is stopped once a large number of generations have passed without significant

294   improvement in fitness.

295

296   We implemented Peterman's (2014) algorithm in R (package ResistanceGA; Peterman 2014)

297   allowing for the independent optimization of each of our landscape layers. The optimal resistance

298   landscapes identified in this way were then used to run a final univariate MLPE.lmm model to

299   identify the association between landscape features and conditional genetic distances between

300   localities. In addition, we run separate simple and partial db-RDA (distance-based Redundancy

301   Analyses; Legendre & Anderson 1999)—the latter uses the geographic distance between

302   populations as a covariate to account for the concurrent increase in cumulative resistance with

303   geographic distance. We opted for db-RDA instead of the most commonly used partial Mantel

304   test given the statistical issues of the latter (Raufaste & Rousset 2001; Guillot & Rousset 2013).

305   We run this latter analysis using the package vegan (Oksanen *et al.* 2015) in R and assessed

306   significance with 9999 permutations. Finally, to identify the relative contribution of natural and

307   human-driven landscape features we ran a Multiple Regression on Distance Matrices (MRDM;

308   Legendre *et al.* 1994), which has been identified as one of the best performing methods for

309   evaluating the interplay between landscape features and genetic connectivity (Balkenhol *et al.*

310   2009). Before running these regressions, we standardized all optimized resistance layers to mean

311   of zero and variance of 1 (Dyer *et al.* 2010). These final regressions included geographic distance

312   as a predictor and were run in R (using package ecodist; Goslee & Urban 2007) using 10,000

313   permutations to assess significance. In none of our analyses did we implement a Bonferroni

314   correction for multiple testing because of the overly conservative nature of this correction

315   (Nakagawa 2004; Glickman *et al.* 2014). Instead we applied a false recovery rate correction

316   (Benjamini & Hochberg 1995) using the function *p.adjust* in R (R Core Development Team

317   2016).

318

319   **RESULTS**

320   **Population structure**

14

321   The set of preliminary genetic analyses indicated that *I. purpurea* sampled localities were

322   in no major violation of Hardy-Weinberg equilibrium, as judged by the small difference between

323   expected and observed heterozygosity (mean He = 0.294±0.014 and 0.250±0.001; mean Ho =

324   0.291±0.009 and 0.260±0.001, respectively for SSR and SNP datasets). Levels of expected and

325   observed heterozygosity for the SSR dataset were only slightly greater than those estimated for

326   the SNP dataset (Table S3). Likewise, the estimated mean effective population size per sampled

327   locality was only slightly greater and more variable for the SSR dataset than for the SNP dataset

328   (13.71±5.59, 9.49±0.13, respectively; Table S3), but in neither case was there salient evidence of

329   a plausible source-sink dynamic, as judged by the similar effective sizes among populations

330   (Table S3; Diffendorfer 1998). Accordingly, spatial structure in this species was limited, as

331   evidenced by the modest F_{ST} estimates (0.151 and 0.140, respectively for SSR and SNP datasets)

332   and the admixed genetic composition of individuals (rather more structured in the SNP dataset;

333   Fig. 2). Furthermore, although TESS results indicated the existence of partially discernable

334   spatial genetic clusters (4 and 5 clusters for the SSR and SNP datasets, respectively; Fig. 2),

335   variation among the inferred spatial genetic clusters explained less than 10% of the variance

336   (Table 1). It is especially noteworthy that the inferred spatial genetic clusters were constrained to

337   geographically contiguous areas in the SNP dataset, but not in the SSR dataset (Fig. 2).

338

339   Despite these overall similarities between datasets, estimates of recent ancestry differed

340   between them. The analysis on the SSR dataset indicated that migration among localities is

341   widespread and hardly geographically constrained, with only four localities being primarily

342   constituted of native individuals (Fig. 3a). Across localities, on average 73.65% of individuals

343   were inferred to be 1st or 2nd generation immigrants. On the other hand, the analysis on the SNP

344   dataset showed that most populations have a limited number of recent immigrants, but that the

345   relatively few inferred immigrants (on average 27.42% of individuals across localities) did not

346   come exclusively from geographically proximate localities (Fig. 3c). Nevertheless, long distance

347   migration events were estimated to be infrequent in this dataset relative to the SSR dataset, in

348   which migration across geographically spread states was common (e.g., between North Carolina

349   and Tennessee, Fig. 3a,c). Furthermore, of the 6 localities shared between datasets only 1 was

350   similarly inferred to comprise mostly native individuals by both sets of markers; inferences for

351   the other five localities differed between datasets. Accordingly, the SSR-based pruned

352   conditional genetic network (Dyer *et al.* 2010) was significantly more interconnected (vertex

353   connectivity = 5 ; White & Harary 2001) than the SNP-based network (vertex connectivity 0)

354   (Fig. 3b,d). This relatively low complexity of the SNP-based network further confirmed the

355   inference of locally constrained migration in this dataset compared to more widespread migration

356   inferred with the SSR dataset.

357

358   In line with these findings, the spatial autocorrelation analyses indicated that genetic

359   clustering extended over wider spatial scales for the SSR dataset than for the SNP dataset (Fig.

360   S1). Specifically, nearby localities showed a significant tendency to be genetically similar to each

361   other up to a considerably greater distance in the case of the SSR dataset relative to the SNP

362   dataset (Fig. 3). Furthermore, at greater distances no significant association was recovered for the

363   SNP dataset, further confirming their genetic isolation from each other, whereas a significant

364   negative association (i.e., statistical tendency to be more different from each other than expected

365   by chance) was recovered in the SSR dataset.

366

16

Landscape genetics

367

368   Unsurprisingly given the distinct geographical and environmental ranges covered by each

369   dataset (Fig. 1), the optimization of landscape resistance layers resulted in alternative optimal

370   solutions for the SSR and SNP datasets (Fig. S2). It is important to note, however, that a formal

371   comparison is not possible as the associations recovered are statistical associations driven by the

372   fit of the resistance parameterization to the data under the statistical model implemented

373   (Martínez-Abraín 2008). While these associations are expected to recapitulate real biological

374   properties of the study system, they are constrained to the data at hand. Nevertheless, association

375   patterns that are robust to the data used are expected to better reflect the actual impact that

376   landscape features have on gene flow, independently of possible biases introduced by expert

377   opinion. Therefore, we focus below on the common biological findings between marker types,

378   while also denoting the most relevant differences. Such differences likely reflect not only the

379   different environmental ranges covered by each dataset, but most importantly, the particular

380   genetic connectivity pattern captured by each one of them (Fig. 3).

381

382   In spite of the underlying differences in inferred population connectivity (Fig. 3), some of

383   the resulting landscape resistance layers roughly resemble their counterpart in the other dataset in

384   terms of their relative resistance allocation (Fig. S2). There were some landscape features that

385   showed consistent resistance patterns between datasets. For example, cotton-dominated areas and

386   woody savannas were recovered as highly permeable landscape features, whereas areas

387   dominated with soybean fields, evergreen forests, open shrublands, and grasslands were

388   recovered as highly resistant landscape features. This difference between the permeability of

389   different crop and landcover features matches the prevalence and distribution of these features

17

390   across the study area. Soybean dominated areas occupied over 8.7% of the study area and were

391   most prevalent in its northwestern portion, where *I. purpurea,* being primarily a subtropical vine

392   (Fang *et al.* 2013)*,* is less prevalent. On the other hand, cotton-dominated areas occupied 0.8% of

393   the study area and were concentrated in its southern portion, including areas of high *I. purpurea*

394   concentration. Likewise, woody savannas represented 10.8% of the study system and were

395   constrained to the southern portion of it, whereas evergreen forests, open shrublands, and

396   grasslands represent together 1.1% of the study system and are prevalent in areas where *I.*

397   *purpurea*, being mainly found on agricultural fields, gardens, and roadsides (Defelice 2001), is

398   absent. Hence, these landscape associations most likely reflected the environmental preferences

399   of *I. purpurea*, as environments where this species thrives coincided with environments with a

400   low resistance assigned, whereas environments unoccupied by this species were assigned greater

401   resistances.

