MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2078–2091 doi: 10.1111/j.1365-294X.2012.05454.x

Adaptation with gene flow across the landscape in a dune sunflower

ROSE L. ANDREW,* KATHERINE L. OSTEVIK,* DANIEL P. EBERT* and LOREN H. RIESEBERG*†
*Department of Botany, University of British Columbia, 3529-6270 University Blvd, Vancouver, BC V6T 1Z4, Canada,
†Department of Biology, Indiana University, 1001 E Third St., Bloomington, IN 47405, USA

Abstract

Isolation by adaptation increases divergence at neutral loci when natural selection against immigrants reduces the rate of gene flow between different habitats. This can occur early in the process of adaptive divergence and is a key feature of ecological speciation. Despite the ability of isolation by distance (IBD) and other forms of landscape resistance to produce similar patterns of neutral divergence within species, few studies have used landscape genetics to control for these other forces. We have studied the divergence of *Helianthus petiolaris* ecotypes living in active sand dunes and adjacent non-dune habitat, using landscape genetics approaches, such as circuit theory and multiple regression of distance matrices, in addition to coalescent modelling. Divergence between habitats was significant, but not strong, and was shaped by IBD. We expected that increased resistance owing to patchy and unfavourable habitat in the dunes would contribute to divergence. Instead, we found that landscape resistance models with lower resistance in the dunes performed well as predictors of genetic distances among subpopulations. Nevertheless, habitat class remained a strong predictor of genetic distance when controlling for isolation by resistance and IBD. We also measured environmental variables at each site and confirmed that specific variables, especially soil nitrogen and vegetation cover, explained a greater proportion of variance in genetic distance than did landscape or the habitat classification alone. Asymmetry in effective population sizes and numbers of migrants per generation was detected using coalescent modelling with Bayesian inference, which is consistent with incipient ecological speciation being driven by the dune habitat.

Keywords: divergence with gene flow, ecological speciation, *Helianthus petiolaris*, isolation by adaptation, landscape genetics

Received 22 September 2011; revision received 3 December 2011; accepted 10 December 2011

Introduction

When populations diverge owing to different ecological conditions, despite continuing gene flow between them, ecological speciation can be the outcome. In such cases, divergence can be hastened not only by the reduction in gene flow between populations that results from their spatial separation but also by the barrier to migration created by selection against maladapted immigrants (Barton & Bengtsson 1986; Gavrilets & Cruzan 1998). This process, often called 'isolation by adaptation' (IBA) or 'ecologically dependent reproductive isolation', is a key component of the picture of ecological speciation that has been emerging for the past several decades (Hendry 2004; Nosil *et al*. 2009).

Selection against immigrants is taxonomically widespread (Mallet & Barton 1989; Nagy & Rice 1997; Sambatti & Rice 2006; Tobler *et al*. 2009) and has important implications as a reproductive barrier (Nosil *et al*. 2005). The impact on divergence at neutral loci that are unlinked to the targets of selection has been observed in a number of species, including walking-sticks (Nosil *et al*. 2008), sticklebacks (McCairns & Bernatchez 2008) and silverswords (Friar *et al*. 2007). Evidence of IBA is most convincing when isolation by distance (IBD) and

Correspondence: Rose Andrew, Fax: 1 604 822 6089;
E-mail: andrew@biodiversity.ubc.ca

other non-adaptive forces are accounted for. Yet in very few cases have the tools of landscape ecology and genetics been applied to studies of ecological divergence (McCairns & Bernatchez 2008; Freedman *et al.* 2010). In the present study, we used neutral molecular markers in a landscape genetics framework to test the hypothesis that IBA contributes to ecotypic divergence.

Identifying IBA requires that other factors that could generate genotype–environment associations be accounted for. In addition to sharing a habitat, members of one population are likely to be closer to one another than to individuals outside the population. For this reason, they are expected to share more genes with each other than with members of other populations as a result of IBD, even in the absence of any non-neutral processes. The tendency of different parts of the landscape to gene flow can also shape genetic connectivity among subpopulations and play a role in the divergence of populations (Manel *et al.* 2003). Landscape geneticists have employed several methods to study this process, mainly in the context of animal conservation ecology (Storfer *et al.* 2010). Prominent among these is a form of circuit analysis that analyses spatially structured gene flow in a way similar to electrical current, which is impeded by materials of high electrical resistance or by narrow conduits (McRae 2006; McRae *et al.* 2008). The appeal of this approach lies partly in its ability to consider a number of possible paths of movement between two points, treating the landscape as an electrical network. The ecological analogue of electrical conductance (the inverse of resistance) is the exchange of migrants, which in animals may be most easily interpreted as resulting from the choices made by individual animals. In annual plants, however, the resistance of a given habitat type is determined by its impact on dispersal distances and the suitability for growth and reproduction of propagules dispersing to it (McRae & Beier 2007). The genetic isolation that is attributable to the resistance landscape is termed isolation by resistance (IBR) (McRae 2006). The key difference between IBR and IBA is that IBR is determined by the landscape and by the intrinsic biology of the organism, whereas IBA results from the interaction between specific genotypes and the environments to which they are well or poorly adapted. IBR is often a better predictor of population differentiation than IBD (McRae & Beier 2007; Row *et al.* 2010; Moore *et al.* 2011), making it a suitable null hypothesis for the identification of IBA. Therefore, it is important to test for associations between habitats and genetic divergence in a model that accounts for both geographic distance and the most likely factors contributing to landscape resistance (Manel *et al.* 2010a).

Associations of genetic distance with environmental gradients might arise without IBA if the edge of the habitat patch coincides with a barrier to gene flow (e.g. endogenous reproductive isolation, Bierne *et al.* 2011). Although this may be unlikely in many cases, the possibility should be considered as an alternative explanation for strong divergence between habitats. Fortunately, this issue can be addressed if habitat patches do not have sharp edges or if the environmental variables that characterize the habitats vary within habitat types. In such cases, the association between genetic structure and environmental differences among sampling locations is a useful way of testing for IBA (McCairns & Bernatchez 2008; Manel *et al.* 2010b).

We have studied an extreme case of ecotypic divergence with gene flow in the prairie sunflower, *Helianthus petiolaris*. This annual sunflower occurs in sandy soils throughout the mid-west and southwest of the USA. Uniquely however, in the Great Sand Dunes National Park and Preserve in Colorado, the species occurs abundantly in active sand dunes; indeed, at this location, it inhabits the tallest dune system in North America. While maintaining the characteristic features of *H. petiolaris*, the dune ecotype is morphologically differentiated in the field and in common garden experiments (R. Andrew, unpublished). For example, the dune form produces large seeds, which are thought to be an adaptation to habitat in the dune endemic, *Helianthus anomalus* (Donovan *et al.* 2010).

The dune ecotype of *H. petiolaris* is intriguing because it appears to be diverging despite ample opportunities for gene exchange with the typical form of the species, which occurs in habitats surrounding the active dune field in the stabilized sandy soils of the sand sheet, vegetated dune and alluvial deposits. Subpopulations of the different ecotypes are in some cases separated by <100 m, which is well within the range of many insect pollinators. Three non-mutually exclusive forces may help to maintain the divergent ecotypes in the face of likely gene flow. First, IBD might reduce gene flow between the core of the dune population and the non-dune form. Second, tall dunes and large stretches of bare sand may act as barriers to dispersal within the dunes, which could add IBR to the effect of geographic isolation on the differentiation of sunflower populations. Third, the dune and non-dune habitats might differ so greatly that immigrants may be unable to survive or reproduce: in effect, IBA. We set out to test which of these hypotheses fit the patterns of neutral genetic divergence within the system by analysing variation at microsatellite loci in the context of geography and environmental variables. In particular, our aim was to test whether IBA remained a good predictor of genetic differentiation when other forces were also considered.

