Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*

Carlos M. Herrera and Pilar Bazaga

Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas (CSIC), Isla de La Cartuja, Avenida Américo Vespucio s/n, E–41092 Sevilla, Spain

Summary

- In plants, epigenetic variations based on DNA methylation are often heritable and could influence the course of evolution. Before this hypothesis can be assessed, fundamental questions about epigenetic variation remain to be addressed in a real-world context, including its magnitude, structuring within and among natural populations, and autonomy in relation to the genetic context.
- Extent and patterns of cytosine methylation, and the relationship to adaptive genetic divergence between populations, were investigated for wild populations of the southern Spanish violet *Viola cazorlensis* (Violaceae) using the methylation-sensitive amplified polymorphism (MSAP) technique, a modification of the amplified fragment length polymorphism method (AFLP) based on the differential sensitivity of isoschizomeric restriction enzymes to site-specific cytosine methylation.
- The genome of *V. cazorlensis* plants exhibited extensive levels of methylation, and methylation-based epigenetic variation was structured into distinct between- and within- population components. Epigenetic differentiation of populations was correlated with adaptive genetic divergence revealed by a Bayesian population-genomic analysis of AFLP data. Significant associations existed at the individual genome level between adaptive AFLP loci and the methylation state of methylation-susceptible MSAP loci.
- Population-specific, divergent patterns of correlated selection on epigenetic and genetic individual variation could account for the coordinated epigenetic-genetic adaptive population differentiation revealed by this study.

Introduction

Variation in nucleotide sequence (i.e. genuine genetic variation) is not the only information in chromosomes that is heritable across generations. Epigenetic information, which is based on variations in DNA methylation or chromatin states, often is also heritable during cell propagation, particularly in plants, where epigenetic states of genes that influence phenotypic traits can be inherited over generations (Henderson & Jacobsen, 2007; Jablonka & Raz, 2009; Verhoeven *et al.*, 2010). In theory, heritable epigenetic variation could directly or indirectly influence the course of evolution in plants, as it can affect the processes of adaptation and divergence through selection of stable epigenetic variants without involvement of genetic variation, or through evolutionary change in which epigenetic modifications would guide the selection of correlated genetic variation (Kalisz & Purugganan, 2004; Rapp & Wendel, 2005; Jablonka & Raz, 2009). Nevertheless, before the hypothesized role of epigenetic variation in plant microevolution can be assessed, fundamental questions remain to be addressed in a real-world context about its magnitude, structuring within and among natural plant populations, possible correlations with genetic or phenotypic traits related to fitness, and degree of autonomy in relation to the genetic context (Kalisz & Purugganan, 2004; Rapp & Wendel, 2005; Richards, 2006). DNA methylation is the best-described epigenetic mechanism, and it is involved in

nearly all well-documented instances of transgenerational epigenetic inheritance in plants (Akimoto *et al.*, 2007; Henderson & Jacobsen, 2007; Jablonka & Raz, 2009; Verhoeven *et al.*, 2010). Given also that relatively simple and cost-effective molecular methods are available to study patterns of genome-wide DNA methylation in nonmodel organisms lacking detailed genomic information (Reyna-López *et al.*, 1997; Xiong *et al.*, 1999), investigating natural patterns of DNA methylation in wild plant populations emerges as a first step towards assessing the potential significance of epigenetic variation in microevolution (Kalisz & Purugganan, 2004; Richards, 2006).

The molecular mechanisms and functions associated with DNA methylation are reasonably well understood for model plants (Finnegan *et al.*, 1998b; Vanyushin, 2006; Zhang, 2008), and a number of studies have also documented intraspecific epigenetic variation in model and cultivated species (Xiong *et al.*, 1999; Ashikawa, 2001; Cervera *et al.*, 2002; Keyte *et al.*, 2006; Vaughn *et al.*, 2007; Salmon *et al.*, 2008). Almost nothing is known, however, on the amount and structuring of standing methylation-based epigenetic variation in wild plant populations (but see Li *et al.*, 2008). Of particular relevance from an evolutionary perspective is to ascertain whether species-level epigenetic variation is, like sequence-based genetic variation, structured into distinct between- and within-population components, and whether such epigenetic structuring can be interpreted in adaptive terms, that is, as maintained by divergent selection (Kalisz & Purugganan, 2004; Bossdorf *et al.*, 2008). For example, under conditions of extensive gene flow connecting discrete populations of a species into a genetically coherent, panmictic unit, a significant correlation linking between-population epigenetic differentiation with adaptive genetic divergence would be suggestive of epigenetic differentiation being directly or indirectly driven by variable selection. In this paper we address these questions by studying the extent and pattern of cytosine methylation in a set of discrete populations of the southern Spanish endemic violet *Viola cazorlensis* (Violaceae), and explore the relationship to adaptive genetic divergence between populations.