402

403   Given the differences in sampling between datasets (Fig. 1), the most revealing patterns

404   were similar landscape variables identified as significant (or marginally significant) predictors of

405   genetic differentiation in *I. purpurea* across datasets (Table 2). Both datasets pointed towards

406   human-impacted landscapes playing an important direct role in shaping genetic connectivity in

407   this species. While in both sets of MLPE.lmm models, natural (climate, elevation, and soils) and

408   human-related landscapes (landcover and human population density) were identified as

409   significant or marginally significant predictors of genetic similarity between localities, the

410   variable with the greatest association coefficient and lowest AICc value in these models was in

411   both cases a variable closely linked to human presence (landcover in the SSR dataset, and human

412   population density in the SNP dataset; Table 2). In contrast, the dbRDA analyses showed that

18

413   none of these features were consistently recovered as a significant predictor of genetic similarity

414   in the SNP dataset, especially after accounting for geographic distance between populations—

415   even though geographic distance by itself was not a significant predictor of genetic structure in

416   this species (i.e., no isolation by distance pattern was recovered). Yet, when considering all

417   variables together in a multivariate manner while accounting for geographic distance, human

418   population density was the only variable that remained a significant or marginally significant

419   predictor of population structure across both SSR and SNP datasets (Table 2). Specifically, these

420   results indicated that, after considering other landscape variables, human-population-density

421   resistance was associated with population differentiation across datasets, with high and low

422   populated areas identified as less favorable areas for migration (although conductance optimum

423   for the SNP dataset was notably lower than for the SSR datasets; Fig. 4). Accordingly, the

424   corresponding resistance-based conductance map (Shah & McRae 2008), showed local clusters

425   of high permeability for both datasets (although slightly less constrained in the SSR dataset),

426   surrounded by vast areas of high resistance (Fig. S3). It is important to note, however, that these

427   multivariate regressions explained a small, non-significant proportion of the variance (MRDM R^2

428   for SSR and SNP dataset were 0.055 (F = 3.776, p-val. = 0.110) and 0.309 (F = 1.275, p-val. =

429   0.261), respectively).

430

431   In addition to population density, climate was also recovered as a significant predictor of

432   genetic dissimilarity across analyses on the SSR dataset but not for the SNP dataset (Table 3) —

433   most likely because of the lack of northern samples, from colder areas, for the latter dataset (Fig.

434   1). Specifically, the first component of the climate PCA (see methods above), which primarily

435   summarized temperature variation, was consistently associated with genetic differentiation (Table

436   3), with intermediates values identified as a barrier to dispersal. Warmer areas were identified as

437   the most conducive to dispersal, followed by colder areas—the latter pattern exclusively driven

438   by connectivity between the northernmost localities (Fig. 4). The corresponding resistance-based

439   conductance map showed a less spatially constrained distribution of permeable areas with

440   connections between southern and between northern localities, and a strong barrier between them

441   (Fig. S3).

442

443

444   **DISCUSSION**

445   Taken together, the population structure analyses indicated limited to moderately high

446   (depending on the molecular marker) global levels of genetic differentiation in *I. purpurea* and

447   non-geographically structured migration. This non-geographically structured migration was

448   inferred to be rather rare when using the SNP dataset (Fig. 3). These results suggest that broadly

449   distributed populations of this agricultural weed are generally genetically distinct, but there is

450   some indication of long-distance and putatively human-mediated migration between localities, as

451   suggested by the recovered association between human population density and genetic similarity.

452   Such levels of differentiation and long-distance migration strongly contrast with this species'

453   patchy distribution, which is tightly linked to isolated agricultural patches that are surrounded by

454   a complex matrix of natural and urbanized areas. As suggested by the landscape genetics

455   analyses, this matrix does seem to impact connectivity in this species. Specifically, climate and

456   human population density were robustly recovered as predictors of genetic connectivity in this

457   species. Of these landscape variables, climate has a stronger effect, as judged by its greater

458   MRDM coefficient, but only when considering the SSR dataset, which covers the northern

459   portion of the range (Table 3). Otherwise, population density was the only variable across

460   datasets with a marginally significant effect—even after accounting for multiple tests. Along with

461   the recovered pattern of landscape features of high and low permeability, in particular in regards

462   to crop types (Fig. S2), this finding highlights the important role that humans play in structuring

463   populations of this species. In addition, the results highlight how inferences about population

464   structure and patterns of connectivity are dataset-dependent, with marked differences becoming

465   apparent only after careful dissection of roughly similar F_{ST} and heterozygosity estimates across

466   molecular markers. Below, we detail each of these novel findings and place them in the context

467   of agricultural weed movement across the landscape, invasive species, and landscape genetics

468   practice.

469

470   **Human impact**

471

472   Given that *I. pupurea* is a naturalized species in the United States that is found primarily

473   associated with cultivated crops and horticultural gardens (Defelice 2001; Baucom & Mauricio

474   2004; Fang *et al.* 2013), the finding that human population density is a predictor of genetic

475   similarity in this species is at first glance intuitive. Yet, because habitat requirements for

476   establishment and migration are not always the same, especially for organisms with distinct

477   migration stages (e.g., pollen or seeds in plants) and dormant stages (Murphy & Lovett-Doust

478   2004), this finding is not as straightforward as it seems. In particular, the fact that human

479   population density is an informative predictor throughout the entire sampled distribution, whereas

480   climate seems to mainly represent a gene flow barrier between southern and northern *I. purpurea*

481   sites (Fig. S1) highlights the influence that humans have on structuring the populations of this

21

482   species and helps to discern the factors involved in the spread of this noxious weed. In this sense,

483   the results point towards humans not only as likely responsible for the introduction of this weed

484   into the United States (Fang *et al.* 2013), but also as responsible for facilitating its current spatial

485   connectivity, and hence its opportunities for thriving in the complex landscapes it inhabits. As

486   evidenced by the inferred landscape conductivity, human population density seems to primarily

487   facilitate connectivity at local to intermediate scales (note the clusters of high conductivity

488   around sampled localities; Figs. 4 and S2), which encompass proximate agricultural fields,

489   suggesting that factors such as sharing of contaminated agricultural machines, trade between

490   nearby farmers, or local distribution of contaminated crop seeds are at play (Dastgheib 1989;

491   Thill & Mallory-Smith 1997; Benvenuti 2007; Boyd & White 2009). At the same time,

492   considering that i) the horticultural trade has been recognized as the main source of invasive

493   introductions and spread in the United States (Lehan *et al.* 2013), ii) that *I. purpurea* is an

494   appreciated horticultural species (Fang *et al.* 2013), and that, given current agricultural practices,

495   crop seed contamination is unlikely to be a major factor (Economic Research Service 1998), it is

496   probable that ornamentals' trade between population centers may help explain both the long

497   distance dispersal events recovered in both datasets (Fig. 3) and the local clustering (Fig. 4).

498   Alternatively, the impact of human populations on the distribution and abundance of bumblebees

499   (Martins *et al.* 2013; Jha 2015), which are *I. purpurea*'s predominant pollinators (Ennos 1981;

500   Baucom & Mauricio 2008), could also be partially responsible for the connectivity patterns

501   recovered as changes in the pollinators community would have strong effects on gene flow (Jha

502   & Kremen 2013). In reality a combination of all these factors may be involved.

