Molecular Ecology (2008) **17**, 5378–5390 doi: 10.1111/j.1365-294X.2008.04004.x

Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet

C. M. HERRERA and P. BAZAGA
Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas (CSIC), Avenida de María Luisa s/n, E-41013 Sevilla, Spain

Abstract

Local adaptation to contrasting biotic or abiotic environments is an important evolutionary step that presumably precedes floral diversification at the species level, yet few studies have demonstrated the adaptive nature of intraspecific floral divergence in wild plant populations. We combine a population-genomic approach with phenotypic information on floral traits to examine whether the differentiation in metric floral traits exhibited by 14 populations of the southern Spanish hawk moth-pollinated violet *Viola cazorlensis* reflects adaptive divergence. Screening of many amplified fragment length polymorphism (AFLP) loci using a multiple-marker-based neutrality test identified nine outlier loci (2.6% of the total) that departed from neutral expectations and were potentially under selection. Generalized analysis of molecular variance revealed significant relationships between genetic distance and population divergence in three floral traits when genetic distance was based on outlier loci, but not when it was based on neutral ones. Population means of floral traits were closely correlated with population scores on the first principal coordinate axis of the genetic distance matrix using outlier loci, and with the allelic frequencies of four of the outlier loci. Results strongly support the adaptive nature of intraspecific floral divergence exhibited by *V. cazorlensis* and illustrate the potential of genome scans to identify instances of adaptive divergence when used in combination with phenotypic information.

Keywords: adaptive divergence, AFLP, floral differentiation, gene flow, genome scan, population genomics

Received 22 May 2008; revision received 28 September 2008; accepted 8 October 2008

Introduction

Local adaptation to contrasting environments, both biotic and abiotic, is an important evolutionary step that presumably precedes floral diversification at the species level (Galen 2000; Patterson & Givnish 2004; Johnson 2006). Studies of intraspecific geographical differentiation in floral features and its potential adaptive nature are therefore crucial for understanding the linkage between the micro- and macroevolution of floral traits (Herrera *et al*. 2006; Hodgins & Barrett 2008), yet few investigations have so far provided reliable evidence for adaptive intraspecific differentiation in floral traits (review in Herrera *et al*. 2006). This gap in our knowledge probably reflects the challenging difficulties associated with proving that floral divergence among conspecific populations of wild plants is adaptive, i.e. is the outcome of variable selection acting on genetically determined floral traits (e.g. Schemske & Bierzychudek 2007). This applies particularly to species with long generation times, for which common-garden experiments and quantitative genetics approaches using experimental crossing schemes are prohibitive because of the time needed. Recent technological and analytical advances, however, provide unprecedented opportunities to circumvent the difficulties posed by long-lived species for testing hypotheses on adaptive floral divergence.

Genomic approaches can be used to identify functional genetic polymorphisms that are related to phenotypic traits of interest (Storz 2005; Vasemägi & Primmer 2005). Among the various research strategies available for dissecting functionally important genetic variation, the 'population-genomic approach' (Black *et al*. 2001) is particularly suitable for the analysis of adaptation processes in natural populations of non-model organisms (Vasemägi & Primmer 2005). By genotyping large numbers of markers in individuals

Correspondence: Carlos M. Herrera, Fax: +34 954 62 11 25;
E-mail: herrera@cica.es

taken from one or more populations, it is possible to identify genomic regions that exhibit deviant patterns of variation relative to the rest of the genome, due to the effects of selection (Black *et al*. 2001; Luikart *et al*. 2003). The population-genomic approach is a 'bottom-up' strategy, whereby the genetic basis of adaptation can be considered directly at the genome level without prior knowledge about the phenotypes or the selectively advantageous genes (Vasemägi & Primmer 2005). In the present study on adaptive floral divergence of the southern Spanish, endemic, hawk moth-pollinated violet *Viola cazorlensis*, we complement the population-genomic approach with information on population means for metric floral traits, some of which have been previously shown to be subject to selection (Herrera 1990a, 1993). In this way, we illustrate the powerful analytical possibilities afforded by genome scans when used in combination with associated phenotypic data. In the next section, we outline the steps followed to address the focal question of this paper, namely whether intraspecific, among-population variation in mean floral traits in *V. cazorlensis* reflects essentially random, non-adaptive phenotypic variation as previously hypothesized by Herrera (1990b) or conversely, whether it most likely reflects adaptive floral divergence mediated by selection. To our knowledge, this is the first study adopting a population-genomic approach to investigate the adaptive nature of floral divergence among populations of an animal-pollinated plant in the wild, and also one of the few attempts so far at elucidating adaptive divergence in plant populations by combining a genome scan with information on population differences in phenotypic traits.

General procedure

We adopt here the following analytical sequence: (i) a genome scan will be performed to distinguish putative loci subject to selection from neutral ones, based on typing individual plants from distinct populations using a large number of amplified fragment length polymorphism (AFLP) markers; (ii) on the basis of results obtained in (i), loci will be split into two non-overlapping subsets, 'selected' and 'neutral' ones, and separate matrices of pairwise population genetic distances will be obtained for each group; and (iii) the generalized analysis of molecular variance (GAMOVA) method proposed by Nievergelt *et al*. (2007) will be applied to explore the relationship across populations between divergence in mean floral phenotype and divergence in genetic features as measured with genetic distance. Separate GAMOVA analyses will be conducted for the selected and neutral loci subsets. Support for adaptive floral divergence will be obtained if some significant relationship across populations is found between phenotypic differences in floral traits and genetic distances based on selected loci, but not between differences in floral traits and genetic distances based on neutral loci. In case the GAMOVA analysis should reveal some overall association between genetic background based on outlier loci and population trait means, then an additional step (iv) will examine genotype-phenotype relationships across populations using 'band-based' and 'allele-frequency-based' AFLP analytical strategies (Bonin *et al*. 2007).

In the first step, a population-genomic approach will be used to distinguish putative loci subject to selection from neutral ones (Luikart *et al*. 2003). Selection perturbs patterns of genetic variation relative to those expected under a standard neutral model. While all the loci across the genome are expected to respond similarly to demography and neutral history of populations, including genetic drift and gene dispersal, a few among them will exhibit the signature of adaptive genetic differentiation caused by selection. Scanning the patterns of DNA polymorphisms at the genomic level through genotyping numerous random loci spread over the entire genome of individuals in several populations, will thus enable to identify 'outlier' loci that behave differently from the rest, presumably because of direct or indirect (through linkage) selection. When a locus shows extraordinary levels of genetic differentiation across populations in comparison with other loci, this can be interpreted as evidence of positive selection (Beaumont & Nichols 1996; Vitalis *et al*. 2001; Luikart *et al*. 2003; Bonin *et al*. 2006). Because AFLP genotyping allows for both a large number of markers and an accurate assessment of baseline levels of neutral genetic variation across the whole genome, AFLP markers are particularly suitable to the population genomics approach (Bonin *et al*. 2007; Meudt & Clarke 2007).