As well as characterizing neutral divergence between the sunflower ecotypes, we used the techniques of landscape genetics to identify the signature of IBA in two ways. We first asked whether genetic divergence between ecotypes is greater than that within ecotypes. We then confirmed that this divergence is greater than would be expected based on geographic distance alone, also taking into account differences between habitats in landscape resistance to gene flow. Our second approach was to measure environmental variables at each site and to search for significant associations with genetic distance, to identify environmental variables that might be driving divergence. By confirming that genetic distance is related to environmental variables, this approach supported to the hypothesis of IBA.

Materials and methods

Sampling

In September 2008, we visited 21 subpopulations of *Helianthus petiolaris* in Great Sand Dunes National Park (Table 1). These included subpopulations growing in the main dune field, the non-dune habitat surrounding the dune field and intermediate habitats at the margin of the dune field. At each sampling location, we sampled seeds from 15 to 30 plants, each separated by at least 1 m.

Plot data

In May 2009, we returned to these sampling locations to deploy plant root simulator (PRS) probes (Western Ag Innovations, Saskatoon, SK, Canada), which use ion exchange membranes to measure soil nutrient availability. Four pairs of probes (anion and cation exchangers) were deployed per site, each at the southeastern corner of a 1×1 m quadrat. For each quadrat, we recorded the percentage cover of grasses, forbs, debris and bare ground. At the time of assessment, most sunflower seedlings had emerged but were still at the cotyledon stage. In September 2009, the nutrient probes were collected and processed. In some cases, markers had been blown away or disturbed by animals, and probes could not be recovered.

In the landscape genetics analysis, subpopulations were represented by mean values of the environmental variables. Soil nutrient probes were pooled during analysis by Western Ag Innovations, so no averaging was necessary. Subpopulation means for the cover data were estimated using generalized linear models with a logit link function. Additional composite variables were generated by principal components analysis (PCA) of standardized soil or cover data. The vegetation cover classes were arcsine-square-root-transformed prior to PCA.

Table 1 Locations of subpopulation samples and plots where cover was assessed and soil nutrient probes located

Site	Habitat type	Northings	Eastings	Samples	Probes recovered	*Helianthus* density May (/m^2)	*Helianthus* density September (/m^2)	*Helianthus* mean survival (%)
970	N	14168520	444836	19	0	4.75		
1003	D	14180849	449296	13	3	94.75	18.00	38.6
1033	D	14182289	449894	13	4	6.75	4.75	90.2
1063	N	14180939	447346	13	4	7.00	0.50	21.2
1094	I	14185951	445744	15	4	1.75	1.00	41.7
1117	D	14185640	447459	21	1	108.75	51.00	36.4
1147	N	14187750	446930	19	4	3.50	2.50	50.0
1180	D	14180747	451327	14	3	32.75	26.00	78.8
1240	D	14182239	453271	14	4	24.25	21.00	88.5
1270	D	14179804	453854	20	4	195.00	51.75	27.2
1300	D	14177695	451904	22	4	16.25	7.25	36.5
1331	I	14175581	451253	13	4	8.75	4.00	70.3
1363	N	14174455	453024	14	4	1.75	1.00	58.3
1500	N	14179922	445771	14				
1547	D	14179472	449829	13	3	31.00	3.00	5.2
1701	D	14184103	453889	13	3	21.25	20.33	66.7
1731	D	14184594	451267	13	4	10.25	4.00	25.9
1761	I	14185327	452941	14	4	8.25	11.50	100.0
1791	N	14185209	454644	13	4	1.50	1.00	66.7
2001	N	14178933	455337	21	4	0.75	0.00	0.0
2250	N	14169533	447833	14	4	6.50	6.00	92.3

Location data are Universal Transverse Mercator coordinates in zone 13 S. Site 1500 was added to the data set after probes were deployed and vegetation assessed.

Molecular data

DNA was extracted from frozen tissue using a modified CTAB method (Doyle & Doyle 1987). Transcriptome-derived microsatellite loci from previous studies (Kane & Rieseberg 2007, 2008; Yatabe *et al.* 2007) were tested in a subset of the Great Sand Dunes NPP samples. The remaining individuals were genotyped at the 15 most reliable loci, plus one anonymous microsatellite located close to a quantitative trait locus for seed weight (ORS847; Lexer *et al.* 2005). Summary statistics are given in Tables S3 and S4 (Supporting information). Microsatellite alleles were amplified following Kane *et al.* (2009), separated using an ABI3730 and scored using ABI GeneMapper (Applied Biosystems, Foster City, CA, USA). In about half of the loci, allele sizes did not always differ by multiples of the repeat unit, which is not unusual for cross-amplified microsatellites.

Analysis

Genetic structure: differentiation of ecotypes. Genetic structure was investigated using a number of approaches: (i) analysis of molecular variance (AMOVA) and F-statistics; (ii) PCA; and (iii) non-hierarchical Bayesian clustering.

Weir & Cockerham's (1984) single-locus and multilocus θ_{ST} were estimated with and without correcting for null alleles in FREENA (Chapuis & Estoup 2007). Using GENALEX (Peakall & Smouse 2006), multilevel AMOVA was performed on all samples or on the dune and non-dune subpopulations only (i.e. omitting the intermediate sampling locations). This allowed differentiation between ecotypes to be assessed while taking into account the subpopulation structure within each habitat type. Shannon's index offers a means to partition diversity within and among groups, which complements AMOVA and F-statistics (Sherwin *et al.* 2006). This analysis, also performed in GENALEX, permits a goodness-of-fit test of differentiation, as well as population-specific indices of information.

We next asked whether the ecotype delineation represents natural groups by performing principal coordinates analysis (PCoA) at the subpopulation level. Based on pairwise Euclidean genetic distance, we performed correlation-standardized PCoA at the subpopulation level in GENALEX.

To explore the genetic groups within the samples, non-hierarchical Bayesian clustering was performed using the program Structure 2.3 (Pritchard *et al.* 2000; Falush *et al.* 2003, 2007). The correlated allele frequency model with admixture was used with a burn-in of 1 million replicates, followed by 4 million replicates, with convergence monitored for each run. Ten runs were performed for each value of *K*, and the output was interpreted with Structure Harvester (Earl 2011), using the methods of Pritchard *et al.* (2000) and Evanno *et al.* (2005); CLUMPP was used to average admixture proportions over runs (Jakobsson & Rosenberg 2007).

Isolation by distance. IBD was tested in two ways. First, a Mantel test was performed in GENALEX between pairwise genetic distance (average Euclidean distance between subpopulations) and geographic distance, with 10 000 permutations. In addition, spatial autocorrelation analysis was performed on genetic distances between subpopulations in GENALEX 6.4 (Smouse & Peakall 1999; Peakall & Smouse 2006). Multiple distance classes were tested to have the best chance of identifying spatial genetic structure, and analyses were repeated at both population and individual levels. Permutation and bootstrap tests were conducted with 10 000 replicates in each case.