503

504   While further analyses are needed to elucidate the ultimate causes behind the recovered

505   association between human population density and genetic dissimilarity in *I. purpurea*, our

506   findings bring much needed information to limit the spread of this noxious weed. Our findings

507   are not only relevant to *I. purpurea* and to the evolution of herbicide resistance in this species

508   (i.e., is herbicide resistance evolving independently across populations or is it being disseminated

509   through human-aided migration?), but also has important implications for other weeds of

510   agricultural concern as well as other human-exploiter species (Blair 2001), such as other

511   invasives. Specifically, in line with previous work (Bataille *et al.* 2011; Auffret *et al.* 2014;

512   Banks *et al.* 2015), the results here point towards the need of better strategies to minimize the

513   impact that humans have on the spread of species. In particular, our results further support that

514   humans may not only facilitate the introduction of invasive species into non-colonized areas, but

515   also contribute to the maintenance of gene flow among naturalized populations (Medley *et al.*

516   2015), which may be critical in providing relevant genetic variants for increased fitness as well as

517   prevent inbreeding depression in these newly colonized areas (Kolbe *et al.* 2004; Edelaar &

518   Bolnick 2012).

519

520   **Landscape genetics practice**

521   Under the common landscape genetics' assumption of an equilibrium between migration

522   and genetic drift (Marko & Hart 2011; Dyer 2015b), our results offer a valuable window into the

523   role that environmental setting plays in structuring genetic diversity. Our analyses take advantage

524   of recent methodological developments i) that surpass the need of arbitrary landscape resistance

525   assignment that make inferences sensitive to subjectivity of expert opinion (Dyer 2015a), ii) that

526   account for the indirect genetic similarity of populations (Dyer & Nason 2004), and iii) that use

527   rigorous statistical inferences (Balkenhol *et al.* 2009; Peterman *et al.* 2014). Furthermore, in

528   contrast to the common practice in the field of using one or a few loci that are either uni-

529   parentally inherited (although a few exceptions exist; e.g. Perry *et al.* 2013), which prevents an

530   assessment of common patterns across the genome (Bohonak & Vandergast 2011), our inferences

531   are derived from common findings among two rather different sets of molecular markers. In

532   doing so, we provide not only a better representation of the genome, and hence, less sensitive

533   inferences to i) ascertainment bias (Brandström & Ellegren 2008), ii) molecular markers'

534   idiosyncrasies (Buschiazzo & Gemmell 2006), and iii) coalescent and mutational variance

535   (Nielsen & Slatkin 2013; Steiner *et al.* 2013), but also the ability to distinguish differences in the

536   underlying population dynamics. Such differences have strikingly important implications. For

537   example, when evaluating plausible approaches to the threat of an invasive species such as *I.*

538   *purpurea*, recommendations would be quite different depending on whether gene flow is believed

539   to be relatively widespread (as inferred by the SSR dataset) or whether it is believed to be

540   minimal (as inferred by the SNP dataset). In this example, it is clear that evidence-based

541   conservation would clearly benefit from recognizing the current uncertainty in regards to the

542   exact population connectivity as opposed to automatically relying on a single-marker story.

543   Given recent advances in next generation sequencing, it seems straightforward to focus on

544   landscape genomics instead of few loci. Hence, development of methods for explicitly integrating

545   inferences from multiple genome regions, as it is customary in population genetics, would be of

546   great value.

547

548   Even more important, the differences identified between markers offer the opportunity to

549   explore the underlying causes for such differences and hence a more in-depth understanding of

24

550   the landscape influences on species' genetic structure. While it is theoretically possible that the

551   differences in inference between the two markers are exclusively driven by the particular

552   geographic sampling of each dataset, the robust differences that we report between the localities

553   and regions common to both datasets renders this possibility unlikely. A second alternative is that

554   the differences are driven by the different mutation rates underlying the two type of markers

555   (Wang 2010, but see Bohonak and Vandergast 2011). This time-differential hypothesis is based

556   on the argument that SSR mutation rates are typically estimated to be in the $10e^{-3}$-$10e^{-4}$ range,

557   whereas SNP mutation rates, although harder to estimate given their occurrence in multiple

558   heterogeneous genomic regions (Lercher & Hurst 2002), are typically presumed to be on average

559   much lower. Nonetheless, given mutational rates are still on the scale of thousand of years and

560   up, most genetic variation seen in current populations would precede the temporal window of the

561   majority of landscape genetics studies. It is then the sorting of standing genetic variation, rather

562   than the mutation rate, what would primarily drive genetic dissimilarity between current

563   subpopulations of a species. Such sorting is thus expected to be specific to different genomic

564   regions. That is, because of the genome-wide coverage of SNPs and the independent evolutionary

565   trajectories of SSR and SNP loci, recombination, effective population sizes, and other relevant

566   evolutionary parameters are expected to be highly heterogeneous across loci, which can play an

567   important role in the pace at which each loci segregates, making it unclear which temporal scale

568   would be reflected by each dataset (Ennos 1994; Bohonak & Vandergast 2011). Of all the factors

569   involved, effective population sizes of each locus, which directly affect the rate of genetic drift of

570   each loci (Storz *et al.* 2001; Piganeau & Eyre-Walker 2009), is probably the most impacted by

571   landscape conditions, and hence it is probably a major determinant of patterns of shared variants

25

572   across the landscape. Yet, traditionally estimates of effective population size are rarely

573   incorporated into landscape genetics studies.

574

575   In reality, both datasets likely encompass a range of temporal scales, and presumably reflect

576   both historical and contemporary processes. In this regard, working with an agricultural weed

577   with a relatively well-known history offers the advantage of better accounting for the historical

578   component of current genetic variation. Specifically, under the time-differential hypothesis

579   (Wang 2010; but see Bohonak & Vandergast 2011), it is plausible that the spatial structure

580   recovered by our SNP dataset is strongly influenced by ancestral subdivision and that the relative

581   lack of structure recovered by our SSR dataset reflect more recent gene flow. However,

582   considering the current understanding of the introduction of *I. purpurea* into the United Sates,

583   this possibility seems unlikely. If this weed was indeed introduced through horticulture from a

584   European bottlenecked population during the European colonization of North America (Fang *et*

585   *al.* 2013), ancestral structure would be unlikely. Likewise, if the species expanded from the point

586   of introduction, a nested pattern of genetic similarity driven by serial founder events would be

587   expected (Ramachandran *et al.* 2005; Slatkin & Excoffier 2012)—not a rather distinct clustering

588   of individual subpopulations as we recovered (Fig. 3). Instead, it seems more likely that the

589   obvious structure recovered in the SNP dataset reflects the isolation driven by the complex

590   agricultural matrix *I. purpurea* inhabits and the role that human intervention has had on its

591   evolutionary trajectory (e.g., by facilitating long distance dispersal events; Fig. 3). Yet, further

592   analyses (see below) are needed to test this hypothesis.

593

26

594 More importantly, the incorporation of both datasets into our analyses highlights a commonly

595 overlooked issue in landscape genetics. That is, any pattern of genetic variation cannot be

596 understood without explicit consideration of species' demographic history as current genetic

597 variation is the result of multiple processes thorough the history of a species (Marko & Hart

598 2011), and different genomic regions reflect different coalescent histories (Nielsen & Slatkin

599 2013). Hence, the importance of incorporating coalescent-based simulations that explicitly model

600 the presumed landscape effects on genetic population structure and that take into account its

601 demographic history (Balkenhol & Landguth 2011; Hoban *et al.* 2012). Advances in this area are

602 already being developed with promising perspectives (He *et al.* 2013; Alvarado-Serrano &

603 Knowles 2014). In this regard, traditional landscape genetics results can be interpreted as a

604 necessary first step towards a more comprehensive understanding of the interaction between

605 landscapes and species evolutionary trajectories. By offering a working hypothesis of the effect

606 of current landscapes on genetic differentiation, traditional landscape genetics results serve the

607 purpose of identifying relevant hypotheses for further testing (Dyer 2015a) and pave the way for

608 rigorous simulation-based assessments of the role of landscape features in promoting or deterring

609 population differentiation.