Materials and methods

Study species

Viola cazorlensis (Violaceae) is a perennial, suffruticose violet endemic to a few contiguous limestone mountain ranges in southeastern Spain, where it occurs in small, discrete populations associated with rocky outcrops, cliffs and 'islands' of sandy soils originating from heavily weathered dolomitic limestone between 900 and 2100 m elevation (Herrera 1993). In the Sierra de Cazorla, where the present study was conducted (see below), the species commonly occurs in widely disjunct, discrete, local populations varying in size from a few dozen to, quite rarely, a few thousand reproductive individuals, separated by several km of unsuitable habitat (various types of coniferous and mixed forest) (Herrera 1993). The flowers have pinkish-purple corollas and are characterized by a long and thin spur (mean length = 25.0 mm, range 8–42 mm; Herrera 1993), the longest for all European species in its genus. At all populations investigated in the Sierra de Cazorla over

1987–2007 by the first author, flowers are pollinated by a single species of long-tongued (mean proboscis length = 26.4 mm), day-flying hawk moth (*Macroglossum stellatarum*, Sphingidae) (Herrera 1993). Although flowers are self-compatible and may occasionally produce some fruits in absence of any pollinators, the activity of pollinators is essential for fruit to set (Herrera 1990a). The plants lack vegetative multiplication, and all reproduction takes place by seed. The seeds lack special dispersal mechanisms and are myxospermous. When moistened, the seeds develop a mucilaginous, sticky coat which after drying glues them to the substrate and precludes further movement. Detailed information on the ecology of *V. cazorlensis* can be found in Herrera (1990a, b, 1993).

Study area and methods

Floral characteristics were studied in the spring of 1988 in 18 populations of *V. cazorlensis* from the Sierra de Cazorla (Jaén province, southeastern Spain). Populations were distributed over an area of approximately 16 × 30 km, at elevations between 850–1860 m above sea level. Thirty flowers were collected from each population, taking a single flower from each of 30 randomly chosen plants. A detailed analysis of patterns of among-population variation in morphometric floral traits based on this sampling was presented by Herrera (1990b), where additional details on methods can be found. For the purposes of the present study, we considered among-population variation in the length of the floral peduncle, floral spur (straightened), and the upper, middle and lower petal blades (see Fig. 1 in Herrera 1990b for details on these measurements), using the populations means for these characters reported by Herrera (1990b). *V. cazorlensis* flowers have two long filiform nectaries originating from the posterior staminal filaments and extending back into the spur. The length of the longest nectary was also included in the floral morphometric traits considered here. Means of all these floral traits were shown by Herrera (1990b) to differ significantly among populations.

The 18 localities studied by Herrera (1990b) in 1988 were revisited in June 2006 for collecting leaf material for genetic analyses. We were unable to locate again four of the original sampling sites. At each of the remaining 14 sites, fresh leaf material was collected from 12–15 reproductive plants. *V. cazorlensis* is an endangered, legally protected species. Since removal of vegetative parts may have negative effects on the subsequent reproductive success of individuals, we sampled a relatively small number of individuals at each site (mean ± SD = 13.4 ± 2.2 plants/site) to minimize the impact of collections on local populations. Leaves were placed in small paper envelopes and dried immediately at ambient temperature in sealed containers with abundant silica gel. *V. cazorlensis* plants are very long-lived, and populations are characterized by slow adult plant turnover. At one of the sites sampled in 1988 that has been monitored annually since as part of other studies, 80% of a total of 75 reproductive individuals tagged in 1988 were still alive in 2006. We are confident that, at each site, leaf samples collected in 2006 for genetic analyses came from approximately the same individuals sampled in 1988 for the study of floral morphometry.

Table 1 Primer combinations used, number of markers (loci) obtained in the size range 150–500 base pairs, and observed polymorphism level, in the amplified fragment length polymorphism (AFLP) analysis of the 14 *Viola cazorlensis* populations considered in this study

Primer combination	Number of markers	% polymorphic*
1. *Eco*RI-AGA/*Mse*I-CTT	74	94.6
2. *Eco*RI-AGC/*Mse*I-CTC	47	91.5
3. *Eco*RI-AGG/*Mse*I-CTG	41	90.2
4. *Eco*RI-ACT/*Mse*I-CTA	38	92.1
5. *Eco*RI-AGG/*Mse*I-CAT	36	91.7
6. *Eco*RI-AGC/*Mse*I-CTT	45	97.8
7. *Eco*RI-AGG/*Mse*I-CAA	44	97.7
8. *Eco*RI-ACA/*Mse*I-CTT	44	95.5

*A locus was considered polymorphic if at least one individual in the sample showed a variant score.

Genomic DNA extraction and analysis

Dried leaf material was homogenized to a fine powder using a Retsch MM 200 mill. Total genomic DNA was then extracted from approximately 35 mg of ground leaf material using DNeasy Plant Mini Kit (Qiagen), following the manufacturer's protocol. DNA concentration of extracts was estimated by running electrophoreses of 5 μL aliquots on 0.8% agarose gels. The AFLP analysis was performed essentially as originally described by Vos *et al*. (1995), with modifications involving the use of fluorescent dye-labelled selective primers following Applied Biosystems (2005). Restriction-ligation was conducted using *Eco*RI/*Mse*I endonuclease mixture and double-stranded adaptors. A total of 12 *Eco*RI + 3/*Mse*I + 3 primer pairs were first screened for selective amplification in a pilot study conducted on a random subsample of 15 individuals from three widely spaced populations. The eight primer pairs providing the most reliable, consistently scorable results were finally chosen (Table 1). Each individual plant sampled was fingerprinted using these eight combinations. Fragment separation and detection was made using an ABI PRISM 3100 DNA sequencer. The presence or absence of each marker in each individual plant was scored manually by visualizing electrophoregrams with GeneMapper 3.7 software.

Several precautions were taken to avoid the appearance of spurious patterns. Samples were processed haphazardly

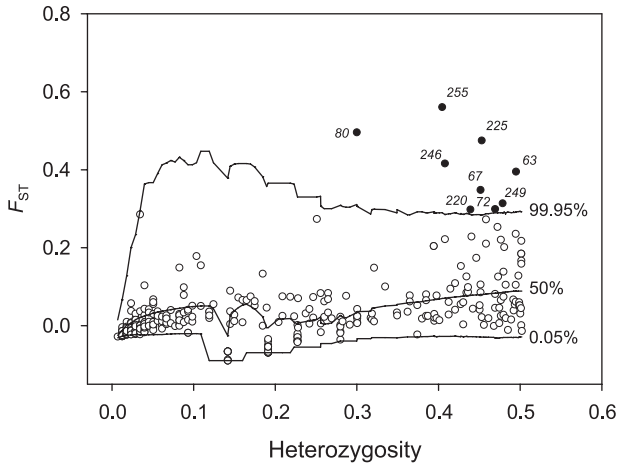

Fig. 1 Plot of F_{ST} values against heterozygosity estimates generated with DFDIST. Each point corresponds to an AFLP marker ($N = 341$). The three lines represent the 0.05, 50% (median) and 99.95% percentiles of the simulated distribution of neutral expectations (with $F_{ST} = 0.0719$) based on 10^6 realizations. The nine outlier loci at a significance level of $P < 0.0005$ are shown as filled dots, and associated numerals denote the corresponding locus number. Loci 63 through 80 were generated from primer Combination 2, and loci 220 through 255 from primer Combination 6 (see Table 1 for primer combinations).

with regard to population of origin. All scoring was done 'blindly' by the same person (PB), who during the process lacked any information on the populations of origin of individual samples, including geographical location, ecological characteristics of the site, or average floral features. Although we did not formally evaluate error rates, we are confident that our protocol should lead to scoring errors being randomly distributed across individuals and populations, which renders it extremely unlikely that the significant associations found in this study [i.e., linking AFLP profiles (both band- and allele-frequency-based), geographical provenance and mean floral traits] represent spurious artefacts derived from systematic and/or unintentional human biases arising during sample processing and scoring. Furthermore, since scoring errors are expected to add just random noise to our data, their effects would fall on the conservative side, tending to obscure underlying biological patterns. Fragment-size homoplasy commonly associated with AFLP markers (Vekemans *et al*. 2002) can introduce biases in estimates of population-genetic differentiation and the detection of loci subject to selection (Caballero *et al*. 2008). Since the magnitude of these biases is expected to be greatest in the small fragment-size classes (Caballero *et al*. 2008), we ignored markers smaller than 150 bp.