Landscape genetics. Comprehensive modelling of connectivity in this system is outside the scope of this study, as there were not sufficient sampling locations to test complex explanatory scenarios. For this reason, we limited our hypotheses to those that could be tested based on simple matrix comparisons with pairwise genetic distances between subpopulations. Several measures of genetic divergence or differentiation were considered, but only the results using the Cavalli-Sforza and Edwards genetic distance are shown here, as it was consistently the most strongly associated with both geographic and environmental distances. To ask whether genetic divergence between ecotypes is greater than that within ecotypes, we generated a pairwise matrix of habitat differences. All within-habitat-type comparisons were assigned a value of 0, whereas all dune vs. non-dune comparisons were assigned a value of 1. Comparisons of dune or non-dune subpopulations with those in intermediate habitat were assigned a value of 0.5. In addition to Mantel and partial Mantel tests (Legendre & Fortin 1989), this matrix was then used as a predictor matrix in an analysis based on these procedures, multiple regression of distance matrices (MRDM), in which genetic distance was the response variable (Lichstein 2007). Briefly, partial regression slopes are estimated using standard multiple linear regression, but the significance of each term is determined by randomly permuting the explanatory variables one at a time while holding the others constant. Both approaches were implemented using the *ecodist* package in the statistical program R (R Core Development Team, 2011), with 100 000 permutations per model.

We wished to assess the power of habitat to predict genetic distance in the face of competing hypotheses. The first of these is that IBD generates patchy spatial

genetic structure, which happens to coincide with the dunes by chance. That is, dune populations are genetically similar to one another simply because of their spatial proximity (IBD). We accounted for this explanation by controlling for geographic distance while testing for an association between genetic distance and habitat difference. If selection against immigrants does indeed present a barrier to gene flow, this association should persist when controlling for IBD.

The second alternative hypothesis we wished to consider was that differences in landscape resistance between the dune and non-dune habitat enhance differentiation. We used circuit theory (McRae & Beier 2007; McRae & Shah 2009) to characterize potential differences in landscape resistance associated with vegetation type. The Great Sand Dunes Vegetation map (Salas *et al.* 2010) was reclassified in ARCGIS 9.3 and exported using the add-in *Export to Circuitscape* by Jeff Jenness (http://www.circuitscape.org/Circuitscape/ArcGIS.html). The map was converted to a raster with 30 × 30 m grid cells and reclassified into three habitat classes: (A) bare dune, (B) vegetated sunflower habitat and (C) impermeable non-habitat (Fig. 1, Table S1, Supporting information). Briefly, stable patches of vegetation in the dune field and ponderosa pine grassy woodland in the foothills were both classified as B as they represent suitable sunflower habitat. Although some sunflowers grow on bare dune faces, their numbers and reproductive output are usually much lower than those growing in the more vegetated patches. *Helianthus petiolaris* is specialized on sandy soils and does not tolerate salt, montane, forest and sabkha vegetation types, which were classified as C. An infinite resistance (conductance = 0) was always assigned to class C. Circuitscape (McRae & Shah 2009) was used to calculate total resistance between pairs of populations, based on average resistance and four-neighbour connections only. Using average conductance and eight-neighbour connections did not affect the results substantially. To parameterize the resistance surface, a range of conductance values were assigned to habitat classes A and B. We first optimized the univariate association with genetic distance before including habitat in the model (Spear *et al.* 2010). Because the resulting pairwise resistance matrix scales with the resistance (or conductance) surface and the magnitude does not affect MRDM on scaled matrices, we controlled the A and B conductances to vary the dune/non-dune conductance ratio. The resulting resistance matrix was considered as a representation of IBR at the scale of the study.

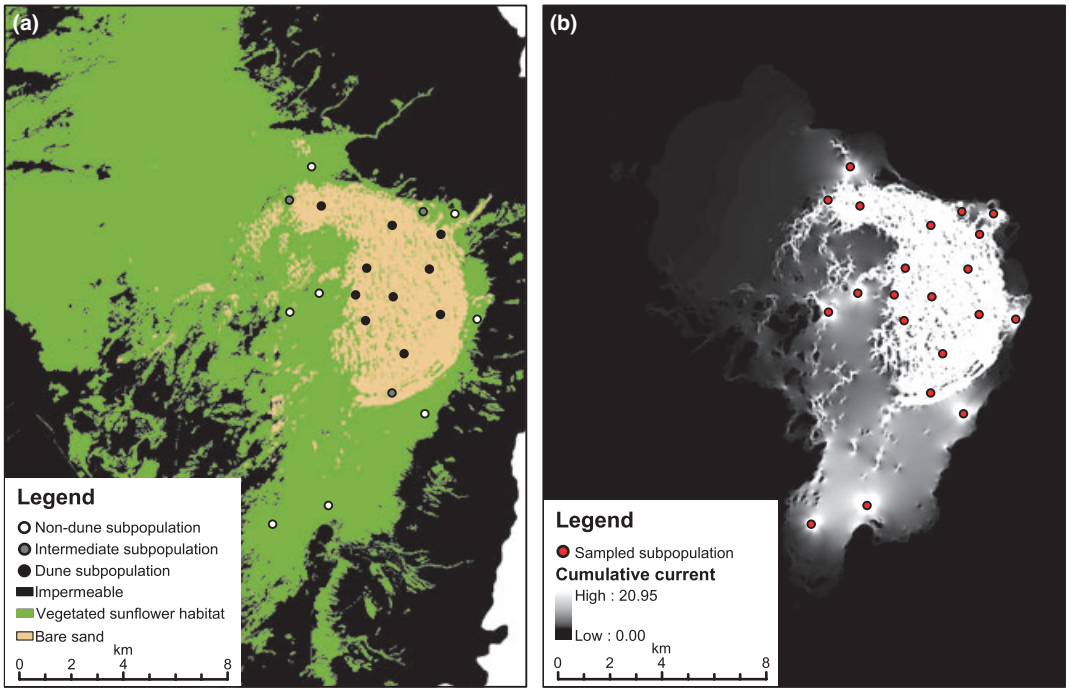

Fig. 1 Maps of Great Sand Dunes NP. (a) Vegetation resistance surface reclassified for circuit analysis. Areas shown in black are considered impermeable (conductance = 0; class C) in all analyses, whereas non-dune and dune habitat (classes B and A, represented by green and tan areas, respectively) are assigned varying levels of conductance to find the best fit to the genetic distance matrix. Dune, intermediate and non-dune subpopulations are indicated by circles of black, grey and white, respectively. (b) Cumulative current map derived from Circuitscape analysis with a conductance ratio of 1000:1 dune/non-dune. Light areas represent higher connectivity; dark areas indicate higher landscape resistance.

MRDM and partial Mantel tests were used to test the effect of habitat on genetic distance while controlling for IBD and IBR. A positive result would be interpreted as evidence that IBA helps to explain the differentiation of the sunflower ecotypes. The advantage of landscape resistance over geographic distance was first tested by omitting habitat from the full model. However, it should be noted that this is not a straightforward test of IBD compared with IBR, as the resistance matrices integrate both distance and landscape resistance when resistance is non-zero. One-sided *P*-values were used throughout the analysis of distance matrices, and all distance matrices (including genetic and habitat distances) were scaled to zero mean and unit variance so that MRDM models yielded standardized partial regression slopes.