610

611 **ACKNOWLEDGEMENTS**

615

616 **DATA ACCESSIBILITY**

27

617   All data generated is in the process of being archived in Dryad. The corresponding doi would be

618   made available upon acceptance.

619

620   **AUTHOR CONTRIBUTIONS**

621   D.F.A.-S. and R.S.B conceived the study. D.F.A.-S. and M.V.E. generated and compiled the

622   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.

623   wrote the manuscript. All authors read and approved the final submission.

LITERATURE CITED

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

Auffret AG, Berg J, Cousins SAO (2014) The geography of human-mediated dispersal. *Diversity and Distributions*, **20**, 1450–1456.

Auffret AG, Cousins SAO (2013) Humans as Long-Distance Dispersers of Rural Plant Communities. *PLoS ONE*, **8**.

Balkenhol N, Landguth EL (2011) Simulation modelling in landscape genetics: on the need to go further. *Molecular ecology*, **20**, 667–70.

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

Banks NC, Paini DR, Bayliss KL, 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.

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

Bataille A, Cunningham AA, Cruz M, Cedeño V, Goodman SJ (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 RS, 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 RS, Mauricio R (2008) The evolution of novel herbicide tolerance in a noxious weed:

29

648   The geographic mosaic of selection. *Evolutionary Ecology*, **22**, 85–101.

649   Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate [mdash] a practical and

650   powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

651   Benvenuti S (2007) Weed seed movement and dispersal strategies in the agricultural

652   environment. *Weed Biology and Management*, **7**, 141–157.

653   Berthouly-Salazar C, Hui C, Blackburn TM *et al.* (2013) Long-distance dispersal maximizes

654   evolutionary potential during rapid geographic range expansion. *Molecular Ecology*, **22**,

655   5793–5804.

656   Blair RB (2001) Birds and butterflies along urban gradients in two ecoregions of the US. In:

657   *Biotic homogenization* (eds Lockwood JL, McKinney ML), pp. 33–56.

658   Bohonak AJ, Vandergast AG (2011) The value of DNA sequence data for studying landscape

659   genetics. *Molecular ecology*, **20**, 2477–9; authors reply 2480–2.

660   Boyd NS, White S (2009) Impact of Wild Blueberry Harvesters on Weed Seed Dispersal within

661   and between Fields. *Weed Science*, **57**, 541–546.

662   Brandström M, Ellegren H (2008) Genome-wide analysis of microsatellite polymorphism in

663   chicken circumventing the ascertainment bias Genome-wide analysis of microsatellite

664   polymorphism in chicken circumventing the ascertainment bias. *Genome Research*, 881–

665   887.

666   Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic

667   genomes. *BioEssays*, **28**, 1040–1050.

668   Chaney L, Baucom RS (2012) The evolutionary potential of Baker's weediness traits in the

669   common morning glory, Ipomoea purpurea (Convolvulaceae). *American Journal of Botany*,

670   **99**, 1524–1530.

671   Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining

672   spatial population structure: A new computer program and a comparison study. *Molecular*

673   *Ecology Notes*, **7**, 747–756.

674   Cheng L, Connor TR, Sir??n J, Aanensen DM, Corander J (2013) Hierarchical and spatially

675   explicit clustering of DNA sequences with BAPS software. *Molecular Biology and*

676   *Evolution*, **30**, 1224–1228.

677   Clobert J, Baguette M, Benton TG, Bullock JM, Ducatez S (2012) *Dispersal Ecology and*

678   *Evolution*. Oxford University Press, Oxford.

679   Consortium; MERPD, Aksoy S, Almeida-Val VMF *et al.* (2013) Permanent Genetic Resources

680   added to Molecular Ecology Resources Database 1 October 2012–30 November 2012.

681   *Molecular Ecology Resources*, **13**, 341–343.

682   Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for

683   the analysis of genetic population structure. *Bioinformatics (Oxford, England)*, **20**, 2363–9.

684   Culley TM, Sbita SJ, Wick A (2007) Population genetic effects of urban habitat fragmentation in

685   the perennial herb Viola pubescens (Violaceae) using ISSR markers. *Annals of Botany*, **100**,

686   91–100.

687   Culpepper AS (2006) Glyphosate-Induced Weed Shifts. *Weed Technology*, **20**, 277–281.

688   Cushman S a, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes:

689   testing multiple hypotheses with causal modeling. *The American naturalist*, **168**, 486–99.

690   Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools.

691   *Bioinformatics*, **27**, 2156–2158.

692   Dastgheib F (1989) Relative importance of crop seed, manure and irrigation water as sources of

693   weed infestation. *Weed Research*, **29**, 113–116.

31

694   Defelice MS (2001) Intriguing World of Weeds Tall Morningglory , Ipomoea purpurea (L .)

695   Roth-Flower or Foe ? *Weed Technology*, **15**, 601–606.

696   Diffendorfer JE (1998) Testing Models of Source-Sink Dynamics and Balanced Dispersal. *Oikos*,

697   **81**, 417–433.

698   Diniz-Filho JAF, Nabout JC, Bini LM *et al.* (2009) Niche modelling and landscape genetics of

699   Caryocar brasiliense ("Pequi" tree: Caryocaraceae) in Brazilian Cerrado: an integrative

700   approach for evaluating central–peripheral population patterns. *Tree Genetics & Genomes*,

701   **5**, 617–627.

702   Do C, Waples RS, Peel D *et al.* (2014) NeEstimator v2: Re-implementation of software for the

703   estimation of contemporary effective population size (Ne) from genetic data. *Molecular*

704   *Ecology Resources*, **14**, 209–214.

705   Durand E, Chen C, Francois O (2009a) Tess version 2.3 - Reference Manual August 2009 ∗.

706   *Available at: memberstimc. imag. fr/Olivier. Francois/tess. html*, 1–30.

707   Durand E, Jay F, Gaggiotti OE, François O (2009b) Spatial inference of admixture proportions

708   and secondary contact zones. *Molecular biology and evolution*, **26**, 1963–73.

709   Dyer RJ (2009) GeneticStudio: a suite of programs for spatial analysis of genetic-marker data.

710   *Molecular ecology resources*, **9**, 110–3.

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

712   Dyer RJ (2015b) Population Graphs and Landscape Genetics. *Annual Review of Ecology,*

713   *Evolution, and Systematics*, **46**, annurev–ecolsys–112414–054150.

714   Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure.

715   *Molecular ecology*, **13**, 1713–27.

716   Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power

717   using conditional genetic distance derived from the topology of population networks.

718   *Molecular ecology*, **19**, 3746–59.

719   Economic Research Service U (1998) Seed Industry Structure Is Characterized by Growth and

720   Consolidation. *AIB*, **786**, 25–29.

721   Edelaar P, Bolnick DI (2012) Non-random gene flow: An underappreciated force in evolution

722   and ecology. *Trends in Ecology and Evolution*, **27**, 659–665.

723   Ennos RA (1981) Quantitative studies of the mating system in two sympatric species

724   ofIpomoea(Convolvulaceae). *Genetica*, **57**, 93–98.

725   Ennos R a (1994) Estimating the relative rates of pollen and seed migration among plant

726   populations. *Heredity*, **72**, 250–259.

727   Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling

728   historical and contemporary influences and inferring change. *Molecular Ecology*, **24**, 6021–

729   6040.