Data analyses

Outlier loci were identified in the AFLP data set using the program DFDIST (available at http://www.rubic.rdg.ac.uk/~mab/stuff/), which is based on the principle that genetic differentiation between populations is expected to be higher for loci under divergent selection than for the rest of the genome (Beaumont & Nichols 1996; Beaumont & Balding 2004). The version of DFDIST used (30 January 2008) analyses dominant markers and implements the Bayesian method of Zhivotovsky (1999) to estimate allelic frequencies from the proportion of recessive phenotypes (AFLP marker absent). A global analysis was done for all the populations sampled using 14 demes. Outlier loci were identified by plotting F_{ST} against heterozygosity under the assumption of Hardy–Weinberg equilibrium. Significance values were obtained by generating a null distribution of F_{ST} values based on 10^6 simulated loci, with a mean F_{ST} similar to the trimmed mean F_{ST} calculated from the empirical distribution. The trimmed mean F_{ST} was computed by removing the 25% highest and lowest F_{ST} values observed in the empirical dataset and is used as an estimate of the average 'neutral' F_{ST} uninfluenced by outlier loci (Beaumont & Nichols 1996; Bonin *et al*. 2006).

The performance of the DFDIST method to identify outlier loci depends decisively on choosing the 'correct' baseline F_{ST}, since this influences the quantiles of the simulated null distribution of F_{ST} conditional on heterozygosity, and consequently the likelihood of identifying a given locus (i.e. an empirical F_{ST} value) as an outlier (Caballero *et al*. 2008). In addition, overall F_{ST} values calculated by DFDIST can exaggerate the differentiation associated with one or a few populations, thus inflating the level of population differentiation for a given locus and artificially increasing its chances of being taken as an outlier. To cross-check the reliability of the outlier loci identified with DFDIST, we computed a separate set of F_{ST} estimates using the hierarchical Bayesian procedure of Beaumont & Balding (2004; see also Balding 2003) as implemented in the BAYESFST software (available at http://www.reading.ac.uk/statistics/genetics/software.html). This method adopts a regression approach and combines information over loci and populations in order to simultaneously estimate F_{ST} at every locus in each population. We simulated 2500 values from the posterior distribution of F_{ST} using the full regression model (i.e. including the locus, population, and locus × population effects) and the program's default settings. From these simulated values, we obtained an average F_{ST} and its associated '*P*-value' for every locus, as defined by the proportion of simulated outputs yielding locus effects < 0 (see Beaumont & Balding 2004 for details).

Another potential shortcoming of the DFDIST approach for identifying outlier loci is the possibility of obtaining an unknown number of false positives (i.e. committing type-I errors; Luikart *et al*. 2003; Bonin *et al*. 2007). This risk, however, can be considerably reduced by using a sufficiently conservative significance level (Beaumont & Balding

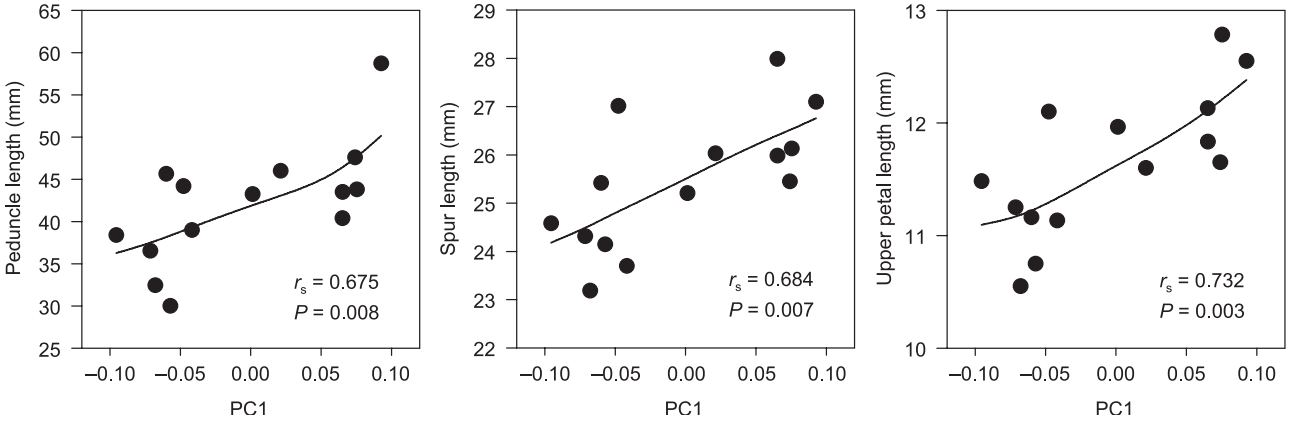

Fig. 2 Relationships between the population means of the three floral traits revealed by GAMOVA to be significantly related to differences in genetic divergence measured using outlier loci (Table 3), and the population scores on the first axis (PC1) from a principal coordinates analysis of the F_{ST} matrix for outlier loci. Each point corresponds to a population. Lines are nonparametric regressions (cubic splines) fitted to the data. Spearman rank correlation coefficients (r_s) and significance levels (P) are also shown.

2004; Caballero *et al*. 2008), as done here by adopting an $\alpha = 0.0005$. To corroborate this presumption, the likelihood of our results being 'contaminated' by some false-positive outlier loci was examined directly by applying Storey & Tibshirani's (2003) q-value method for estimation of false discovery rates to the set of P-values for individual loci provided by DFDIST. In the present context, the q-value obtained for a given locus is the expected proportion of false positives incurred for the whole dataset when that particular locus is considered as an outlier. The expected proportion of false positives in our results was obtained by calculating the q-values for all loci and thresholding them at $P = 0.0005$. The QVALUE package was used for computations (Storey & Tibshirani 2003).

F_{ST} matrices of pairwise genetic distances between populations were computed separately for outlier and neutral loci with the program AFLPSURV 1.0 (Vekemans *et al*. 2002), using Zhivotovsky's (1999) Bayesian method with non-uniform prior distribution of allele frequencies and assuming Hardy–Weinberg proportions. This method provides estimates of differentiation close to the real value even for moderate-to-small samples per population, as those used here (Bonin *et al*. 2007: Fig. 2 therein). Each of the F_{ST} matrices was then regressed on the population means for the six metric floral traits considered, using the GAMOVA method proposed by Nievergelt *et al*. (2007), which provides a flexible regression-based extension of the analysis of molecular variance (Excoffier *et al*. 1992). Gamova is particularly suited to identifying and characterizing the strength of relationships and genetic variations between populations and variables collected on the same populations (e.g. phenotypic traits, population-level features, environmental parameters). The strength of the relationship between genetic background and floral characters was explored by running matrix regressions for the six floral traits considered using the web-based GAMOVA tool available at http://polymorphism.scripps.edu/~cabney/cgi-bin/mmr.cgi (Nievergelt *et al*. 2007). Permutation tests based on 10 000 repetitions were used to assess the statistical significance of the relationships.

For those floral traits whose among-population variation was found significantly related to genetic background, the nature of the relationship was further explored using two complementary approaches based, respectively, on 'band-based' and 'allele-frequency-based' strategies (Bonin *et al*. 2007). The first method involved relating population character means to population mean scores on the first two axes from a principal coordinate analysis (PCA) of the genetic distance matrix based on outlier loci, as implemented in GENALEX 6.1 (Peakall & Smouse 2006). The second approach involved seeking significant relationships across populations between floral character means and allelic frequencies of the dominant allele of each outlier loci, as estimated with AFLPSURV.