Our second approach was to test the association of genetic distance with environmental variables measured at each site. These give a more detailed description of the habitat variation and vary on a more continuous scale than the coarse habitat classification. Thus, genetic distance may be better predicted by environmental variation than by habitat classification alone. Using MRDM with Euclidean distances along soil and vegetation cover axes included as predictor variables, we tested whether these variables were significantly associated with genetic distance while controlling for geographic distance and vegetation resistance. We then asked whether this association persisted when habitat class was also included in the model. As multiple tests were conducted, we adjusted *P*-values using the Benjamini and Hochberg (1995) procedure for controlling false discovery rate. We aimed to identify (i) candidate variables that may be driving differentiation between ecotypes and (ii) proxy variables that may be used as functional measures of dune-like characteristics for future studies.

Asymmetric migration. In order to assess the direction of gene flow between the ecotypes, we obtained Bayesian estimates of migration rates and effective population sizes using MIGRATE-N (Beerli & Felsenstein 2001; Beerli 2006). This program fits a coalescent island model to the data, allowing asymmetric migration rates. The subpopulations exhibiting the least admixture were selected and pooled for each ecotype. These were dune subpopulations 1003, 1180, 1240 and 1300 and non-dune subpopulations 970, 1363, 1500 and 2250, yielding total sample sizes of 63 and 61, respectively. Three microsatellite loci (HT536, HT520 and HT279) were excluded because many of their alleles differed by only a single base pair. Most of the alleles at the remaining loci varied in size by multiples of the repeat unit length. The few that did not were adjusted to fit a stepwise model of repeat mutation with as little loss of information as possible.

Uniform prior distributions ranging from 0 to 100 were used for both theta and *M*, with starting values estimated from F_{ST}. Relative mutation rates were estimated from the data, and a Brownian mutation model was employed. The Metropolis–Hastings algorithm was used to sample from the prior distributions and generate posterior distributions. Two replicate runs with 500 000 recorded steps and a sampling increment of 100, following 1 million burn-in trees. A static heating scheme was employed with four chains at temperatures of 1, 1.5, 3 and 1 million.

Migration models constraining mutation-scaled theta and/or migration rate to be symmetric between ecotypes were compared to the full model (containing four parameters). The Bezier approximation to the marginal likelihood was used to test which model fits the data best (Beerli & Palczewski 2010).

Results

Field measurements

The non-dune sites varied greatly but tended to be richer in the major nutrients than the dune or intermediate sites (Fig. 2). The difference between dune and non-dune habitats (excluding the intermediate subpopulations) was significant for total N, NO_3-N, P and Ca only; however, this difference was not supported when the false discovery rate was accounted for (Table S2, Supporting information). All cover variables differed significantly between dune and non-dune habitats (Fig. 2, Fig. S1, Table S2, Supporting information). The first three PCoA of the soil nutrients explained 34.5%, 23.2% and 10.6% of the total variation, respectively, while those of the cover variables explained 67.7%, 21.8% and 7.7%, respectively.

Molecular data

Estimates of Weir & Cockerham's (1984) θ_{ST} strongly suggested that one locus may be influenced by selection. The θ_{ST} estimate of HT285 (0.252) was 10 standard deviations greater than the mean of the other 15 loci (mean = 0.035, SD = 0.021) and showed much lower diversity within the dunes than elsewhere (0.09 vs. 0.50). The log-ratio of genetic diversity (Schlötterer & Dieringer 2005) in the dune ecotype relative to the non-dune ecotype was −3.9, suggesting that a selective sweep had reduced variation in the region around HT285. When standardized relatively by the overall mean and standard deviation (including HT285), lnRH for HT 285 is highly statistically significant ($P = 0.0014$). This locus was therefore omitted from all further analyses. ORS847, HT333 and HT440 displayed high F_{IS}, suggesting null alleles;

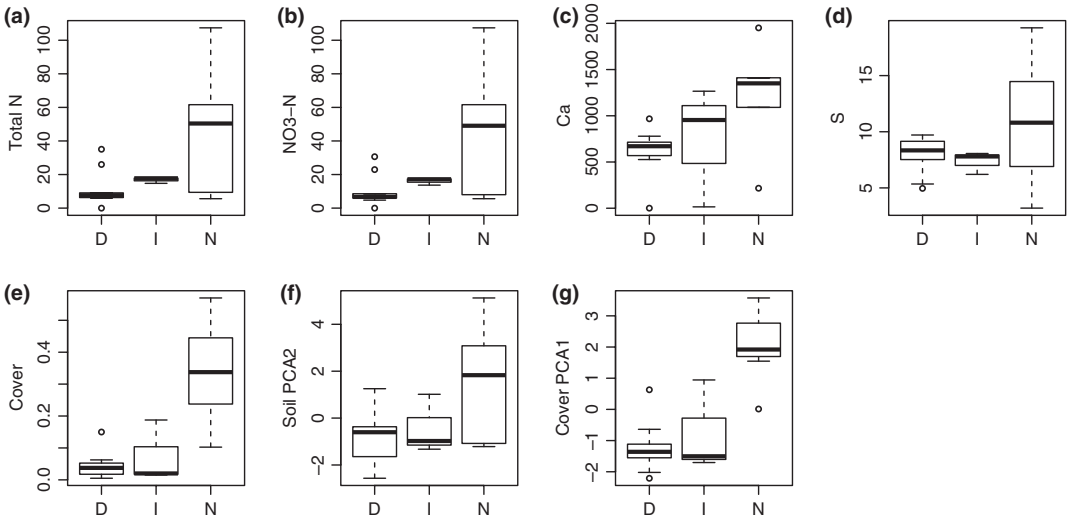

Fig. 2 Comparison of selected variables among habitat types. Shown here are the five variables that are significantly correlated with genetic distance after correction for multiple comparisons (see Fig. S1, Supporting information for other variables). Soil nutrient availability during the growing season was measured using ion exchange probes, and the total absorption was scaled to the maximum burial time (123 days), such that in each case, the unit of measurement was mg/10 cm^2/123 days. D, dune; I, intermediate; N, non-dune populations.

however, no substantive differences were observed in analyses conducted with or without these loci. Global θ_{ST} was little affected by the inclusion of estimated null allele frequencies (Chapuis & Estoup 2007), as were the landscape genetics analysis described below.

Genetic structure

AMOVA and diversity statistics. Hierarchical analysis of molecular variance suggested low, but significant, differentiation between ecotypes; overall, 2.3% of variation occurred among habitat types ($P < 0.001$), 4.1% among subpopulations ($P < 0.001$) and 93.5% within subpopulations ($P < 0.001$). Excluding the intermediate subpopulations, a greater proportion of variance (3.3%) occurred between ecotypes and among subpopulations within ecotypes (4.4%).

Shannon's diversity index indicated significant differentiation ($P < 0.05$) for 15 of the 16 loci and much stronger differentiation at HT285 (Table S3, Supporting information). For most loci, the within-group diversity was lower for the dune samples (mean sHa = 1.14) than for the non-dune ones (mean sHa = 1.29). This pattern was much more pronounced for HT285 (0.35 and 2.14, respectively), but most measures of overall diversity were also slightly higher in the non-dune samples than in the dune samples for other loci (Table S4, Supporting information).