730   Everman E, Klawinski P (2013) Human-facilitated jump dispersal of a non-native frog species on

731   Hawai'i Island. *Journal of Biogeography*, **40**, 1961–1970.

732   Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From

733   Metric Distances Among DNA Haplotypes: Application. , **491**, 479–491.

734   Fang Z, Gonzales AM, Durbin ML *et al.* (2013) Tracing the geographic origins of weedy

735   Ipomoea purpurea in the Southeastern United States. *Journal of Heredity*, **104**, 666–677.

736   Fountain T, Duvaux L, Horsburgh G, Reinhardt K, Butlin RK (2014) Human-facilitated

737   metapopulation dynamics in an emerging pest species, Cimex lectularius. *Molecular*

738   *Ecology*, **23**, 1071–1084.

739   Glickman ME, Rao SR, Schultz MR (2014) False discovery rate control is a recommended

740   alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical*

741   *Epidemiology*, **67**, 850–857.

742   Goslee SC, Urban DL (2007) The ecodist Package for Dissimilarity-based Analysis of Ecological

743   Data. *Journal Of Statistical Software*, **22**, 1–19.

744   Guillot G, Rousset F (2013) Dismantling the Mantel tests (L Harmon, Ed,). *Methods in Ecology*

745   *and Evolution*, **4**, 336–344.

746   Harris SE, Xue AT, Alvarado-serrano D *et al.* (2016) Urbanization shapes the demographic

747   history of a native rodent (the white- footed mouse , Peromyscus leucopus) in New York

748   City.

749   He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to

750   the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.

751   Hoban S, Bertorelle G, Gaggiotti OE (2012) Computer simulations: tools for population and

752   evolutionary genetics. *Nature Reviews Genetics*.

753   Hodkinson DJ, Thompson K, Journal T, Dec N (2007) Plant Dispersal : The Role of Man. , **34**,

754   1484–1496.

755   Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in*

756   *Plant Science*, **15**, 675–683.

757   Jha S (2015) Contemporary human-altered landscapes and oceanic barriers reduce bumble bee

758   gene flow. *Molecular Ecology*, **24**, 993–1006.

759   Jha S, Kremen C (2013) Urban land use limits regional bumble bee gene flow. *Molecular*

760   *Ecology*, **22**, 2483–2495.

761   Kolbe JJ, Glor RE, Rodríguez Schettino L *et al.* (2004) Genetic variation increases during

762   biological invasion by a Cuban lizard. *Nature*, **431**, 177–81.

763   Kuester A, Chang S-M, Baucom RS (2015) The geographic mosaic of herbicide resistance

764   evolution in the common morning glory, *Ipomoea purpurea* : Evidence for resistance

765   hotspots and low genetic differentiation across the landscape. *Evolutionary Applications*,

766   n/a–n/a.

767   Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: Testing multispecies

768   responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.

769   Legendre P, Lapointe FJ, Casgrain P (1994) Modeling Brain Evolution from Behavior - a

770   Permutational Regression Approach. *Evolution*, **48**, 1487–1499.

771   Lehan NE, Murphy JR, Thorburn LP, Bradley BA (2013) Accidental introductions are an

772   important source of invasive plants in the continental United States. *American Journal of*

773   *Botany*, **100**, 1287–1293.

774   Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology and*

775   *Evolution*, **17**, 183–189.

776   Lercher MJ, Hurst LD (2002) Human SNP variability and mutation rate are higher in regions of

777   high recombination. *Trends in Genetics*, **18**, 337–340.

778   Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139–220.

779   Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping

780   and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**,

781   2987–2993.

782   Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and

783   SAMtools. *Bioinformatics*, **25**, 2078–2079.

784   Mack RN, Lonsdale WM (2001) Humans as global plant dispersers: Getting more than we

785   bargained for. *BioScience*, **51**, 95.

786   Magwene PM (2001) New Tools for Studying Integration and Modularity. *Evolution*, **55**, 1734–

787   1745.

788   Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*,

789   **28**, 614–621.

790   Margosian ML, Garrett K a., Hutchinson JMS, With K a. (2009) Connectivity of the American

791   Agricultural Landscape: Assessing the National Risk of Crop Pest and Disease Spread.

792   *BioScience*, **59**, 141–151.

793   Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in*

794   *Ecology and Evolution*, **26**, 448–456.

795   Martínez-Abraín A (2008) Statistical significance and biological relevance: A call for a more

796   cautious interpretation of results in ecology. *Acta Oecologica*, **34**, 9–11.

797   Martins AC, Goncalves RB, Melo G a R (2013) Changes in wild bee fauna of a grassland in

798   Brazil reveal negative effects associated with growing urbanization during the last 40 years.

799   *Zoologia*, **30**, 157–176.

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

801   McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations.

802   *Proceedings of the National Academy of Sciences of the United States of America*, **104**,

803   19885–90.

804   Medley KA, Jenkins DG, Hoffman EA (2015) Human-aided and natural dispersal drive gene

805   flow across the range of an invasive mosquito. *Molecular Ecology*, **24**, 284–295.

806   Meehan TD, Werling BP, Landis D a, Gratton C (2011) Agricultural landscape simplification and

807   insecticide use in the Midwestern United States. *Proceedings of the National Academy of*

808   *Sciences of the United States of America*, **108**, 11500–11505.

809   Menchari Y, Délye C, Le Corre V (2007) Genetic variation and population structure in black-

810   grass (Alopecurus myosuroides Huds.), a successful, herbicide-resistant, annual grass weed

811   of winter cereal fields. *Molecular Ecology*, **16**, 3161–3172.

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

813   Murphy HT, Lovett-Doust J (2004) Context and connectivity in plant metapopulations and

814   landscape mosaics: Does the matrix matter? *Oikos*, **105**, 3–14.

815   Nakagawa S (2004) A farewell to Bonferroni: The problems of low statistical power and

816   publication bias. *Behavioral Ecology*, **15**, 1044–1045.

817   Nielsen R, Slatkin M (2013) *An introduction to population genetics : theory and applications*.

818   Sinauer Associates, Sunderland Mass.

819   Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick

820   ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**,

821   316–36.

822   Oksanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: community ecology package. ,

823   http://CRAN.R–project.org/package=vegan.

824   Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic

825   software for teaching and research--an update. *Bioinformatics (Oxford, England)*, **28**, 2537–

826   9.

827   Perry GH, Louis EE, Ratan A *et al.* (2013) Aye-aye population genomic analyses highlight an

828   important center of endemism in northern Madagascar. *Proceedings of the National*

829   *Academy of Sciences of the United States of America*, **110**, 5823–5828.

830   Peterman WE (2014) ResistanceGA: An R package for the optimization of resistance surfaces

831   using genetic algortihms. *bioRxiv*.

832   Peterman WE, Connette GM, Semlitsch RD, Eggert LS (2014) Ecological resistance surfaces

833   predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular*

834   *Ecology*, **23**, 2402–2413.

835   Piganeau G, Eyre-Walker A (2009) Evidence for variation in the effective population size of

836   animal mitochondrial DNA. *PLoS ONE*, **4**, 2–9.

837   Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs

838   associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–

839   288.

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

841   3.2.4. , **0**, [http://www.r–project.org].

842   Ramachandran S, Deshpande O, Roseman CC *et al.* (2005) Support from the relationship of

843   genetic and geographic distance in human populations for a serial founder effect originating

844   in Africa. *Proceedings of the National Academy of Sciences of the United States of America*,

845   **102**, 15942–7.

846   Raufaste N, Rousset F (2001) Are Partial Mantel Tests Adequate? *Evolution*, **55**, 1703–1705.

847   Ricciardi A (2007) Are modern biological invasions an unprecedented form of global change?

848   *Conservation Biology*, **21**, 329–336.

849   Rogers JOYB, Murray DONS, Verhalen LM, Claypool PL (1996) Ivyleaf Morningglory (

850   Ipomoea hederacea) Interference with Cotton (Gossypium hirsutum)'. , **10**, 107–114.

851   Rowe HC, Renaut S, Guggisberg a (2011) RAD in the realm of next-generation sequencing

852   technologies. *Molecular ecology*, **20**, 3499–502.