Inferences on population genetic structure obtained from the subset of neutral loci alone will be more reliable than if the whole set of loci were used, since in this latter case even a few non-neutral outlier loci subject to selection can substantially bias estimates of genetic structuring and gene flow (Luikart *et al*. 2003). Patterns of genetic structuring were explored by applying the STRUCTURE 2.2 program to the subset of neutral loci. This program uses Bayesian methods to infer population structure and generate posterior probabilities of assignment of individuals to each of a given number of populations (K) without prior information on the number of locations at which the populations were sampled (Pritchard *et al*. 2000; Falush *et al*. 2003). Version 2.2 of STRUCTURE provides a full implementation of the admixture model with correlated allele frequencies (the 'F' model of Falush *et al*. 2003) for dominant markers (Falush

et al. 2007; Pritchard *et al*. 2007). We performed 25 replicates of each simulation with K = 1–14, with a burn-in of 30 000 and 70 000 post burn-in Markov chain Monte Carlo (MCMC) iterations, assuming admixture and correlated allele frequencies. The most likely number of populations represented in our sample (optimal K) was determined by examination of the estimates of model log likelihood [log $P(X|K)$, Ln $P(D)$ in the program output] for K = 1–14. As shown below, our results departed noticeably from the monotonically increasing relationship ordinarily relating Ln $P(D)$ and K, so the approach of Evanno *et al*. (2005) based on the rate of change of Ln $P(D)$ with increasing K could not be applied to our data.

Additional evidence on genetic structuring was obtained by examining the relationship between population pairwise F_{ST} computed for the set of neutral loci against geographical distance (Hutchison & Templeton 1999, and references therein). This analysis was mainly aimed at determining the relative importance of drift and gene flow in determining genetic structuring, rather than at elucidating spatial patterns of genetic variation. Untransformed F_{ST} and distance values were used in the analysis to allow a direct comparison between our results and the various theoretical scenarios proposed by Hutchison & Templeton (1999). Statistical significance of the F_{ST}-distance relationship was assessed using the Mantel test as implemented in GENALEX 6.1.

Results

Levels of polymorphism

A total of 369 AFLP fragments (loci) in the range 150–500 bp could be unambiguously scored from the eight primers combinations for the 188 plants sampled. There was little variation among primer combinations in both the number of scorable loci and levels of polymorphism, the latter being particularly high for all combinations (range = 90.2–97.8%, Table 1). Loci that were monomorphic for the entire data set (N = 17), and those that were present in < 1% or > 99% of individuals (N = 11), were excluded from all subsequent analyses. Seven plants that continued to produce noisy sequencer electrophoregrams for some primer combinations after running the analyses several times, were also excluded from the sample. Our final data set consisted of a total of 181 individuals, each scored for presence/absence of 341 polymorphic loci. All individuals exhibited unique AFLP profiles.

Outlier loci

The joint expected neutral distribution of F_{ST} against heterozygosity was simulated with DFDIST using 14 demes and a weighted mean of F_{ST} = 0.0719 over the 341 loci. Nine loci (2.6% of the total) fell outside the 99.95% confidence interval constructed from the simulated distribution (Fig. 1). All these outlier loci exhibited F_{ST} values significantly greater than expected, ranging 0.298–0.561 (Table 2; mean F_{ST} ± SE = 0.400 ± 0.031), and are therefore likely to be under directional selection or linked to a locus under selection. The remaining 332 loci, unlikely to be subject to selection, exhibited considerably smaller levels of differentiation (mean F_{ST} ± SE = 0.0267 ± 0.0035). All the outlier loci revealed by DFDIST were also highly significant outliers when the Bayesian regression method of Beaumont & Balding (2004) implemented in BAYESFST was applied to our dataset (Table 2). The F_{ST} values obtained by the two methods were very similar and linearly correlated (Table 2, r = 0.926, N = 9, P = 0.0004), which denotes that the high levels of population differentiation associated with the nine outlier loci were independent of computation method and underlying model.

Table 2 Comparison of estimates of F_{ST} and associated P-values obtained using DFDIST and the Bayesian regression method implemented in BAYESFST for the nine loci that were identified by DFDIST as being under selection (Fig. 1). Loci numbers as in Fig. 1

	DFDIST		BAYESFST	
Locus	F_{ST}	P-value	F_{ST}	P-value
63	0.395	< 10^{-6}	0.376	0.0002
67	0.348	< 10^{-4}	0.366	0.0013
72	0.299	0.0003	0.301	0.003
80	0.496	< 10^{-6}	0.396	0.0002
220	0.298	0.0003	0.286	0.0004
225	0.475	< 10^{-6}	0.467	0.0002
246	0.416	< 10^{-6}	0.368	0.0002
249	0.314	0.0002	0.286	0.0004
255	0.561	< 10^{-6}	0.582	0.0002

The maximum estimated q-value among all P-values less than or equal to α = 0.0005 was 0.011. This figure represents the expected probability of occurrence of false positives when all loci with P < 0.0005 are considered to be significant, as we did in Fig. 1 and the preceding paragraph. In view of this small q-value, we conclude that it is very unlikely that there is any false positive among the nine loci identified as outliers.

The nine outlier loci were non-randomly distributed among primer combinations. Six of the eight combinations used did not produce any outlier loci, while the other two combinations produced all the outlier loci detected. Combination 2 produced four outliers (loci 63–80, Fig. 1), or 9.3% of its polymorphic loci. Combination 6 produced five loci (loci 220–255, Fig. 1), or 11.4% of its polymorphic loci (see Table 1 for combination codes). The observed variation among primer combinations in the proportion of polymorphic loci that exhibited outlier behaviour was highly significant (P = 0.0005, Fisher's exact test).

Table 3 Results of the generalized analyses of molecular variance (GAMOVA) testing for relationships between population differences in mean values of floral traits and genetic distance between populations, conducted separately for the sets of neutral and outlier loci. Significant and nearly-significant *P*-values are shown in bold type

	Neutral loci (N = 332)			Outlier loci (N = 9)		
Floral trait (length)	Pseudo-F	Proportion of variance	P-value	Pseudo-F	Proportion of variance	P-value
Peduncle	1.79	0.13	0.16	5.45	0.31	**0.009**
Spur	1.64	0.12	0.19	5.57	0.32	**0.013**
Upper petal	1.46	0.11	0.25	7.14	0.37	**0.005**
Middle petal	1.15	0.09	0.37	2.44	0.17	0.111
Lower petal	1.23	0.09	0.33	3.23	0.21	**0.056**
Nectary	1.06	0.08	0.42	2.05	0.15	0.153

Table 4 Spearman-rank correlations (r_s) across populations (N = 14) between floral character means and the estimated frequencies of the dominant allele for each of the nine outlier loci. Loci numbers as in Fig. 1. Only the three floral traits shown by the GAMOVA to be significantly related to genetic divergence between populations in outlier loci were subjected to this analysis. Primer combination codes as in Table 1. Significant and nearly-significant correlation coefficients are shown in bold type

		Locus (Primer combination; AFLP marker size, bp)								
Floral trait (length)		63 (2; 157)	67 (2; 182)	72 (2; 219)	80 (2; 279)	220 (6; 205)	225 (6; 248)	246 (6; 318)	249 (6; 380)	255 (6; 497)
Peduncle	r_s	– 0.260	0.040	–0.339	0.103	**0.521**	**0.502**	**0.710**	0.371	**0.697**
	P-value	0.37	0.89	0.24	0.72	0.056	0.068	0.004	0.19	0.006
Spur	r_s	– 0.269	–0.176	–0.167	–0.035	**0.697**	**0.719**	**0.793**	0.424	**0.631**
	P-value	0.35	0.55	0.57	0.90	0.006	0.004	0.0007	0.13	0.016
Upper petal	r_s	– 0.363	–0.236	–0.284	–0.233	**0.692**	**0.625**	**0.714**	0.433	**0.666**
	P-value	0.20	0.42	0.32	0.42	0.006	0.017	0.004	0.12	0.009

Relationship between genetic and floral phenotypic divergence

Two separate F_{ST} distance matrices between populations were obtained using neutral and outlier loci, and GAMOVA matrix regressions were then run relating each of these matrices to the vectors of population means for the six floral traits considered. Results are summarized in Table 3. None of the regressions between mean floral traits and the genetic distance between populations based on neutral loci approached statistical significance. In contrast, three of the regressions of floral traits with genetic distance between populations based on outlier loci were statistically significant (length of peduncle, spur and upper petal), and a fourth (length of lower petal) was marginally significant.