We use the technique of methylation-sensitive amplified polymorphism (MSAP), a modification of the amplified fragment length polymorphism method (AFLP) which takes advantage of the differential sensitivity of a pair of iso-schizomeric restriction enzymes to site-specific cytosine methylation (McClelland *et al.*, 1994; Reyna-López *et al.*, 1997). By allowing the determination of the methylation status of anonymous regions of the genome susceptible to methylation, the MSAP method enables the identification of methylation-based epiallelic markers in wild populations of nonmodel plants in absence of detailed genomic information. Populations of *V. cazorlensis* are ideally suited to a study of patterns and possible adaptive correlates of

epigenetic variation. Genetic analyses of wild populations of this narrowly-distributed plant have shown that, despite their spatial discreteness, populations are interconnected by extensive gene flow, neutral genetic variation is not spatially structured across populations and there is a lack of regional drift–gene flow equilibrium because of an overwhelming predominance of gene flow over drift (Herrera & Bazaga, 2008). In the face of extensive gene flow, however, populations of *V. cazorlensis* exhibit distinctive signatures of adaptive genetic divergence (Herrera & Bazaga, 2008). These circumstances combine favourably to render *V. cazorlensis* an excellent study system to investigate possible adaptive epigenetic variation by searching for correlations between methylation-based epigenetic differentiation and adaptive genetic divergence between populations.

Materials and Methods

Study system

Viola cazorlensis (Violaceae) is a perennial, suffruticose violet endemic to a few contiguous limestone mountain ranges in south-eastern Spain, where it generally occurs as discrete populations varying in size from a few dozen to a few thousand reproductive individuals, separated by a few kilometres of unsuitable habitat (various types of coniferous and mixed forest; Herrera, 1993). The species is a strict habitat specialist, with populations restricted to rocky outcrops, cliffs and 'islands' of sandy soils in weathered dolomitic limestone. Plants have a woody rootstock, are very long-lived (> 25 yr), lack vegetative multiplication, and all reproduction takes place by seeds.

This study was conducted on the same 14 populations of *V. cazorlensis* from the Sierra de Cazorla, south-eastern Spain, considered in an earlier investigation on adaptive genetic divergence in this species (Herrera & Bazaga, 2008). The nearest and most distant populations were 1.3 km and 25 km away, respectively, and elevation ranged between 850 m and 1860 m. At each site, fully expanded fresh leaves were collected from 12 to 15 reproductive plants, placed in small paper envelopes and dried immediately at ambient temperature in sealed containers with abundant silica gel. Plants from all sites were sampled on approximately the same dates (range of sampling dates = 29 May–8 June 2006) and at the same phenological stage (bearing fully expanded leaves and some open flower), so that possible developmental variation in DNA methylation would not confound individual or population differences in methylation patterns. The 186 plants considered in this study included all the individuals considered in the earlier investigation on adaptive genetic divergence based on conventional AFLP markers (Herrera & Bazaga, 2008; *n* = 181 plants). For each plant, the same DNA sample was used as the starting material in both investigations, which was

obtained by extracting total genomic DNA from *c.* 35 mg of ground leaf material using DNeasy Plant Mini Kit (Qiagen) and the manufacturer's protocol.

Epigenetic analyses

Methylation-sensitive amplified polymorphism (MSAP) analysis was used to identify methylation-susceptible anonymous 5′-CCGG sequences and assess their methylation status. Methylation-sensitive amplified polymorphism is a modification of the standard AFLP technique that uses *Eco*RI as rare cutter and the methylation-sensitive restriction enzymes *Hpa*II and *Msp*I as frequent cutters, the latter being a pair of isoschizomers that recognize the same tetranucleotide 5′-CCGG but have differential sensitivity to methylation at the inner or outer cytosine (Reyna-López *et al.*, 1997). *Hpa*II is inactive if one or both cytosines are methylated at both DNA strands, but cleaves when one or both cytosines are methylated in only one strand. *Msp*I, by contrast, cleaves $C^{5m}CGG$ but not $^{5m}CCGG$ (McClelland *et al.*, 1994). Differences in the products obtained with *Eco*RI–*Hpa*II and *Eco*RI–*Msp*I should thus reflect different methylation states at the cytosines of the CCGG sites recognized by *Hpa*II or *Msp*I, which renders MSAP an efficient method for detecting alterations in cytosine methylation in plants (Cervera *et al.*, 2002; Keyte *et al.*, 2006).

The standard AFLP analyses corresponding to each of the six *Eco*RI–*Hpa*II and six *Eco*RI–*Msp*I primer combinations assayed (Table 1) were performed as in Herrera & Bazaga (2008). Fragment separation and detection was made using an ABI PRISM 3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA). The presence or absence of *Eco*RI–

*Hpa*II and *Eco*RI–*Msp*I fragments in each individual plant was scored manually by visualizing electropherograms with GENEMAPPER 3.7 software (Applied Biosystems). Several precautions were taken to avoid the appearance of spurious patterns. Samples were processed haphazardly with regard to population of origin. All scoring was blindly done by the same person (PB), who, during the process, lacked information on the populations of origin of individual samples, including geographical location and ecological characteristics of the site. Only fragments ≥ 150 bp in size were considered to reduce the potential impact of size homoplasy (Caballero *et al.*, 2008). Genotyping error rates were determined for each primer combination by running repeated, independent analyses for 12 individual plants, and estimated as the ratio of the total number of loci with discordant scores (all individuals combined) to the product of the number of individuals by the total number of scored loci (Herrera & Bazaga, 2009).