Principal coordinates analysis. The proportion of variance explained by each principal coordinates axis dropped off after axis 1, which represented 42% of the variation (axes 2 and 3 explained 19.9% and 17.35%, respectively). Principal coordinate 1 was strongly associated with ecotypes (Fig. 3), and intermediate subpopulations were nested within the dune subpopulations. In particular, subpopulations 1701 and 1731, from the northeastern part of the dune field, were more similar to the non-dune subpopulations than were the intermediate populations.

Non-hierarchical clustering of individuals. The posterior probability was higher for two subpopulations than for all other values of *K*, and the Evanno method also supported this model (Fig. S2a,b, Supporting information).

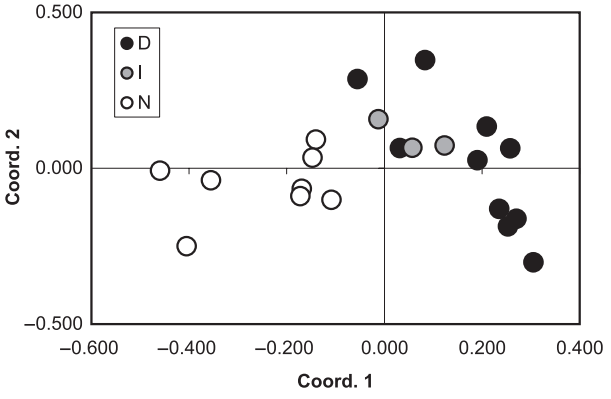

Fig. 3 Principal components analysis of microsatellite data at the subpopulation level. The first two axes are shown, representing 42% and 19% of the variation, respectively. The marker colours indicate dune (black), intermediate (grey) and non-dune (white) sampling locations.

Most samples with ancestry assigned to cluster 1 were from the non-dune habitat, while most cluster 2 individuals were from dune populations (Fig. 4, Fig. S3a, Supporting information). However, the 90% probability interval for each individual included either 1 or 0 (Fig. S3b, Supporting information), indicating that this analysis does not provide fine resolution of individual admixture, possibly due to the low level of overall differentiation. Separate analysis of the dune and non-dune subpopulations did not support any further subdivision (Fig. S4, Supporting information). Mapping the ancestry proportions onto sampling locations illustrated three features of the Structure results (Fig. 4): (i) there is a clear association between clusters and habitats; (ii) most dune and non-dune subpopulations show some evidence of admixture; and (iii) subpopulations that are closer to the margin of the dune field have higher proportions of admixture than those in the centre of the dune field or more distant from the dunes (Fig. 5).

At the subpopulation level, the proportion of dune ancestry was negatively correlated with Shannon's index (Fig. S5a, Supporting information). This indicates that a greater proportion of variation is unique to each non-dune population than to each dune population. Indeed, Shannon's index was low in the centre of the dune field and highest in the non-dune subpopulations most distant from the dunes (Fig. S5b, Supporting information).

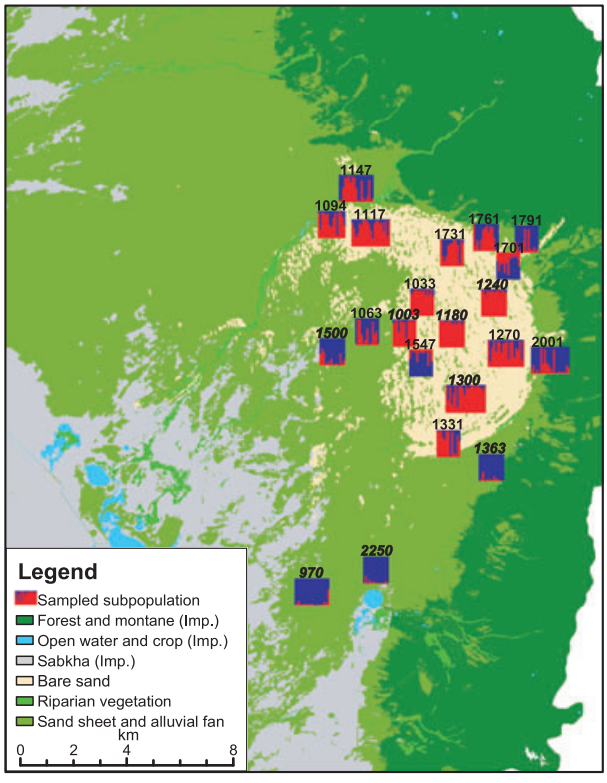

Fig. 4 Inferred ancestry of progeny, based on Structure analysis with $K = 2$, grouped by subpopulation and mapped onto sampling locations. Impermeable (Imp.) vegetation types include unsuitable habitat for *Helianthus petiolaris*, whereas bare sand and other non-dune vegetation types are considered permeable to sunflower migration. Subpopulations included in the Migrate-N analysis are shown in bold and italic script (dune: 1003, 1180, 1240, 1300; non-dune: 970, 1363, 1500, 2250).

Isolation by distance

The Mantel statistic based on pairwise Euclidean genetic distance and geographic distance was positive (Rxy = 0.126, $P = 0.003$), indicating that IBD explains a significant proportion of the variance in genetic distance among pairs of subpopulations. Both permutation and bootstrap tests detected significant spatial genetic autocorrelation in the 0- to 2.5 -km distance class, but not at greater distances (Fig. 6). This indicates increased similarity between neighbouring pairs of subpopulations. At the individual level, autocorrelation coefficients for distance classes including zero were substantially higher than the autocorrelation coefficients for greater distance classes (Fig. S6, Supporting information), confirming the significant subpopulation-level differentiation detected using AMOVA.

Landscape genetics

Genetic distance was positively associated with both habitat (Mantel's $r = 0.482$, $P < 0.001$) and geographic distance (Mantel's $r = 0.462$, $P = 0.001$). Genetic distances were highest in comparisons between dune and non-dune subpopulations, and intermediate subpopulations were typically most similar to dune subpopulations (Fig. S7, Supporting information). The partial

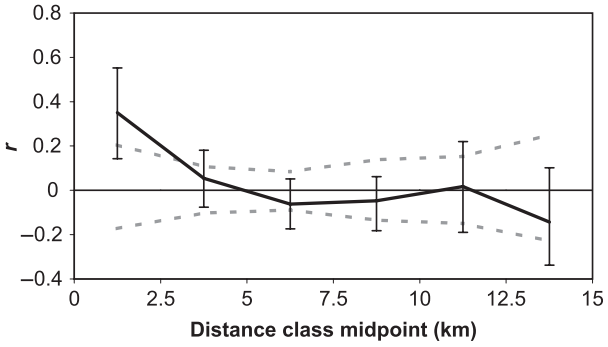

Fig. 6 Summary of subpopulation-level spatial autocorrelation analysis with 2500 -m distance classes. The autocorrelation coefficient (r, solid line) for the distance class (in m) was bias-corrected and tested against a null hypothesis of no spatial structure using both bootstraps (95% error bars) and permutations (95% confidence limits, dashed grey lines) with 10 000 replicates.