Figure 2 depicts the relationships between population means of the three floral traits revealed by GAMOVA to be significantly related to genetic divergence in outlier loci (peduncle, spur and upper petal length) and population scores on the first axis from the principal coordinates analysis of the F_{ST} matrix for outlier loci (Principal component 1, PC1). The three relationships are statistically significant, with around 50% of the among-population variance in floral trait means being accounted for by the multilocus genetic background described by PC1. The three floral dimensions vary in unison with PC1, increasing linearly across populations with increasing PC1. No significant relationship was found between any population trait mean and population scores on the second axis from the PCA ($P > 0.65$ in all cases).

Results of the rank correlations relating variation among populations in mean peduncle, spur and upper petal length and estimated allelic frequencies for each of the nine outlier loci are shown in Table 4. There was a highly significant concordance between the three floral traits in their correlations with the allelic frequencies of the different outlier loci ($W = 0.952$, $P < 0.004$, Kendall's concordance test). Irrespective of the floral trait considered, trait-allele-frequency correlations were positive and statistically significant for loci 220, 225, 246 and 255, and statistically nonsignificant for loci 63, 67, 72, 80 and 249. As an example of the significant relationships between mean floral traits and allelic frequencies summarized in Table 4, Fig. 3 plots these relationships for locus 246.

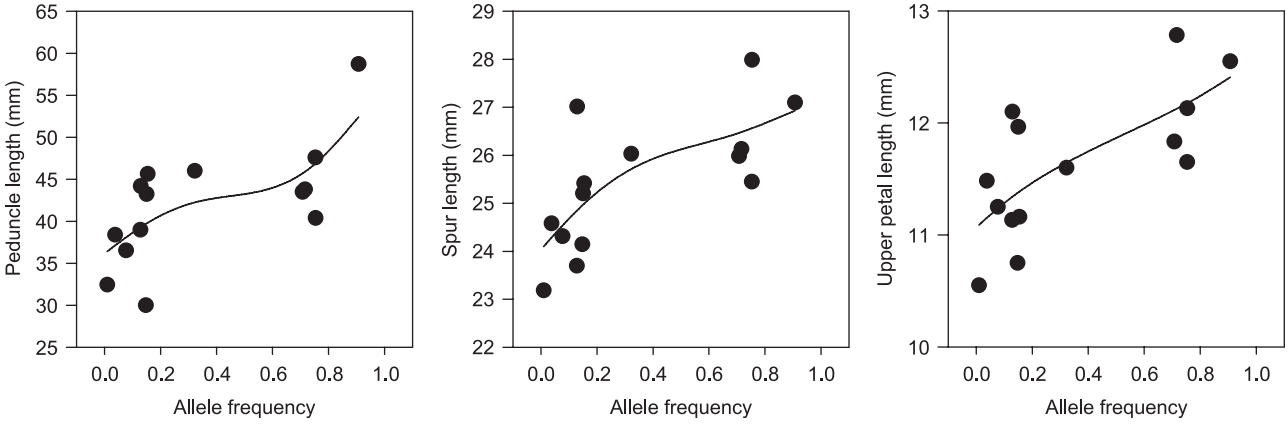

Fig. 3 Relationships between the population means of the three floral traits revealed by GAMOVA to be significantly related to differences in genetic divergence measured using outlier loci (Table 3), and the estimated population frequencies of the dominant allele of locus 246. Each point corresponds to a population. Lines are nonparametric regressions (cubic splines) fitted to the data. See Table 4 for significance levels of trait-allele frequency correlations.

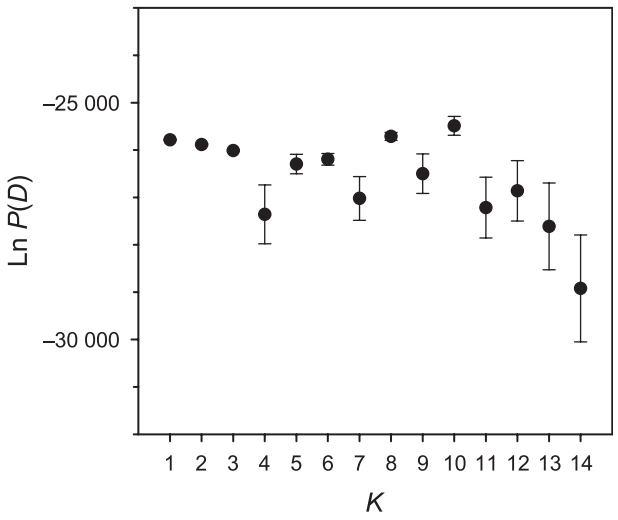

Fig. 4 Mean (±SE) model log likelihood (Ln $P(D)$) over 25 runs of STRUCTURE for each K-value from 1 to 14. Runs used a burn-in of 30 000 and 70 000 post burn-in Markov chain Monte Carlo (MCMC) iterations, assuming admixture and correlated allele frequencies.

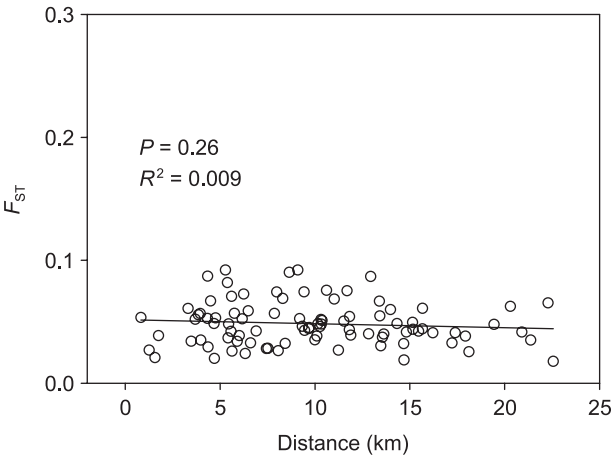

Fig. 5 Relationship between pairwise F_{ST} and geographical distance for the 14 populations of *Viola cazorlensis* studied. The line shown is the ordinary least-squares regression ($y = 0.051 - 0.00033x$). Significance level of the regression, as determined using a Mantel test with 10 000 permutations, and proportion of variance explained are also shown.

Genetic structure based on neutral loci

Genetic structure was investigated by applying STRUCTURE to the AFLP data matrix for the set of 332 neutral loci in the 181 individuals assayed. Results are summarized in Fig. 4 as a plot of mean model log likelihood (Ln $P(D)$) against K. The relationship is essentially flat up to $K = 10$, with Ln$P(D)$ subsequently decreasing up to $K = 14$. This variation pattern departs from the monotonic increase from $K = 1$ up to the optimal K that is theoretically expected, and is commonly found, when individuals are structured into distinct genetic groups (Pritchard *et al*. 2000, 2007; Evanno *et al*. 2005). Of particular interest is the decline of mean Ln$P(D)$ over the interval from $K = 1$ (–25 784) through $K = 2$ (–25 884) to $K = 3$ (–26 012). Following Falush *et al*. (2003) and Pritchard *et al*. (2007), and given that $K = 1$ is the smallest value of K that captures the major structure in our data, the most parsimonious explanation for the STRUCTURE results summarized in Fig. 4 is that there is a lack of genetic structuring (based on neutral loci) in the sample of individuals studied, despite their coming from 14 widely spaced locations.