Data analysis

Analyses of MSAP results were based on element-wise comparisons of fragment presence–absence matrices for individual plants obtained with *Eco*RI–*Hpa*II and *Eco*RI–*Msp*I primer combinations. For every individual and particular fragment, it was first determined whether the fragment was: (1) present in both *Eco*RI–*Hpa*II and *Eco*RI–*Msp*I products; (2) absent from both *Eco*RI–*Hpa*II and *Eco*RI–*Msp*I products; or (3) present only in either *Eco*RI–*Hpa*II or *Eco*RI–*Msp*I products. Condition (1) denotes a nonmethylated state, condition (3) corresponds to a methylated state, and condition (2) is uninformative, as

Table 1 Primer combinations used, number of markers (loci) in the size range 150–500 bp, scoring error rates, and frequency and polymorphism of nonmethylated and methylation-susceptible markers, in the methylation-sensitive amplified polymorphism (MSAP) analysis of 186 individuals of *Viola cazorlensis* from 14 populations considered in this study

Primer combination	Number of AFLP markers	Scoring error rate (%)[1]		Estimated probability of erroneous *Hpa*II–*Msp*I mismatch[2]	Non-methylated markers		Methylation-susceptible markers	
		*Hpa*II	*Msp*I		*n*	% Polymorphic[3]	*n*	% Polymorphic[4]
1. *Eco*RI + AGA/*Hpa*II–*Msp*I + TG	63	3.84	1.85	0.0554	36	50.0	27	100.0
2. *Eco*RI + ACC/*Hpa*II–*Msp*I + TT	49	3.40	3.91	0.0704	25	76.0	24	91.7
3. *Eco*RI + AG/*Hpa*II–*Msp*I + TGA	46	2.90	5.07	0.0767	26	65.4	20	100.0
4. *Eco*RI + ACA/*Hpa*II–*Msp*I + TA	69	2.54	2.17	0.0459	21	66.7	48	97.9
5. *Eco*RI + ACT/*Hpa*II–*Msp*I + TC	41	3.05	5.28	0.0801	14	71.4	27	85.2
6. *Eco*RI + AC/*Hpa*II–*Msp*I + TAA	75	4.82	4.17	0.0858	40	67.5	35	94.3
All combined	343	3.51	3.56	0.0682	162	64.8	181	95.0

[1]Calculated from the 12 individuals that were reassayed as 100 × (number of discordant scores on two independent analyses)/(number of scored markers × number of individuals).
[2]Estimated average probability of obtaining discordant *Eco*RI–*Hpa*II and *Eco*RI–*Msp*I scores for an individual × locus combination owing to scoring error alone, estimated from the scoring error rates for *Eco*RI–*Hpa*II (= e_{Hpa}) and *Eco*RI–*Msp*II (= e_{Msp}) as $e_{Hpa} + e_{Msp} - 2e_{Hpa}e_{Msp}$.
[3]A nonmethylated marker was considered polymorphic if at least two individuals in the sample showed a variant score.
[4]A methylation-susceptible marker was considered polymorphic when both methylated and nonmethylated states occurred in the sample of individuals studied.
AFLP, amplified fragment length polymorphism method; *n*, number of markers.

it could be caused by either fragment absence or hyper-methylation (Xiong *et al.*, 1999; Ashikawa, 2001). As noted below, condition (2) was treated as a missing score. Individual fragments were classed as either 'methylation-susceptible' or 'nonmethylated', depending on whether the observed proportion of discordant *Hpa*II–*Msp*I scores suggestive of methylation (i.e. number of individuals with contrasting *Hpa*II–*Msp*I scores for the fragment divided by total number of individuals assayed) exceeded a given threshold. This threshold was specific for each primer combination and set equal to the expected per-individual probability of obtaining a mismatch of *Hpa*II and *Msp*I scores owing to errors associated with *Hpa*II scoring (= e_{Hpa}), *Msp*I scoring (= e_{Msp}), or both (= $2e_{Hpa}e_{Msp}$) (i.e. drawing a false inference of methylation), estimated as $e_{Hpa} + e_{Msp} - 2\,e_{Hpa}e_{Msp}$. The threshold varied among primer combinations, ranging from 0.046 to 0.086 (Table 1). Nonmethylated loci were scored as dominant binary markers, as usually done for AFLP markers (1 and 0, for fragment presence and absence, respectively). Instances of discordant *Hpa*II–*Msp*I scores in nonmethylated fragments were resolved as fragment presence. Methylation-susceptible fragments were scored as if the methylated state were an imperfectly assessed dominant marker: 1 for the methylated state (condition 3 above), 0 for the nonmethylated state (condition 1 above) and unknown (i.e. score missing) for uninformative condition 2 above.

A band-based strategy (*sensu* Bonin *et al.*, 2007) was adopted for the statistical analyses of MSAP results. Genetic and epigenetic diversity was assessed with the Shannon's diversity index. Single-locus and multilocus epigenetic population differentiation were tested using likelihood ratio χ^2 tests for population heterogeneity in methylation frequency, and analyses of molecular variance (AMOVA; Excoffier *et al.*, 1992), respectively. Given the large number of simultaneous χ^2 tests, we applied Storey & Tibshirani (2003) *q*-value method for estimating the false discovery rates of the set of *P*-values for individual fragments obtained from the battery of simultaneous likelihood ratio tests. Using the QVALUE package (Storey & Tibshirani, 2003), we calculated the *q*-values for all the tests, ranked them, and found the largest *q*-value leading to an expectation of ≤ 1 falsely significant between-population heterogeneity in methylation frequency (i.e. *q*-value × (number of tests accepted as significant) ≤ 1) (Herrera & Bazaga, 2009).