853   Sexton JP, Hangartner SB, Hoffmann A a (2014) Genetic isolation by environment or distance:

854 which pattern of gene flow is most common? *Evolution; international journal of organic*

855 *evolution*, **68**, 1–15.

856 Shah VB, McRae BH (2008) Circuitscape : A Tool for Landscape Ecology. *Proceedings of the*

857 *7th Python in Science Conference*, 62–65.

858 Simberloff D (2013) *Invasive species : what everyone needs to know*.

859 Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and*

860 *Systematicsystematics*, **16**, 393–430.

861 Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of

862 genetic drift. *Genetics*, **191**, 171–181.

863 Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for

864 landscape genetic studies: Considerations for parameterization and analysis. *Molecular*

865 *Ecology*, **19**, 3576–3591.

866 Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation genomics of threatened

867 animal species. *Annual Review of Animal Biosciences*, **1**, 261–281.

868 Storz JF, Ramakrishnan U, Alberts SC (2001) Determinants of effective population size for loci

869 with different modes of inheritance. *The Journal of heredity*, **92**, 497–502.

870 Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity Is a Vital Element of Landscape

871 Structure. *Nordic Society Oikos* , **68**, 571–573.

872 Thill DC, Mallory-Smith CA (1997) The Nature and Consequence of Weed Spread in Cropping

873 Systems. , **45**, 337–342.

874 Tiffin P, Rausher M (1999) Genetic Constraints and Selection Acting on Tolerance to Herbivory

875 in the Common Morning Glory Ipomoea purpurea. *The American naturalist*, **154**, 700–716.

876 Trakhtenbrot A, Nathan R, Perry G, Richardson DM (2005) The importance of long-distance

877   dispersal in biodiversity conservation. *Diversity and Distributions*, **11**, 173–181.

878   Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and

879   phylogeography. *Molecular ecology*, **19**, 2605–8.

880   Wang IJ, Bradburd GS (2014) Isolation by Environment. *Molecular Ecology*, n/a–n/a.

881   Wang IJ, Savage WK, Bradley Shaffer H (2009) Landscape genetics and least-cost path analysis

882   reveal unexpected dispersal routes in the California tiger salamander (Ambystoma

883   californiense). *Molecular Ecology*, **18**, 1365–1374.

884   Webster TM, Nichols RL (2012) Changes in the Prevalence of Weed Species in the Major

885   Agronomic Crops of the Southern United States: 1994/1995 to 2008/2009. *Weed Science*,

886   **60**, 145–157.

887   White DR, Harary F (2001) The cohesiveness of blocks in social networks: node connectivity

888   and conditional density. *Sociological Methodology*, **31**, 305–359.

889   Wichmann MC, Alexander MJ, Soons MB *et al.* (2009) Human-mediated dispersal of seeds over

890   long distances. *Proceedings. Biological sciences / The Royal Society*, **276**, 523–532.

891   Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus

892   genotypes. *Genetics*, **163**, 1177–1191.

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

894

895

896   **FIGURES AND TABLES - Landscape connectivity of a noxious weed**
897
898   Table 1. Analysis of Molecular Variance (AMOVA) of SSR and SNP data. The contribution of
899   spatial clusters (region), individuals.
900

Effect	F-statistic	Variance explained		F-value		P-value	
		SSR	SNP	SSR	SNP	SSR	SNP
Regions	F_{RT}	3.94 %	8.51%	0.039	0.085	0.001	0.001
Localities (Regions)	F_{SR}	9.05 %	6.10 %	0.094	0.067	0.001	0.001
Individuals (Localities)	F_{ST}	0.67 %	24.85 %	0.130	0.146	0.001	0.001
Individuals	F_{IS}	86.33%	60.54 %	0.008	0.291	0.252	0.001
Total	F_{IT}	100%	100 %	0.137	0.395	0.001	0.001

901
902
903

904 Table 2. Summary of landscape genetics models. Model coefficients are reported followed
905 by associated p-value (in parenthesis) and, for MLPE.lmm models, followed by AIC
906 difference and ranking (in square brackets). Because of multiple testing , p-values reported
907 for MLPE.lmm and dbRDA analyses are adjusted using a false recovery rate correction.
908 Significant coefficients after a false-recovery-rate correction are in bold; marginally
909 significant ones are denoted with an asterisk.
910

| Variable	MLPE.lmm		dbRDA		partial-dbRDA		MRDM	
	SSR	SNP	SSR	SNP	SSR	SNP	SSR	SNP
Geographic distance	-	-	1.234 (0.173)	2.503 (0.202)	-	-	0.006 (0.969)	0.321 (0.542)
Environmental layers								
Elevation	**0.244** (0.028) +10.742 [3]	0.840* (0.054) +1.423 [4]	1.236 (0.104)	1.369 (0.348)	1.195 (0.206)	1.256 (0.438)	-0.629 (0.303)	-1.778 (0.685)
PC1 – climate	**0.242** (0.034) +10.356 [2]	**0.967** (0.050) +0.032 [2]	**1.454** (0.009)	1.513 (0.302)	**1.402** (0.099)	1.031 (0.438)	**1.645** (0.002)	-0.288 (0.847)
PC1 – soil	**0.208** (0.031) +11.378 [5]	**1.044** (0.050) +0.471 [3]	**1.599** (0.009)	3.847 (0.194)	1.315 (0.146)	2.644 (0.340)	-0.191 (0.651)	1.383 (0.670)
Human-impact layers								
Crops	-0.226 (0.134) +14.491 [6]	0.858* (0.054) +15.371 [5]	0.819 (0.947)	2.484 (0.202)	0.836 (0.880)	1.270 (0.438)	-0.240 (0.279)	1.414 (0.297)
Landcover	**0.582** (<0.001) 0.000 [1]	**1.358** (0.006) +39.775 [6]	1.272* (0.054)	3.392 (0.194)	1.249 (0.201)	4.423 (0.318)	-0.222 (0.407)	-0.182 (0.816)
Population density	**0.227** (0.028) +10.821 [4]	0.912* (0.053) -0000 [1]	**1.656** (0.007)	3.808 (0.194)	1.401* (0.099)	2.873 (0.318)	-0.717* (0.091)	-2.071* (0.092)