In concordance with the previous finding, no statistically significant relationship was found at the spatial scale of this study between population pairwise F_{ST} based on neutral loci and geographical distance (Fig. 5). Pairwise F_{ST} values were similarly low over the whole range of between-site

distances represented in our sample (range = 0.8–22.6 km). The scatter of points around the fitted regression was also similarly narrow over the whole range of distances sampled (Fig. 5).

Discussion

Genetic diversity and structure

Plants with restricted distributions typically exhibit low genetic diversity at the species level (Cole 2003). In addition, fragmentation into isolated, discrete demes composed of relatively few individuals contributes to a depletion of overall genetic diversity (Ellstrand & Elam 1993). With fewer than 150 known local populations scattered over an area of ~750 km², and the vast majority of them consisting of < 100 adult plants (CMH, unpublished results), *Viola cazorlensis* is a perfect example of a rare plant with a small geographical range and heavily fragmented, small local populations. Fragmentation does not stem from anthropogenic disturbance, but it is instead the natural consequence of specialization on a scarce habitat type (isolated limestone outcrops and dolomitic sand patches), which is discontinuously distributed amid a sea of fairly undisturbed but ecologically unsuitable habitat (pine and oak forests). Given its ecological characteristics, it was unexpected to find that *V. cazorlensis* harbours so much genetic variation, as denoted by the extensive polymorphism exhibited by all AFLP markers (90–98%). Species-level percent polymorphism for AFLP markers is considerably higher in *V. cazorlensis* than in central European *Viola* species with fragmented populations but much broader geographical distributions (23–72%; Cieślak *et al*. 2006; Eckstein *et al*. 2006), and higher also than in other western Mediterranean narrow endemics (53–69%; Juan *et al*. 2004; Vilatersana *et al*. 2007). This apparent anomaly should be related to another unanticipated result of this study, namely the absence of spatial structuring of neutral genetic variation. As discussed below, the populations of *V. cazorlensis* studied apparently belong to a single panmictic unit and are interconnected by extensive gene flow. The effective size of the *V. cazorlensis* population is thus probably much greater than one would expect from exclusively considering the small size of local subpopulations. This would greatly reduce the incidence of drift, the main factor responsible for the loss of species-level genetic diversity in rare plants with small and fragmented populations (Ellstrand & Elam 1993).

When applied to simulated or real data with an underlying genetic structure, the STRUCTURE admixture model with correlated allele frequencies led to a monotonic increase of model log-likelihood from $K = 1$ up to some optimal K-value that represents the number of distinct genetic subpopulations in the sample of individuals (Falush *et al*. 2003; Evanno *et al*. 2005; Pritchard *et al*. 2007). Our results for neutral loci in *V. cazorlensis* depart noticeably from this pattern by showing a shallow decline of model log-likelihood (log $P(X \mid K)$) with increasing K. This pattern of variation of log $P(X \mid K)$ replicates that theoretically expected for a single panmictic population under the '*F*' model (Falush *et al*. 2003: Table 1 therein). Our STRUCTURE results should therefore be best interpreted as denoting that, despite their discreteness and spatial isolation, all local populations of *V. cazorlensis* sampled for this study actually behave as a single panmictic unit. This interpretation is further strengthened by the analysis of the relationship between population pairwise geographical and genetic distance based on neutral loci. At the spatial scale of this study, which roughly encompasses one-third of the whole distribution range of *V. cazorlensis*, we failed to find a significant relationship between population pairwise F_{ST} and geographical distance suggestive of isolation by distance. Instead, levels of genetic differentiation (based on neutral loci) between populations were independent of geographical distance and consistently low over the entire range of distances sampled, and there was not much scatter of the points around the fitted regression line. This pattern matches a Case II scenario in Hutchison & Templeton's (1999) theoretical classification, corresponding to a situation of lack of regional drift–gene flow equilibrium because of the overwhelming predominance of gene flow over drift.

Our results provide compelling evidence that populations of *V. cazorlensis* have been recently interconnected by extensive gene flow, a finding that disproves Herrera's (1990b) tentative suggestion of limited gene flow among populations of this species. The alternative explanation, that the species was formerly more abundant and continuously distributed in the region, and that its present rarity and fragmented distribution is the outcome of recent disturbances or habitat loss, is implausible, given the species' strict ecological specialization on a geomorphologically and edaphically defined habitat type whose abundance and distribution should vary little historically. Seed dispersal presumably contributes little to gene flow in *V. cazorlensis*, since seeds and fruits lack special mechanisms for dispersal and the myxospermy of seeds actually behaves as an anti-dispersal device (Grubert 1974). Extremely far-reaching pollen dispersal, in contrast, will most likely account for extensive gene flow. The hawk moth *Macroglossum stellatarum*, the only pollinator of *V. cazorlensis*, is a strong flier well-known for its long-distance migratory habits (Pittaway 1997–2008). Abundant migratory waves of *M. stellatarum* have been frequently observed in the Sierra de Cazorla, coincident with the time of *V. cazorlensis* flowering (CMH, unpublished observations 1987–2007). The abundance, wandering behaviour and strong flight capacity of migrating hawk-moth individuals can contribute to regularly dispersing the pollen of *V. cazorlensis* among populations widely apart, but proving this hypothesis will require further investigation.

Outlier loci

The general strengths and limitations of the population-genomic approach to identifying loci subject to selection, including those specific to AFLP markers, have been thoroughly considered in several recent reviews (Luikart *et al*. 2003; Beaumont & Balding 2004; Nielsen 2005; Storz 2005; Bonin *et al*. 2006, 2007; Caballero *et al*. 2008). We will not discuss our results in this context, except to remark on two methodological aspects that can affect the conclusions of genomic scans like the present one, namely the choice of an adequate statistical significance level and the use of a sufficient number of AFLP primer combinations.

In our study, the DFDIST analysis revealed nine AFLP loci (2.6% of the total investigated) departing from neutral expectations and thus potentially under selection. This proportion falls in the range reported by previous AFLP-based genomic scans (3.2%, Wilding *et al*. 2001; 8%, Scotti-Saintagne *et al*. 2004; 1.0%, Bonin *et al*. 2006; 3.2%, Smith *et al*. 2008). Nevertheless, the probability level set in our study to consider loci as outliers ($\alpha = 0.0005$) was considerably more restrictive than those used by these and other investigations ($\alpha = 0.05$ or 0.01; Caballero *et al*. 2008). Had we used the more liberal $\alpha = 0.01$ or $\alpha = 0.05$ in the DFDIST analysis, the proportion of loci with F_{ST} values significantly greater than neutral expectations would have risen to 17 (5.0%) and 24 (7.0%), respectively (results not shown). This may suggest that the proportion of non-neutral loci is actually greater in *V. cazorlensis* than in other species studied so far with AFLP markers, but also that earlier studies using less restrictive α's include an unknown fraction of false positives. In our study, the *q*-values associated with $\alpha = 0.01$ and $\alpha = 0.05$ are 0.199 and 0.596, respectively, which implies very high false discovery rates and a definite expectation of obtaining a non-trivial number of false positives (~3 and ~15, respectively), had a more liberal significance level been chosen. Our results for *V. cazorlensis* thus empirically confirm the conclusions of Caballero *et al*. (2008) based on simulations and show the critical importance of choosing sufficiently conservative significance levels for making the right inferences on the proportion of selective loci in the genome.