Adaptive genetic divergence was documented previously for the same *V. cazorlensis* populations and individuals studied here by means of a population genomic analysis of individual genotypes for 341 polymorphic conventional AFLP loci (Herrera & Bazaga, 2008). To take advantage of recent theoretical advances allowing for a more rigorous identification of candidate loci presumably subject to selection ('outlier loci'), the AFLP dataset used in that earlier work was reanalysed here using the Bayesian method proposed by Foll & Gaggiotti (2008) and implemented in the program BAYESCAN 1.0 (http://www-leca.ujf-grenoble.fr/logiciels.htm). Allelic frequencies of outlier loci at each population were obtained from the proportion of recessive phenotypes (AFLP marker absent) with the Bayesian method of Zhivotovsky (1999). These allelic frequencies were used as predictor variables in a distance-based redundancy analysis (dbRDA; Anderson, 2003) testing for the relationship between epigenetic differentiation and adaptive genetic divergence of populations. For this analysis, a population-by-population matrix of epigenetic distances was obtained using the AMOVA-based population differentiation parameter Φ_{ST} (Excoffier *et al.*, 1992). Conditional (sequential) tests of the effect of population differentiation in allelic frequencies of outlier loci on epigenetic population differentiation were then performed with the program DISTLM-*forward* (Anderson, 2003), and statistical significance was determined using 10^5 permutations. Two analyses were performed to verify the biological basis of the statistical relationship between epigenetic and adaptive genetic population differentiation revealed by the population-level dbRDA. First, an individual-based dbRDA was conducted on the 181 individual plants with epigenetic and genetic data simultaneously available. An individual-by-individual epigenetic distance matrix was computed using a simple-matching coefficient, in which any comparison with the same state yields a value of zero while any comparison of different states yields a value of unity. Binary scores for those AFLP outlier loci whose allelic frequencies had been found to be significantly related to epigenetic population differentiation were included in the individual-based dbRDA analysis as predictor variables. Second, likelihood ratio χ^2 tests were used to test for significant associations across individuals between presence/absence of each AFLP outlier locus and the methylation state of each of the methylation-susceptible loci whose methylation frequencies varied significantly among populations. Statistical significance of associations between binary states of epigenetic (methylated/nonmethylated) and adaptive genetic (present/absent) loci was assessed using the *q*-value method as described earlier.

Results

Methylation-susceptible AFLP loci

The six primer combinations assayed in the MSAP analysis produced a total of 343 AFLP fragments ('loci' hereafter) that could be unambiguously scored for the 186 plants of *V. cazorlensis* from the 14 populations sampled (Table 1). Each locus was classed as either 'methylation-susceptible' or 'nonmethylated', depending on whether the proportion of discordant scores obtained with the *Hpa*II and *Msp*I isoschizomers (i.e. number of individuals with contrasting

HpaII–MspI scores divided by total number of individuals assayed) exceeded the expected per-individual probability of obtaining a mismatch of HpaII and MspI scores caused solely by scoring errors (i.e. drawing a false inference of methylation; Table 1). The proportion of HpaII–MspI discordances was lower than or equal to the corresponding combination-specific threshold for 162 loci (nonmethylated loci hereafter, 47.2% of total) and exceeded the threshold for 181 loci (methylation-susceptible loci hereafter, 52.8% of total). About half of the loci from the MSAP analysis were therefore in a methylated state in a significant fraction of the individuals sampled.

Genetic and epigenetic diversity

For all primer combinations combined, 64.8% of the 162 nonmethylated loci from MSAP were polymorphic in the set of individuals sampled (Table 1). Polymorphism level was substantially higher for the 181 methylation-susceptible loci, 95% of which exhibited variable methylation states in the individuals studied (Table 1). The frequency distributions of Shannon's diversity index (I) for nonmethylated and methylation-susceptible polymorphic loci differed markedly in central trend (Fig. 1). Mean I for methylation-susceptible polymorphic loci (Mean ± SE = 0.456 ± 0.013, n = 172) was significantly higher than the corresponding figure for nonmethylated ones (0.204 ± 0.016, n = 105) (χ^2 = 99.4, df = 1, $P < 0.0001$, Wilcoxon rank-sum test). These results reveal extensive methylation-based, epigenetic diversity in the sample of $V.$ $cazorlensis$ individuals studied, and that methylation-based epigenetic diversity exceeded conventional genetic diversity when the two magnitudes are compared using the same index.
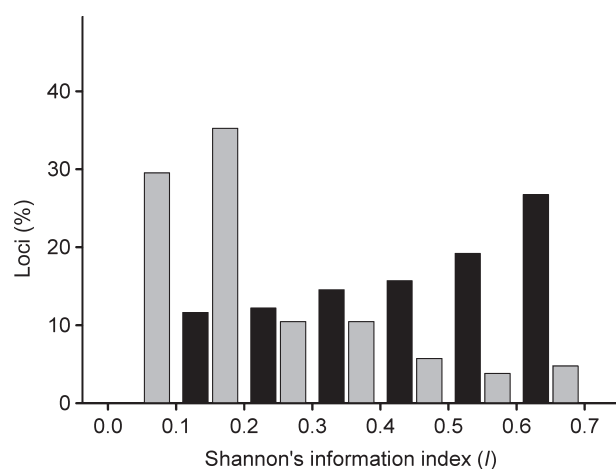

Fig. 1 Comparative diversity of individual methylation-susceptible (closed bars; n = 172 loci) and nonmethylated (tinted bars; n = 105 loci) polymorphic loci in the set of 186 $Viola$ $cazorlensis$ individuals sampled for this study, as measured with Shannon's information index.