911
912

42
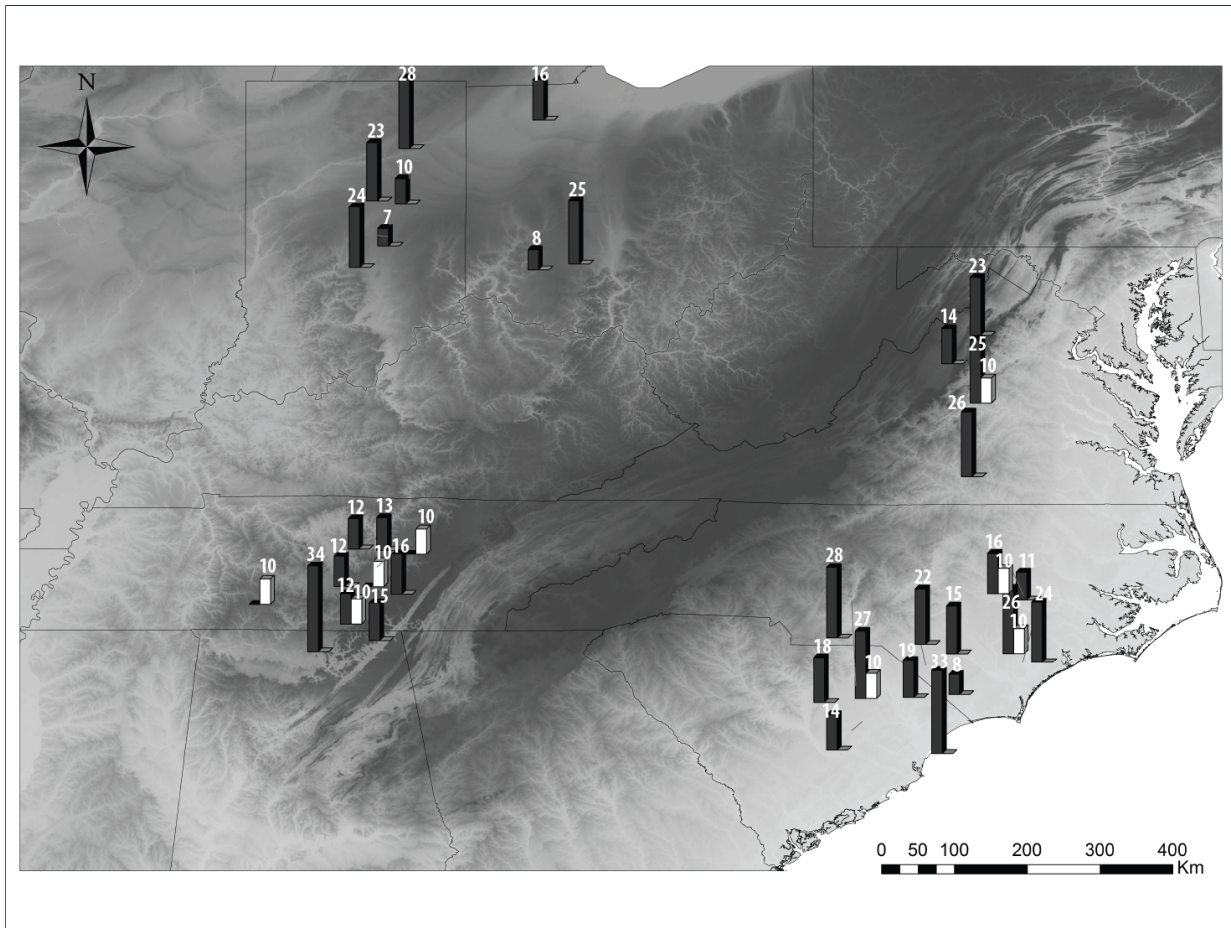
913
914 Fig. 1. Distribution of *Ipomoea purpurea*'s sampled localities. Sample sizes for both SSR
915 (black bars) and SNP (white bars) datasets are indicated (numbers on top indicate the
916 actual number of individuals used in our analyses). Elevation is provided as background.
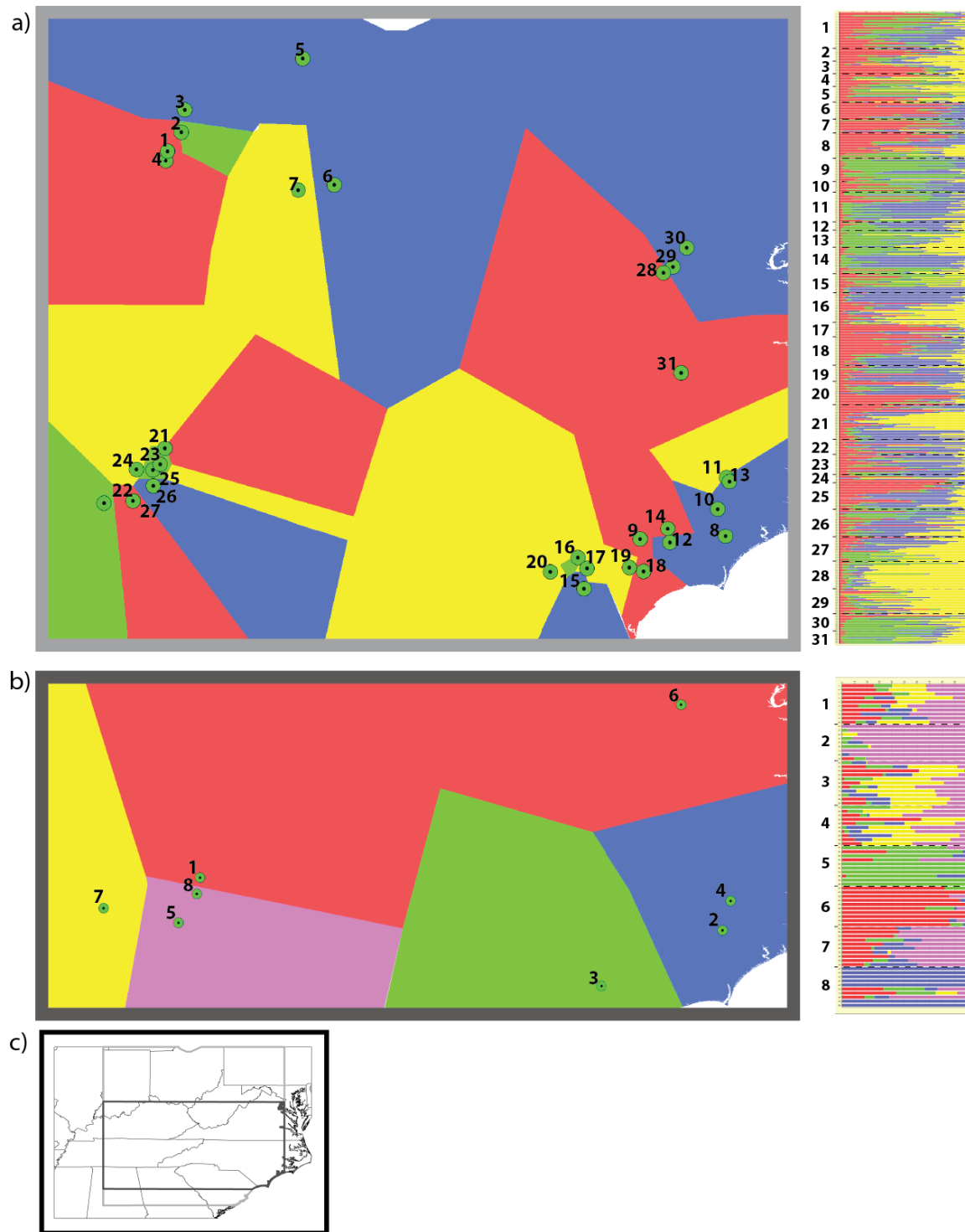
Fig. 2. Admixture and spatial clusters identified in *I. purpurea*. Results for the SSR (a) and SNP (b) datasets are shown together with an inset map (c) denoting the location of those clusters. Individuals in the admixture plot are sorted by sample locality.
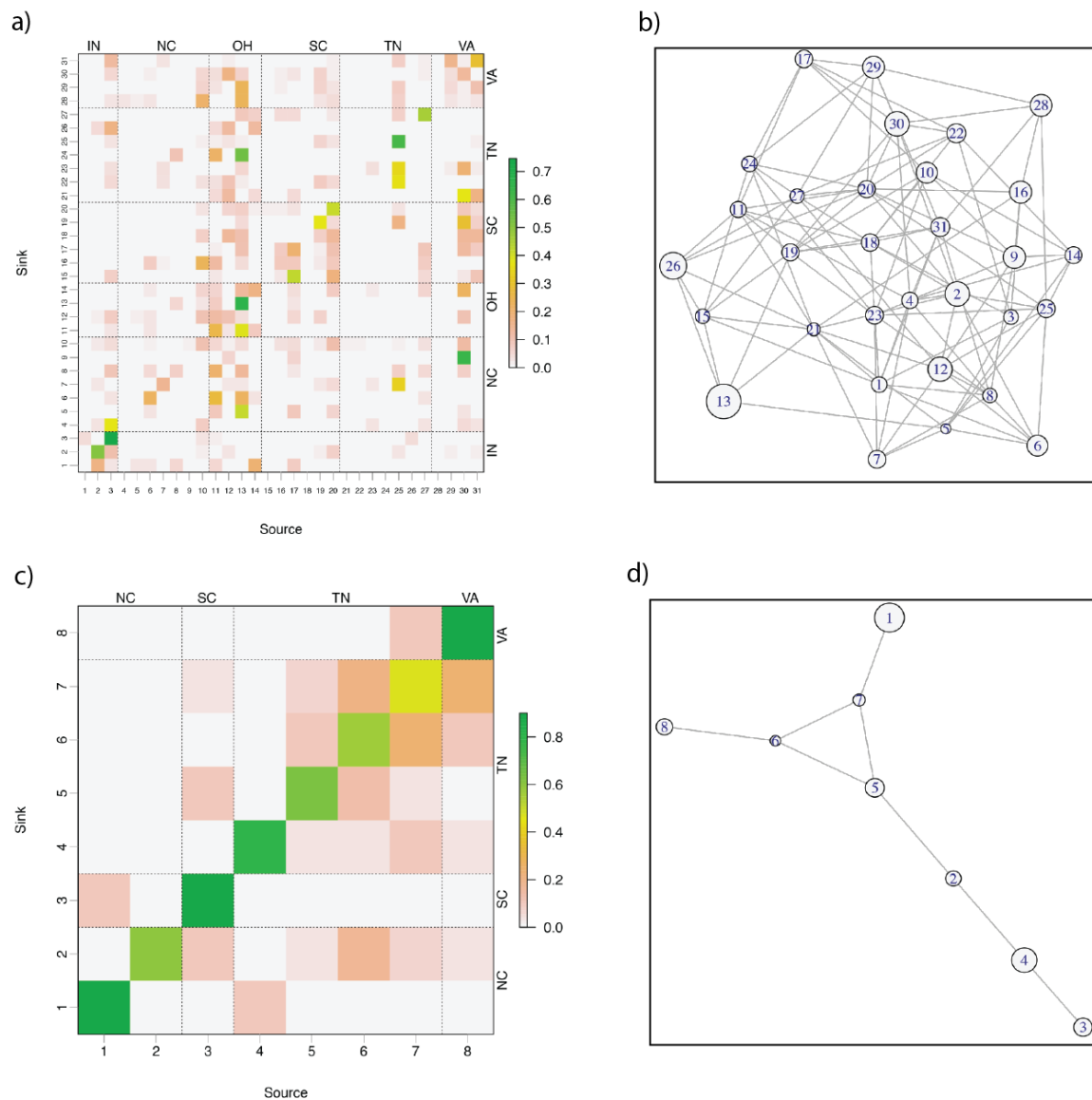
Fig. 3. Inferred population connectivity. (a,c) The estimated origin of individuals for each sampled locality (i.e., sink) is depicted according to the locality they were inferred to have originated from (source). The color of each cell in these plots depicts the proportion of individuals in the sink 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. Localities are sorted by State. (b,d) Pruned conditional