The nine loci that were most likely linked to genes under the influence of directional selection were non-randomly distributed among primer combinations. All were obtained from only two combinations (Combination 2 and 6), while the remaining six combinations produced none. Similar non-random distribution and clustering of outlier loci on particular AFLP-primer combinations have been found in other studies (Campbell & Bernatchez 2004; Scotti-Saintagne *et al*. 2004). The observation that AFLP loci from a given primer combination are often not distributed randomly among chromosomes (Peng *et al*. 2000; Strommer *et al*. 2002; Rogers *et al*. 2007) has led to the hypothesis that outlier loci can sometimes be clustered across a few linkage groups (Campbell & Bernatchez 2004). The concordant variation across localities of the allelic frequencies of outlier loci found in this study, as denoted by the concordance of their correlations with population means for floral trait, is compatible with that interpretation. The non-random distribution of outlier loci among primer combinations is important from a practical viewpoint (Campbell & Bernatchez 2004). If the likelihood of detecting loci showing the signature of selection in an AFLP-based genome scan is contingent on the primer combinations used, then it becomes particularly important to use as many different primer combinations as possible to obtain an unbiased sample of the genome and a reliable estimate of the frequency of selected loci. In a similar vein, the observation that AFLP markers obtained with different restriction enzymes differ in their patterns of distribution within and among chromosomes or linkage groups (Castiglioni *et al*. 1999; Scotti-Saintagne *et al*. 2004; Rogers *et al*. 2007) also suggest that more informative sets of outlier loci could be identified if AFLP-based genome scans were to use a broader variety of restriction enzymes.

Adaptive floral divergence

The finding of significant relationships linking variation in outlier loci and variation in environmental and/or phenotypic characteristics provides evidence that, despite its correlative nature, it is crucial to support *both* the non-neutral nature of the loci concerned and the adaptive nature of the environmental or phenotypic divergence with which their variation is associated (Wilding *et al*. 2001; Luikart *et al*. 2003; Campbell & Bernatchez 2004; Bonin *et al*. 2006). Our results prove particularly strong in this respect. For three metric floral traits, the GAMOVA analysis revealed significant relationships between multilocus population-genetic distance and divergence in phenotypic characteristics when genetic distance was based on outlier loci, but not when it was based on neutral ones. The disparity between the GAMOVA results for outlier and neutral loci is an important aspect of our results, and one lending particular support to the interpretation that population phenotypic divergence was the result of selection. If roughly similar correlation patterns between phenotypic and genetic distances had been obtained for both sets of markers, then factors other than selection could also account for them (e.g. variation in population history affecting equally to neutral and adaptive polymorphisms). The interpretation of adaptive floral divergence is further strengthened by the band-based and allele-frequency-based analyses of relationships across populations between mean phenotypes and genetic characteristics. Variation among populations in mean peduncle, spur and upper petal length were closely related to both the population scores on the

first axis of the PCA, and the allelic frequencies of five out of nine outlier loci. Taken together, these findings lead us to conclude that the differences among *V. cazorlensis* population in metric floral traits have a genetic basis and reflect local adaptations arisen through divergent selection. This conclusion disproves Herrera's (1990b) hypothesis that floral variation among these populations reflects random local differentiation due to genetic drift following the infrequent founding of populations by a few initial colonizers and subsequent restricted gene flow. The two hypothesized premises that prompted that suggestion, namely restricted gene flow between populations and a predominant role of drift in determining genetic structure and phenotypic variation, have also been disproved by this study. Recently, Schemske & Bierzychudek (2007) have similarly shown that spatial differentiation of flower colour in the desert plant *Linanthus parryae* most likely reflects the action of natural selection rather than genetic drift and restricted gene flow as originally envisaged by Wright (1943). In both *V. cazorlensis* and *L. parryae*, strong selective forces – and not genetic drift as previously hypothesized – seem responsible for maintaining intraspecific differentiation in floral traits in the face of extensive gene flow.

It is not possible at present to ascertain the selective agents and mechanisms of selection ultimately responsible for adaptive floral divergence in *V. cazorlensis*, or to establish whether the size of floral parts are actually the targets of selection. It must be noted, however, that pollinator-mediated selection on components of flower size and shape, including two of the traits shown here to exhibit adaptive divergence among populations (upper petal and peduncle length), has been demonstrated in one of our study populations (Herrera 1990a, 1993). Although all populations studied are similar in having the same pollinator, the strength of selection on particular floral traits and/or the shape of the trait-specific fitness functions can be geographically variable, as found for other species with relatively constant pollinators (Maad & Alexandersson 2004; Rey *et al*. 2006). Furthermore, and perhaps more importantly, other biotic and abiotic factors in addition to pollinators can exert direct or indirect selection on floral features (e.g. water stress, resources, herbivores; Galen 2000; Herrera 2000, 2005; Rey *et al*. 2006). Consistent between-population differences in the shape, sign or strength of selection exerted by one or more of these factors on metric floral traits or some pleiotropically linked characters could ultimately give rise to adaptive floral divergence despite the constancy in pollinator composition across populations. Studies are currently underway to examine these possibilities for *V. cazorlensis*.

Despite the remarkable lack of spatial genetic structuring and extensive levels of gene flow among populations of *V. cazorlensis* revealed by the analyses based on neutral loci, nine loci exhibited strong non-neutral signatures and were presumably selected or linked to selected regions of the genome. Our study joins an increasing number of investigations based on genome scans of plants and animals that demonstrate considerable within-genome heterogeneity in 'permeability' to gene flow: differentiation is maintained in a small portion of the genome, while extensive gene exchange continues to prevent divergence at most loci (Wilding *et al*. 2001; Campbell & Bernatchez 2004; Emelianov *et al*. 2004; Scotti-Saintagne *et al*. 2004; Minder & Widmer 2008; Smith *et al*. 2008). As discussed by these and other authors, within-genome patchiness in susceptibility to the homogenizing effect of gene flow has important implications, for example, for divergence-with-gene-flow models, the genic concept of speciation and the long-running debate on the relative importance of allopatric and parapatric modes of speciation (Wu & Ting 2004). When genome regions resistant to gene flow are associated with phenotypic traits that may directly influence mating system and sexual reproduction, as it happens with the floral features considered in this study, then initial adaptive divergence can facilitate the appearance of reproductive isolation, for instance via divergent shifts in pollinators (Bradshaw & Schemske 2003), thus setting the stage for incipient speciation.

Conservation implications

Apart from the implications in relation to speciation, within-genome heterogeneity in levels of population differentiation is also important because, for a given set of populations, neutral and adaptive genetic diversity can be organized very differently, and estimates of genetic differentiation based on neutral loci can differ widely from those obtained using selected loci – or those linked to selected loci (reviewed in Luikart *et al*. 2003). These are two critical aspects when conservation efforts of endangered species base their decisions on genetic data from sets of neutral genetic markers or on mixtures of markers containing an unknown but presumably overwhelming proportion of neutral ones. Our results for the endangered *V. cazorlensis* illustrate the potential perils involved very well. The F_{ST} estimates for either the whole set of 341 loci (0.072) or the 332 neutral loci alone (0.027) would lead to the deceptively reassuring conclusions: (i) that each local population harbours nearly all genetic diversity (~95%) of this endangered species; (ii) that a few local extinctions would have little impact on species-level genetic diversity; and (iii) that extinct populations could be easily restored by importing genetically similar individuals from other populations. Nevertheless, the F_{ST} estimate obtained from the nine outlier loci (0.400) provides quite a different picture: (i) only about 60% of total genetic diversity occurs within populations; (ii) a substantial amount of (presumably) adaptive genetic variation would be lost following even a few local population extinctions; and (iii) such extinctions would

imply the irreplaceable loss of locally-adapted genotypes. Similar contrasts between F_{ST} estimates with and without outlier loci seem to be widespread (Luikart *et al*. 2003), and local adaptation in the face of population-level genetic homogeneity assessed with neutral markers is known for other rare, endangered plants (McKay *et al*. 2001). Extraordinary caution should thus be exercised when genetic differentiation estimates based on neutral loci are used to inform critical conservation decisions. By affording at least a tentative dissection of genetic markers into neutral and (presumably) selected ones, and allowing the formulation of hypotheses on the levels of local adaptation of populations of endangered species, the population-genomic approach can also contribute to better inform management decisions in conservation biology.