Epigenetic differentiation

Both single-locus and multilocus analyses revealed extensive epigenetic differentiation in the set of $V.$ $cazorlensis$ populations sampled. A total of 54 methylation-susceptible polymorphic loci (31.4% of total) exhibited statistically significant variation among populations in the proportion of methylated and nonmethylated states (Likelihood ratio $\chi^2 \geq 27.53$, $P \leq 0.0174$, q-value ≤ 0.0180; expected number of false positives = 0.0180×54 = 0.97). Molecular analysis of variance similarly revealed significant multilocus epigenetic differentiation between populations. The AMOVA-based estimate of epigenetic population differentiation, obtained from the 186 × 186 pairwise matrix of individual epigenetic distances, was Φ_{ST} = 0.130, which is statistically significant ($P < 0.0001$; permutation test with 10 000 repetitions).

Epigenetic differentiation in relation to adaptive genetic divergence

Adaptive genetic divergence of the $V.$ $cazorlensis$ populations studied was examined by screening for loci presumably subject to selection (or linked to selected loci), a dataset consisting of individual plant scores for 341 polymorphic, conventional AFLP markers obtained with eight different EcoRI–MseI primer combinations (see Herrera & Bazaga, 2008 for details), two restriction enzymes unaffected by cytosine methylation (McClelland et $al.$, 1994). These genetic data correspond to the same individual plants used in the MSAP analysis. The Bayesian population genomic scan of conventional AFLP loci data identified 23 loci ('outlier loci' hereafter, 6.7% of total) with a decisive posterior probability of being subject to divergent selection across local populations (or linked to selected loci), as denoted by their Bayes factors > 100 (Fig. 2). For these outlier loci, population differences in estimated allelic frequencies are therefore attributable to divergent selection acting in a population-specific manner (Foll & Gaggiotti, 2008; Herrera & Bazaga, 2008).

A dbRDA, in which the 14 × 14 matrix of AMOVA-based epigenetic distances between populations was regressed on the estimated population allelic frequencies for the 23 outlier AFLP loci, revealed a highly significant relationship between epigenetic population differentiation and adaptive genetic divergence. Ten outlier loci had significant or marginally significant effects on epigenetic differences between populations, and their allelic variation between populations accounted altogether for > 90% of between-population epigenetic variance (Table 2), which reveals a strong link between epigenetic differentiation and adaptive genetic divergence in the set of populations studied.

The statistical association between epigenetic and genetic population differentiation has a biological basis, as shown
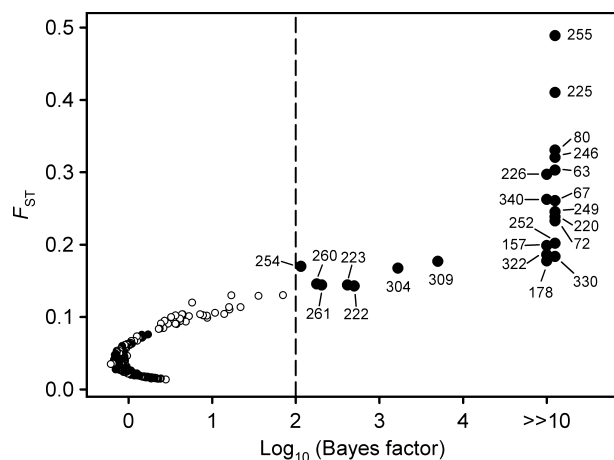
Fig. 2 Bayesian genomic scan of the *Viola cazorlensis* individuals sampled for this study to identify outlier loci presumably subject to selection (or linked to selected ones). Each point corresponds to a conventional amplified fragment length polymorphism method (AFLP) marker (*n* = 341). For each marker, an estimate of F_{ST} (fixation index, a measure of population differentiation) is plotted against the log-transformed Bayes factor (BF), which provides a scale of evidence in favour of the hypothesis that the locus is subject to selection vs the alternative hypothesis that it is not. The vertical dashed line corresponds to a posterior probability of 0.99 and BF = 100, a cut-off beyond which the evidence in favour of selection is considered decisive. The 23 outlier loci identified by the analysis are shown as closed dots, and associated numerals denote the corresponding locus number.

by the results of individual-level analyses on the 181 plants with epigenetic and genetic data simultaneously available. First, individual-based dbRDA conducted on the individual-by-individual epigenetic distance matrix revealed that the relationship between individual plants' epigenotypes and their nonneutral, adaptive genotypes was significant or marginally significant for 7 of the 10 outlier loci which

yielded significant epigenetic effects in the population-level analysis (Table 2). Second, AFLP–MSAP loci association analyses revealed six statistically significant associations in individual genomes between presence/absence of five of the significant outlier loci and the methylation state of four methylation-susceptible loci (Likelihood ratio $\chi^2 \geq 10.29$, $P \leq 0.0021$, *q*-value ≤ 0.1629; expected number of false positives = $0.1629 \times 6 = 0.98$). In three of these instances, the presence in an individual genome of the outlier locus fragment was directly correlated with the methylated state of the methylation-susceptible locus (ϕ coefficients = +0.267, +0.252, and +0.251), and inversely correlated in the other three instances (ϕ coefficients = −0.360, −0.280, and −0.255).