genetic networks. Note the great difference in complexity and interconnectedness between networks. The top row shows SSR-based results, the bottom, the SNP-based results. Locality numbers follow Fig. 2.
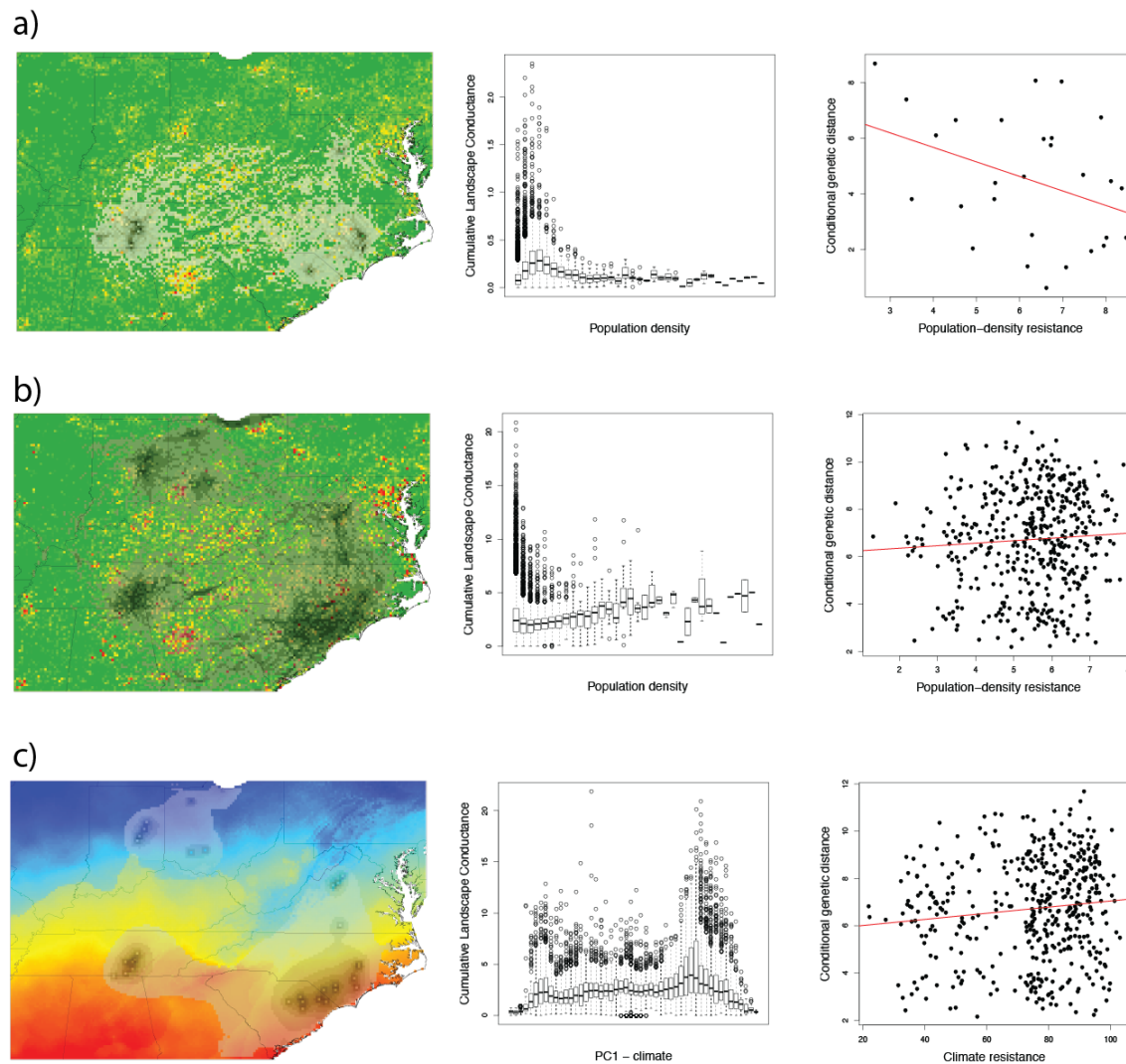
45

933
934
935 Fig. 4. Landscape conductivity patterns recovered. Conductance maps for variables
936 consistently recovered as significant predictors of genetic dissimilarity are overlaid on
937 maps of those variables to indicate the characteristics of the landscape that facilitate
938 connectivity between sampled localities (left column). Boxplots showing the association
939 between each variable values (binned into categories for simplicity) and conductance
940 (center column), and regression between landscape resistance corresponding to each
941 variable and genetic dissimilarity (right column) are also shown. The 3 associations
942 portrayed are: (a) human population density – SNP differentiation association, (b) human
943 population density – SSR differentiation association, and (c) PC1 climate and SSR
944 differentiation.
945

46