Acknowledgements

We are grateful to Conchita Alonso, Antonio Castilla, José L. Garrido, Mónica Medrano, Clara de Vega and anonymous reviewers for helpful comments on the manuscript. Computer-intensive analyses reported in this paper were carried out at the TITAN and SVG computing clusters of the University of Oslo Bioportal and Centro de Supercomputación de Galicia (CESGA), respectively. Permission to work in the Sierra de Cazorla was provided by the Consejería de Medio Ambiente, Junta de Andalucía. This work was supported by grants 2005-RNM-156 (Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía) and CGL2006-01355 (Ministerio de Educación y Ciencia, Gobierno de España).

References

Applied Biosystems (2005) AFLP plant mapping protocol. Applied Biosystems, Foster City, CA.

Balding DJ (2003) Likelihood-based inference for genetic correlation coefficients. *Theoretical Population Biology*, **63**, 221–230.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619–1626.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.

Caballero A, Quesada H, Rolán-Alvarez E (2008) Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, **179**, 539–554.

Campbell D, Bernatchez L (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theoretical and Applied Genetics*, **99**, 425–431.

Cieślak E, Paul W, Ronikier M (2006) Low genetic diversity in the endangered population of *Viola uliginosa* in its locus classicus at Rzaska near Cracow (Southern Poland) as revealed by AFLP markers. *Acta Societatis Botanicorum Poloniae*, **75**, 245–251.

Cole CT (2003) Genetic variation in rare and common plants. *Annual Review of Ecology Evolution and Systematics*, **34**, 213–237.

Eckstein RL, O'Neill RA, Danihelka J, Otte A, Kohler W (2006) Genetic structure among and within peripheral and central populations of three endangered floodplain violets. *Molecular Ecology*, **15**, 2367–2379.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **271**, 97–105.

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.

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

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.

Galen C (2000) High and dry: drought stress, sex-allocation tradeoffs, and selection on flower size in the alpine wildflower *Polemonium viscosum* (Polemoniaceae). *American Naturalist*, **156**, 72–83.

Grubert M (1974) Studies on the distribution of myxospermy among seeds and fruits of Angiospermae and its ecological significance. *Acta Biologica Venezuelica*, **8**, 315–551.

Herrera CM (1990a) The adaptedness of the floral phenotype in a relict endemic, hawkmoth-pollinated violet. 1. Reproductive correlates of floral variation. *Biological Journal of the Linnean Society*, **40**, 263–274.

Herrera CM (1990b) The adaptedness of the floral phenotype in a relict endemic, hawkmoth-pollinated violet. 2. Patterns of variation among disjunct populations. *Biological Journal of the Linnean Society*, **40**, 275–291.

Herrera CM (1993) Selection on floral morphology and environmental determinants of fecundity in a hawk moth-pollinated violet. *Ecological Monographs*, **63**, 251–275.

Herrera CM (2000) Measuring the effects of pollinators and herbivores: evidence for non-additivity in a perennial herb. *Ecology*, **81**, 2170–2176.

Herrera CM, Castellanos MC, Medrano M (2006) Geographical context of floral evolution: towards an improved research programme in floral diversification. In: *Ecology and Evolution of Flowers* (eds Harder LD, Barrett SCH), pp. 278–294. Oxford University Press, Oxford.

Herrera J (2005) Flower size variation in *Rosmarinus officinalis*: individuals, populations and habitats. *Annals of Botany*, **95**, 431–437.

Hodgins KA, Barrett SCH (2008) Geographic variation in floral morphology and style-morph ratios in a sexually polymorphic daffodil. *American Journal of Botany*, **95**, 5–195.

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

Johnson SD (2006) Pollinator-driven speciation in plants. In: *Ecology and Evolution of Flowers* (eds Harder LD, Barrett SCH), pp. 295–310. Oxford University Press, Oxford.

Juan A, Crespo MB, Cowan RS, Lexer C, Fay MF (2004) Patterns of variability and gene flow in *Medicago citrina*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). *Molecular Ecology*, **13**, 2679–2690.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, **4**, 981–994.

Maad J, Alexandersson R (2004) Variable selection in *Platanthera bifolia* (Orchidaceae): phenotypic selection differed between sex functions in a drought year. *Journal of Evolutionary Biology*, **17**, 642–650.

McKay JK, Bishop JG, Lin JZ, Richards JH, Sala A, Mitchell-Olds T (2001) Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **268**, 1715–1721.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice ? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106–117.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552–1563.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nievergelt CM, Libiger O, Schork NJ (2007) Generalized analysis of molecular variance. *Plos Genetics* **3**, e51. doi: 10.1371/journal.pgen.0030051.

Patterson TB, Givnish TJ (2004) Geographic cohesion, chromosomal evolution, parallel adaptive radiations, and consequent floral adaptations in *Calochortus* (Calochortaceae): evidence from a cpDNA phylogeny. *New Phytologist*, **161**, 253–264.

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

Peng J, Korol AB, Fahima T *et al*. (2000) Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. *Genome Research*, **10**, 1509–1531.

Pittaway AR, (1997–2008) Sphingidae of the Western Palaearctic. Available at http://tpittaway.tripod.com/sphinx/list.htm.

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

Pritchard JK, Wen X, Falush D (2007) *Documentation for Structure Software*: Version 2.2. Department of Human Genetics, University of Chicago. Available at http://pritch.bsd.uchicago.edu/structure.html.

Rey PJ, Herrera CM, Guitián J *et al*. (2006) The geographic mosaic in pre-dispersal interactions and selection on *Helleborus foetidus* (Ranunculaceae). *Journal of Evolutionary Biology*, **19**, 21–34.

Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the *dwarf* and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics*, **175**, 375–398.

Schemske DW, Bierzychudek P (2007) Spatial differentiation for flower color in the desert annual *Linanthus parryae*: was Wright right? *Evolution*, **61**, 2528–2543.

Scotti-Saintagne C, Mariette S, Porth I *et al*. (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. & *Q. petraea* (Matt.) Liebl.]. *Genetics*, **168**, 1615–1626.

Smith TB, Milá B, Grether GF *et al*. (2008) Evolutionary consequences of human disturbance in a rainforest bird species from Central Africa. *Molecular Ecology*, **17**, 58–71.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, **100**, 9440–9445.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Strommer J, Peters J, Zethof J, de Keukeleire P, Gerats T (2002) AFLP maps of *Petunia hybrida*: building maps when markers cluster. *Theoretical and Applied Genetics*, **105**, 1000–1009.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Vilatersana R, Susanna A, Brochmann C (2007) Genetic variation in *Femeniasia* (Compositae, Cardueae), an endemic and endangered monotypic genus from the Balearic Islands (Spain). *Botanical Journal of the Linnean Society*, **153**, 97–107.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.

Vos P, Hogers R, Bleeker M *et al*. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Wright S (1943) An analysis of local variability of flower color in *Linanthus parryae*. *Genetics*, **28**, 139–156.

Wu CI, Ting CT (2004) Genes and speciation. *Nature Review Genetics*, **5**, 114–122.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.