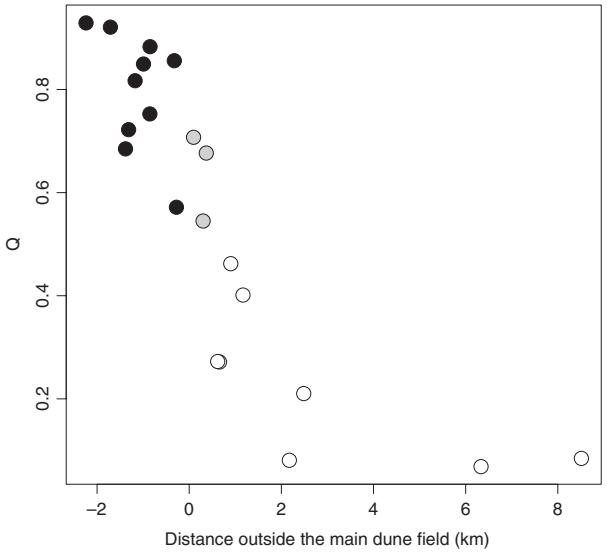

Fig. 5 Proportion of dune ancestry (admixture proportion, *Q*) is influenced by proximity to the interface between the dune and non-dune habitats. The *x*-axis measures distance from the dune edge, with negative values inside the dune field. Symbols are the same as in Fig. 3.

Table 2 Mantel and partial Mantel correlation coefficients used to test association of genetic distance with habitat while controlling for the effects of isolation by distance (IBD) and isolation by resistance (IBR)

Mantel test	r	*P*
G ~ Hab	0.482	<0.001
G ~ Res	0.546	<0.001
G ~ Dist	0.462	0.001
Hab ~ Dist	0.136	0.029
Hab ~ Res	0.166	<0.001
Dist ~ Res	0.685	<0.001
G ~ Hab | Dist	0.477	<0.001
G ~ Dist | Hab	0.457	0.001
G ~ Dist | Res	0.145	0.081
G ~ Res | Dist	0.354	0.007
G ~ Hab | Res	0.473	<0.001
G ~ Res | Hab	0.538	<0.001
G ~ Hab | Res, Dist	0.474	<0.001
G ~ Dist | Hab, Res	0.148	0.077
G ~ Res | Hab, Dist	0.350	0.006

Correlation coefficients between genetic distance (G), habitat (Hab), landscape resistance (Res, representing IBR) and distance (Dist, representing IBD) are shown with probabilities based on 100 000 permutations. The vertical bar symbol is used as shorthand for 'given' or 'controlling for'.

Mantel's correlation between genetic and habitat distance remained significant (Mantel's $r = 0.447$, $P < 0.001$) even when controlling for IBD (Table 2).

Landscape resistance was positively associated with genetic distance when the conductance of bare sand was set much higher than that of vegetated sunflower habitat (Table S5, Supporting information). The strength of this association continued to increase with this ratio to up to 100 000:1 (Mantel's $r = 0.546$, $P = 0.0007$), but the fit changed very little beyond 100:1 (Mantel's $r = 0.544$, $P = 0.0009$). We chose the resistance matrix generated by Circuitscape based on a 1000:1 ratio as it represented the signal effectively. A cumulative current map based on this analysis is shown in Fig. 1.

When controlling for landscape resistance, genetic distance remained positively associated with habitat distance, but not with geographic distance (Table 2). This implies that landscape resistance effectively captures the influence of Euclidean geographic distance on genetic differentiation, illustrating one of the useful features of circuit theory for landscape genetics. Although linear modelling of distance matrices can only account for the linear components of associations, autocorrelation of genetic, environmental and habitat distance matrices decreased approximately linearly with increasing distance (Fig. S8, Supporting information), so that the linear model was likely to capture the relationships well (Goslee & Urban 2007). To confirm this, we also performed rank-based MRDM, which gave results very similar to those of the linear models.

Genetic distance was more strongly associated with several environmental variables than with habitat distance in univariate models (Table 3). Controlling for landscape resistance in MRDM models, total soil nitrogen, NO_3 nitrogen and the second principal components axis based on the soil data had significant partial regression coefficients after adjusting *P*-values to account for multiple comparisons, as did total cover (Fig. 7, Table S6, Supporting information). Soil calcium, soil sulphur and the first vegetation cover PCA displayed significant univariate associations with genetic distance, but these associations were not significant when controlling for landscape resistance (Table S6, Supporting information).

We further examined the association between genetic distance and the environmental distance matrices by controlling for habitat type as well as landscape resistance. This association remained significant ($q < 0.05$) for total nitrogen, NO_3 nitrogen, and total cover but was marginal for the soil PCA1 (Table 3). Rank-based analysis gave similar results and identified the same variables of interest.

Asymmetric migration

A model allowing theta to differ between ecotypes but estimating a single symmetrical migration rate was superior to all of the other models, with a Bayes factor

Table 3 Results of multiple regression of distance matrices estimating the association of genetic distance (G) with environmental (Env, landscape resistance (Res) and habitat (Hab) distance matrices

	G ~ Res		G ~ Res + Hab + Env						
	Res		Res		Hab		Env		
Variable	*b*	*P*	*b*	*P*	*b*	*P*	*b*	*P*	*R*2
Total N	0.705	<0.001**	0.161	0.352	0.258	<0.001	0.546	<0.001*	0.570
NO_3-N	0.714	<0.001**	0.138	0.429	0.253	<0.001	0.565	<0.001*	0.575
Ca	0.382	0.010*	0.499	0.004	0.310	<0.001	0.047	0.699	0.415
S	0.563	0.004*	0.343	0.079	0.314	<0.001	0.319	0.053	0.470
Total cover	0.700	<0.001**	0.228	0.092	0.212	0.005	0.455	0.003*	0.558
Soil PCA2	0.581	<0.001*	0.351	0.035	0.285	<0.001	0.344	0.009***	0.494
Cover PCA1	0.311	0.009*	0.499	0.003	0.451	<0.001	−0.088	0.381	0.452

Only the environmental variables with a significant ($q < 0.05$) effect in the univariate model after controlling for false discovery rate are shown here (Table S6, Supporting information for further detail). Stars indicate significant partial regression coefficients: **$q < 0.001$; *$q < 0.05$; ***$q < 0.1$.

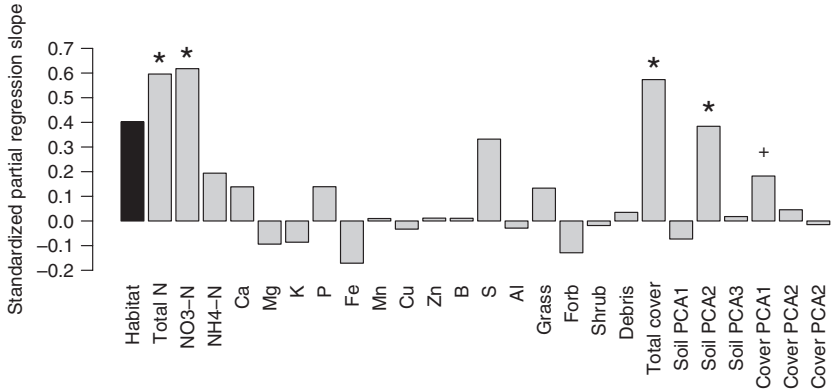

Fig. 7 Partial regression coefficients representing associations between genetic distance and environmental distance matrices (grey bars), controlling for landscape resistance using multiple regression of distance matrices. Variates were standardized prior to analysis. Asterisks indicate significant tests after controlling for false discovery rate ($q < 0.05$), while + indicates significance of uncorrected tests only ($P < 0.05$). The partial regression coefficient for habitat, controlling for landscape resistance, is also shown for comparison (black bar); although a q value was not estimated, its unadjusted P was 0.0006, similar to that of total N.

of at least 900 in each comparison. In this model, the posterior distributions for theta had means of 2.7 (95% CI from 0.6 to 4.4) in the dune ecotype and 4.6 (95% CI from 2.4 to 6.6) in the non-dune ecotype. The mutation-scaled migration rate was estimated as 3.8 (95% CI from 1.6 to 6.0), implying that the effective number of immigrants per generation was 2.5 into the dune population and 4.4 into the non-dune population. The symmetrical estimate based on Shannon's diversity decomposition was 7.2, while the F_{ST}-based estimate was 7.1 (using $\Phi_{RT} = 0.034$).