Discussion

The MSAP technique can underestimate genome-wide levels of DNA methylation, because it detects only methylation at CCGG sites and is unable to discriminate between methylation and fragment absence when the two cytosines are methylated on both strands, as neither *Hpa*II nor *Msp*I cleave such hypermethylated sites. Despite this limitation of the method (Cervera *et al.*, 2002; Salmon *et al.*, 2008), our study has revealed extensive levels of DNA methylation in the genome of wild-growing *V. cazorlensis* plants. Roughly half of the loci from the MSAP analysis were in a methylated state in a significant fraction of the individuals sampled. Ninety-five per cent of these methylation-susceptible loci were polymorphic, thus considerably higher than the 65% polymorphism exhibited by nonmethylated loci. Shannon's index measurements of methylation-associated, epiallelic diversity occurring in the sample of individuals studied were greater than those corresponding to sequence-based genetic

Table 2 Distance-based redundancy analysis (dbRDA) testing for: first, the relationship between epigenetic differences among *Viola cazorlensis* populations and the allelic frequencies at each population of the amplified fragment length polymorphism method (AFLP) outlier loci presumably subject to divergent selection ('Population-level analysis'); and second, the relationship between individual epigenotypes and outlier loci genotypes across *n* = 181 individual plants ('Individual-level analysis')

AFLP outlier locus	Population-level analysis (*n* = 14)			Individual-level analysis (*n* = 181)	
	Pseudo-*F*	*P*-value	Proportion of epigenetic variance explained	Pseudo-*F*	*P*-value
304	4.11	0.019	0.255	2.11	0.0011
249	3.69	0.024	0.187	2.26	0.0004
220	4.40	0.006	0.170	1.05	0.38
252	2.77	0.033	0.091	1.43	0.059
80	3.07	0.022	0.082	2.74	0.0001
340	2.37	0.081	0.054	1.44	0.053
157	2.57	0.100	0.048	1.45	0.049
246	3.30	0.097	0.045	1.05	0.39
309	4.74	0.076	0.037	1.26	0.15
330	7.67	0.065	0.011	1.38	0.073

Results correspond to conditional (sequential) tests of individual loci effects with a forward selection procedure that used the proportion of the total sum of squares explained by each locus as the criterion for selection.

diversity. Variation in methylation levels among different organs or developmental stages has been observed for several plant species (Xiong *et al.*, 1999; Portis *et al.*, 2004), but such effects should be disregarded as important sources of the extensive individual variation in methylation found in this study, as all DNA samples analysed were obtained from the same organ and collected at identical developmental stage (newly expanded, fully-grown leaves). Polymorphism levels of methylation-susceptible and nonmethylated loci found here for *V. cazorlensis* are virtually identical to those obtained in a MSAP study of 30 populations and lines of wild and cultivated morphotypes of *Brassica oleracea* (96 vs 63%, respectively; Salmon *et al.*, 2008). Similarly high methylation levels, and a greater polymorphism and/or diversity of methylation-susceptible loci relative to non-methylated ones, have been also reported for other wild (Li *et al.*, 2008; Marfil *et al.*, 2009) and cultivated plants (Keyte *et al.*, 2006; Fang *et al.*, 2008). Together, these studies (see also Cervera *et al.*, 2002; Riddle & Richards, 2002; Vaughn *et al.*, 2007) suggest that extensive intra-specific epigenetic variation of the sort found here for *V. cazorlensis* is probably widespread in wild plant populations.

Recent studies have stressed the nearly complete absence of information on the organization of epigenetic variation in natural plant populations (Kalisz & Purugganan, 2004; Rapp & Wendel, 2005; Bossdorf *et al.*, 2008). One important result of our study was therefore the discovery that naturally occurring, methylation-based epigenetic variation in *V. cazorlensis* was structured into distinct between- and within-population components, in a manner that could be considered analogous to the structuring of sequence-based, genuine genetic variation. There existed extensive within-population variance owing to individual epigenotypic variation (87% of total), a critical prerequisite for epigenetic variation to have some microevolutionary potential (Kalisz & Purugganan, 2004). In addition, there was significant epigenetic population differentiation at both single-locus and multilocus levels. About one-third of methylation-susceptible loci exhibited statistically significant variation across populations in the proportion of methylated and nonmethylated states, and the AMOVA revealed significant multilocus epigenetic differentiation among populations, the between-population component accounting for 13% of total epigenetic variance in the sample. Although no comparable quantitative data are available, epigenetic differences between populations have been also reported for another species (Li *et al.*, 2008).