Discussion

The process of adaptive divergence is central to evolution and diversification. It is important to biologists because it can be an early step towards ecological speciation and can also help us to understand the ecological significance of traits that vary between habitats. The role of gene flow has been the subject of much study, and it has emerged that gene flow can both constrain and promote adaptive divergence (Hendry 2004; Garant *et al.* 2007). This relationship is a complex one, not least because adaptive divergence itself can reduce gene flow between habitats through IBA (Nosil *et al.* 2005). Using landscape genetics—especially circuit theory—we demonstrate how the effects of IBA can be partially disentangled from positively correlated and potentially confounding factors such as geographic distance and landscape resistance. Also, by integrating our landscape genetic analyses with characterization of habitat differences, we identify key environmental variables (or their proxies) that are responsible for IBA in dune sunflowers.

Differentiation and IBD

Although dune-adapted sunflowers, including those closely related to *Helianthus petiolaris* (Heiser *et al.* 1969; Rieseberg *et al.* 1990), are well known, this is the first case of intraspecific divergence in this habitat to be studied. Our investigation into the neutral genetic structure of *H. petiolaris* at Great Sand Dunes confirmed that the dune and non-dune ecotypes are indeed differentiated at neutral microsatellite loci. This conclusion was supported by AMOVA, Shannon's diversity, PCoA and Bayesian non-hierarchical clustering. Shannon's diversity indices provided further insight, suggesting that the genetic diversity that is unique to the dune ecotype is less than that of the non-dune ecotype. This pattern is reiterated at the subpopulation level and is consistent with asymmetrical selection and/or migration. The latter was supported by coalescent analysis of the ecotypes in an island model. Nevertheless, these analyses do not prove that IBA contributes to this differentiation between ecotypes because habitat differentiation covaries with geographic distance.

Two additional results emerging from these analyses highlight important caveats. First, there appears to be substantial admixture between ecotypes. As expected, based on the spatial scale of the system, genes appear to be moving between habitats, possibly via the movement of pollinators. Second, the genetic variation within and between habitats is clearly spatially structured. This was apparent from the Bayesian clustering approach, as well as spatial autocorrelation analysis and Mantel testing. The intermediate subpopulations displayed the most admixture, followed by subpopulations close to the edge of the dune field, while the populations furthest from the edge showed the least evidence of admixture. However, it is important to note that the marker system used here did not permit the estimation of precise admixture proportions for each individual, possibly because the overall degree of differentiation is insufficient. With greater marker resolution, it may be possible to ask whether individuals with migrant ancestry do indeed possess admixed genomes and whether they are first- or early-generation immigrants.

The divergence between the dune and non-dune ecotypes ($\Phi_{RT} = 0.033$) is substantially lower than that observed in most studies that have also identified IBA; however, it is within the range observed with similar markers (Friar *et al.* 2007; McCairns & Bernatchez 2008; Nosil *et al.* 2009).

Landscape resistance and divergence between habitats

A key test of our hypothesis that selection reduces gene flow between the dune and non-dune ecotypes was the association of genetic structure with habitat while controlling for geographic distance. However, the influence of geographic distance on dispersal is mediated by landscape features and may be nonlinear (Manel *et al.* 2003). Alternatives to Euclidean geographic distance have long been used in population genetics as more realistic predictors of dispersal (Kudoh & Whigham 1997; Lugon-Moulin *et al.* 1999). Landscape genetics has provided examples of intuitively appealing approaches to integrating the effects of distance and barriers or habitat types on gene flow, including cirguit theory (McRae & Beier 2007; McRae *et al.* 2008). In a simple application of this approach, we considered whether the bare sand of the dunes might represent a barrier to dispersal, which might enhance divergence between the ecotypes. Instead, we found that the circuit model parameterized with higher resistance in the vegetated habitat than over bare sand produced resistance matrices that were better predictors of genetic distance than Euclidean geographic distance.

Reduced resistance over bare sand, compared with vegetated sunflower habitat, is surprising given the proportion of the dune surface that is unfavourable for sunflowers. We expected bare sand to present a barrier to both pollen and seed dispersal, notwithstanding the typical leptokurtic dispersal curves of seed plants (Ellstrand 1992) and evidence for long-distance dispersal (Cain *et al.* 2000; Crisp *et al.* 2009). However, high-speed wind is potentially a very effective disperser of seeds in the dunes, provided that sunflower patches provide sufficient seed entrapment for the establishment of migrants (Fort & Richards 1998). In contrast, the shrubby vegetation of the non-dune areas is likely to limit wind-borne seed dispersal. Given sufficiently variable organelle markers, it may be possible in future to assess the relative role of seed and pollen in gene flow in the dune and non-dune habitats. Another potential explanation is that the genetic distances between dune subpopulations are smaller because of the lower genetic diversity of the dune gene pool, which could result from demography or selection. As asymmetric selection is a likely cause of the asymmetric migration and different effective population sizes between the ecotypes, the apparent resistance of the landscape may also reflect asymmetric selection in the different habitats. This underlines the fact that, although IBA and IBR are conceptually different, we are unable to completely disentangle their effects from patterns of genetic divergence in this system. Nevertheless, this resistance matrix is an effective way to account for differences in gene flow across the landscape that might shape divergence between habitat types, allowing us to identify the signature of adaptive divergence in the neutral portions of the genome.

Habitat was significantly associated with genetic distance when controlling for landscape resistance, supporting our hypothesis of IBA. However, our investigation into landscape resistance to gene flow was limited by the spatial scale of the sampling. Sampling at a finer scale would allow more detailed investigation into the influence of vegetation on the dispersal of sunflowers within and between ecotypes.

Association of neutral genetic variation with environments

A second test of IBA involved the association of genetic distance with finer-grained measures of environmental variation than those provided by habitat classifications. This comparison is important to rule out an intrinsic barrier to gene flow that coincides with the habitat edge (Bierne *et al*. 2011), in which case no association between environments and genetic distance is expected once habitat classification and landscape are accounted for. We found several plot characteristics that had significant effects on genetic distance even under these stringent conditions, representing soil nutrient availability and vegetation cover. These results are congruent with the identification of soil nitrogen as an important selective pressure for *H. anomalus*, a hybrid species that grows in the dunes of Utah and northern Arizona exclusively (Ludwig *et al*. 2006). Vegetation cover controls sand movement and thus may be acting as a proxy for important unmeasured variables, such as burial and exposure of seedlings (Donovan *et al*. 2010). Nonetheless, both environmental factors appear to be good indicators of general dune characteristics for future work, although caution should be used as, in this study, they did not fully account for the effect of habitat on genetic distance.