Viola cazorlensis plants are very long-lived and cannot be propagated vegetatively, so that direct verification of the transgenerational constancy of observed epigenetic variation between populations and individuals by means of controlled crosses or common garden experiments was not possible. Some constancy, however, can be safely assumed in view of the data showing that inheritance of DNA methylation-based epigenetic marks across generations seems to be the rule in plants (Henderson & Jacobsen, 2007; Jablonka & Raz, 2009; Verhoeven *et al.*, 2010), and that intraspecific differences in DNA methylation patterns can be stably maintained for long periods, as inferred from differences between *Arabidopsis thaliana* ecotypes separated by relatively large evolutionary distances (Vaughn *et al.*, 2007). Under that assumption, and provided that the populations of *V. cazorlensis* studied form a single panmictic unit connected by extensive gene flow (Herrera & Bazaga, 2008), the maintenance of a between-population component of epigenetic variation can be interpreted as reflecting stable population differences presumably maintained by variable selection, as discussed in later text.

The population genomic approach (Black *et al.*, 2001) adopted in this study to examine adaptive genetic divergence between *V. cazorlensis* populations has proven useful for dissecting functionally important genetic variation in natural populations of nonmodel organisms (Vasemägi & Primmer, 2005; Foll & Gaggiotti, 2008). Scanning patterns of DNA polymorphisms at the genomic level through geno-typing numerous random anonymous loci spread over the entire genome of individuals in several populations, it is possible to identify genomic regions that exhibit deviant patterns of variation relative to the rest of the genome, presumably because of direct or indirect (through linkage) selection (Black *et al.*, 2001; Luikart *et al.*, 2003; Foll & Gaggiotti, 2008). Amplified fragment length polymorphism genotyping allows for both a large number of markers and an accurate assessment of baseline levels of neutral genetic variation across the whole genome, and markers from methylation-insensitive *Eco*RI-*Mse*I primer combinations are thus well suited to the population genomics analysis conducted in this study to detect loci showing signatures of divergent selection (Bonin *et al.*, 2007; Meudt & Clarke, 2007; Foll & Gaggiotti, 2008). The Bayesian method used here revealed that *c.* 6% of conventional AFLP loci had a decisive posterior probability of being subject to selection (or being linked to selected loci). This proportion nearly triplicates the figure obtained in an earlier analysis of the same data set using a frequentist procedure (Herrera & Bazaga, 2008), although all the outlier loci revealed by the earlier study have been corroborated here. This finding agrees with results reported by Foll & Gaggiotti (2008) for another reanalysis of AFLP data using their Bayesian approach, further stressing the superiority of this method over frequentist ones in population genomic scans for detecting candidate loci subject to selection.

The present study has shown that, in the set of *V. cazorlensis* populations studied, methylation-based epigenetic differentiation of populations is associated with adaptive genetic divergence. By contrast, earlier studies also using the MSAP technique generally found weak or nonsignificant correlations between epigenetic and genetic variation across

populations, ecotypes or accessions of the same species (Ashikawa, 2001; Cervera *et al.*, 2002; Keyte *et al.*, 2006; Li *et al.*, 2008; Salmon *et al.*, 2008). Differences in analytical approaches probably account for these contrasting results. In earlier studies, genetic characterization of the groups under comparison used multilocus approaches based on the whole set of nonmethylated AFLP fragments. As most of the latter presumably correspond to neutral, non-coding portions of the genome, their failure to find clear relationships between multilocus genetic and epigenetic differences is not surprising. By contrast, by focusing only on the small subset of adaptive, presumably nonneutral (or linked to nonneutral) AFLP loci, our analytical approach presumably enhanced the likelihood of detecting biologically significant relationships between genetic and epigenetic variation. At the between-population level, the dbRDA revealed that variation in allelic frequencies of 10 outlier loci accounted statistically for most epigenetic variance between populations. This strong statistical association between epigenetic and adaptive genetic divergence between population has a biological basis, as shown by dbRDA and association analyses at the individual plant level. These results provide compelling evidence for a close functional association in the genome of individual *V. cazorlensis* plants between the methylation state of some methylation-susceptible loci and the allelic condition of some nonneutral, adaptive AFLP loci. As outlined later in this text, this result has implications in relation to the evolutionary mechanisms underlying the adaptive differentiation of populations (Kalisz & Purugganan, 2004; Jablonka & Raz, 2009).

In plant genomes, DNA methylation controls gene expression levels through interfering with transcription and influencing the formation of transcriptionally silent heterochromatin (Finnegan *et al.*, 1998b; Zilberman *et al.*, 2007; Jablonka & Raz, 2009). When methylation affects large-effect genes, methylation changes can induce drastic discontinuous phenotypic alterations, including modification of floral symmetry, homeotic transformation of floral organs, and inhibition of fruit ripening (Finnegan *et al.*, 1998a; Cubas *et al.*, 1999; Manning *et al.*, 2006; Marfil *et al.*, 2009). However, methylation changes can also induce less dramatic phenotypic variation involving continuous traits such as flowering time, plant size, fecundity and resistance to pathogens or toxins (Finnegan *et al.*, 1996, 1998a; Sha *et al.*, 2005; Giménez *et al.*, 2006; Akimoto *et al.*, 2007; Jablonka & Raz, 2009), which suggests that genic methylation also affects suites of small-effect genes involved in the determination of complex quantitative traits (Johannes *et al.*, 2009). Results of the present study on *V. cazorlensis* are compatible with a scenario in which natural phenotypic variation among individuals and populations in quantitative traits (Herrera, 1990, 1993) could be accounted for by a combination of adaptive genetic variation and genetically controlled epigenetic differences: first, the significant