Conclusions

Despite the low level of differentiation among subpopulations, we have shown that habitat and environmental variables are better predictors of neutral genetic distance in our study system than either IBD or isolation by landscape resistance. This supports our hypothesis that adaptation is itself acting as a barrier to gene flow at these neutral loci. A possible caveat is that demography could generate similar patterns; for example, expansion along environmental gradients can produce associations between environments and genetic distance. However, this explanation seems unlikely because levels of gene flow between the dune and non-dune populations are high, and associations owing to demographic expansions would be quickly erased. RAD sequencing is currently underway to identify the genomic regions responsible for IBA.

Acknowledgements

We thank Matthew King, Connor Fitzpatrick, Phyllis Bovin, Fred Bunch, Paul Robertson, Great Sand Dunes National Park and Preserve and The Nature Conservancy for help with this project. We thank the editor and anonymous reviewers for their comments on the manuscript. This work was supported by a Killam Postdoctoral Fellowship to RLA and NSERC grant (327475) to LHR.

References

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **57**, 357–376.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

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

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Crisp MD, Arroyo MTK, Cook LG *et al*. (2009) Phylogenetic biome conservatism on a global scale. *Nature*, **458**, 754–756.

Donovan LA, Rosenthal DR, Sanchez-Velenosi M, Rieseberg LH, Ludwig F (2010) Are hybrid species more fit than ancestral parent species in the current hybrid species habitats? *Journal of Evolutionary Biology*, **23**, 805–816.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Journal*, **19**, 11–15.

Earl DA (2011) Structure harvester v0.6.1. Available at http://taylor0.biology.ucla.edu/structureHarvester/; accessed on 6 April, 2011.

Ellstrand NC (1992) Gene flow among seed plant populations. *New Forests*, **6**, 241–256.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data:

dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fort KP, Richards JH (1998) Does seed dispersal limit initiation of primary succession in desert playas? *American Journal of Botany*, **85**, 1722–1731.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.

Friar EA, Cruse-Sanders JM, McGlaughlin ME (2007) Gene flow in *Dubautia arborea* and *D. ciliolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing hybridization. *Molecular Ecology*, **16**, 4028–4038.

Garant D, Forde SE, Hendry AP (2007) The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, **21**, 434–443.

Gavrilets S, Cruzan MB (1998) Neutral gene flow across single locus clines. *Evolution*, **52**, 1277–1284.

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

Heiser CBJ, Smith DM, Clevenger SB, Martin WCJ (1969) The North American sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club*, **22**, 1–218.

Hendry AP (2004) Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research*, **6**, 1219–1236.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics*, **175**, 1823–1834.

Kane NC, Rieseberg LH (2008) Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Molecular Ecology*, **17**, 384–394.

Kane NC, King MG, Barker MS *et al.* (2009) Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*, **63**, 2061–2075.

Kudoh H, Whigham DF (1997) Microgeographic genetic structure and gene flow in *Hibiscus moscheutos* (Malvaceae) populations. *American Journal of Botany*, **84**, 1285–1293.

Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.

Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH (2005) Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **169**, 2225–2239.

Lichstein J (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Ludwig F, Jewitt RA, Donovan LA (2006) Nutrient and water addition effects on day- and night-time conductance and transpiration in a C_3 desert annual. *Oecologia*, **148**, 219–225.

Lugon-Moulin N, Brünner H, Balloux F, Hausser J, Goudet J (1999) Do riverine barriers, history or introgression shape the genetic structuring of a common shrew (*Sorex araneus*) population? *Heredity*, **83**, 155–161.

Mallet J, Barton NH (1989) Strong natural selection in a warning-color hybrid zone. *Evolution*, **43**, 421–431.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

Manel S, Joost S, Epperson BK *et al.* (2010a) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010b) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3835.

McCairns RJS, Bernatchez L (2008) Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment. *Molecular Ecology*, **17**, 3901–3916.

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

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, **104**, 19885–19890.

McRae BH, Shah VB (2009) *Circuitscape User's Guide*. The University of California, Santa Barbara. http://www.circuitscape.org.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.

Moore JA, Tallmon DA, Nielsen J, Pyare S (2011) Effects of the landscape on boreal toad gene flow: does the pattern–process relationship hold true across distinct landscapes at the northern range margin? *Molecular Ecology*, **20**, 4858–4869.

Nagy ES, Rice KJ (1997) Local adaptation in two subspecies of an annual plant: implications for migration and gene flow. *Evolution*, **51**, 1079–1089.

Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: ''isolation by adaptation'' and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Peakall R, Smouse PE (2006) GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

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

R Core Development Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, **44**, 1498–1511.

Row JR, Blouin-Demers G, Lougheed SC (2010) Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. *Molecular Ecology*, **19**, 5157–5171.

Salas DE, Stevens J, Schulz K *et al.* (2010) *Vegetation Classification and Mapping Project Report: Great Sand Dunes National Park and Preserve*. Natural Resource Technical Report NPS/ROMN/NRR—2010/179. National Parks Service, Fort Collins, Colorado.

Sambatti JBM, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution*, **60**, 696–710.

Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (Ed. Nurminsky D), pp. 55–64. Landes Bioscience, Georgetown, TX.

Sherwin WB, Jabot F, Rush R, Rossetto M (2006) Measurement of biological information with applications from genes to landscapes. *Molecular Ecology*, **15**, 2857–2869.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner KIM (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.

Tobler M, Riesch R, Tobler CM, Schulz-Mirbach T, Plath M (2009) Natural and sexual selection against immigrants maintains differentiation among micro-allopatric populations. *Journal of Evolutionary Biology*, **22**, 2298–2304.

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

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–1893.

Data accessibility

Microsatellite data and environmental variables: Dryad, entry doi: 10.5061/dryad.158pb518.

Supporting information

Table S1 Vegetation map units (Salas *et al.* 2010) and reclassification for the present study (e.g. Fig. 3).

Table S2 Analysis of variance summary for soil nutrient and quadrat data, which were averaged prior to analysis.

Table S3 Summary statistics for 16 *Helianthus* microsatellite loci and analysis of molecular variance (AMOVA) results.

Table S4 Locus summary statistics averaged across sunflower subpopulations, grouped by habitat type.

Table S5 Multiple regression of distance matrices (MRDM) to identify the optimal ratio of landscape conductance between barren sand and vegetated sunflower habitat.

Table S6 Multiple regression of distance matrices (MRDM) to identify environmental variables that are associated with genetic distance.

Fig. S1 Comparison of variables among habitat types.

Fig. S2 Summary of Structure analysis of microsatellite data.

Fig. S3 Non-hierarchial clustering of individuals.

Fig. S4 Mean likelihood of the data given K clusters within (a) dune and (b) non-dune samples, with 10 replicates at each value of K.

Fig. S5 Subpopulation estimates of Shannon's diversity index (Ha) negatively related to the proportion of dune ancestry, *Q*, (a) and positively related to the distance outside the dunes (b).

Fig. S6 Summary of individual-level spatial autocorrelation analysis with 1000 m distance classes.

Fig. S7 Genetic distances between pairs of populations in each habitat type.

Fig. S8 Relationship of genetic distance matrix with geographic distance, landscape distance and with differences in soil nitrogen and vegetation cover.