associations found across plants between the presence/absence of some adaptive AFLP loci and methylation state of some methylation-susceptible loci, is suggestive of functional genetic–epigenetic connections within individual genomes; second, the allelic frequencies of some of the adaptive AFLP loci found here to be related to epigenetic population differentiation (loci 220 and 246; Table 2) are also correlated with population differences in quantitative floral traits (Table 4 in Herrera & Bazaga, 2008); third, the broad phenotypic differences in average metric floral traits that characterize *V. cazorlensis* populations (Herrera, 1990) may be partly accounted for by the epigenetic population differentiation documented in this study.

Natural variation in gene methylation can be under complex genetic and epigenetic control (Kalisz & Purugganan, 2004; Zhang, 2008; Jablonka & Raz, 2009), and establishing the degree to which epigenetic variation is autonomous from genetic variation is central to evaluating the evolutionary relevance of the former as an additional, rather than redundant, inheritance system (Richards, 2006; Bossdorf *et al.*, 2008). A key aspect thus remaining to be elucidated is the evolutionary mechanism(s) underlying the three-way relationship linking the adaptive genetic, epigenetic and phenotypic divergence of the populations of *V. cazorlensis* studied here and by Herrera & Bazaga (2008). Assuming that phenotypic and genetic differentiation of populations have been driven by variable selection acting on genetically determined phenotypic traits (Herrera & Bazaga, 2008), such selection could be responsible for shaping the distributions of genotypes, epigenotypes or both. The first situation would exemplify a classical scenario of local adaptation through divergent selection on genetically determined traits. The other two possibilities would correspond to unexplored, but theoretically possible scenarios where evolution could proceed through selection of epigenetic variants without involvement of genetic variation, or through simultaneous, correlated selection on functionally linked, causative epigenetic and genetic variation (Kalisz & Purugganan, 2004; Rapp & Wendel, 2005; Richards, 2006; Jablonka & Raz, 2009). The data currently available do not allow one to discriminate among these alternatives for the *V. cazorlensis* study system, but our results suggest that correlated selection on epigenetic and genetic variation is a plausible evolutionary pathway. This intriguing possibility would simultaneously account for the association between epigenetic and adaptive genetic differentiation of populations, and the within-genome association between presence/absence of adaptive loci and the methylation state of methylation-susceptible loci. As selection is a process occurring within populations, this hypothesis could be tested through studies dissecting the contributions to fitness differences of individual epigenetic variation, individual genetic variation and their interaction in natural plant populations. Such expanded genetic–epigenetic framework could provide some of the

'many pieces missing from the epigenetic puzzle' (Bossdorf *et al.*, 2008).

Acknowledgements

References

Akimoto K, Katakami H, Kim HJ, Ogawa E, Sano CM, Wada Y, Sano H. 2007. Epigenetic inheritance in rice plants. *Annals of Botany* 100: 205–217.

Anderson MJ. 2003. *DISTLM forward: a FORTRAN computer program to calculate a distance-based multivariate analysis for a linear model using forward selection.* Auckland, New Zealand: Department of Statistics, University of Auckland.

Ashikawa I. 2001. Surveying CpG methylation at 5′-CCGG in the genomes of rice cultivars. *Plant Molecular Biology* 45: 31–39.

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, 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.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. *Ecology Letters* 11: 106–115.

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.

Cervera MT, Ruiz-García L, Martínez-Zapater JM. 2002. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. *Molecular Genetics and Genomics* 268: 543–552.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157–161.

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.

Fang J, Song C, Zheng Y, Qiao Y, Zhang Z, Dong Q, Chao CT. 2008. Variation in cytosine methylation in Clementine mandarin cultivars. *Journal of Horticultural Science and Biotechnology* 83: 833–839.

Finnegan EJ, Genger RK, Kovac K, Peacock WJ, Dennis ES. 1998a. DNA methylation and the promotion of flowering by vernalization. *Proceedings of the National Academy of Sciences, USA* 95: 5824–5829.

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES. 1998b. DNA methylation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 49: 223–247.

Finnegan EJ, Peacock WJ, Dennis ES. 1996. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proceedings of the National Academy of Sciences, USA* 93: 8449–8454.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.

Giménez C, Palacios G, Colmenares A. 2006. *Musa* methylated DNA sequences associated with tolerance to *Mycosphaerella fijiensis* toxins. *Plant Molecular Biology Reporter* 24: 33–43.

Henderson IR, Jacobsen SE. 2007. Epigenetic inheritance in plants. *Nature* 447: 418–424.

Herrera CM. 1990. 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, Bazaga P. 2008. Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology* 17: 5378–5390.

Herrera CM, Bazaga P. 2009. Quantifying the genetic component of phenotypic variation in unpedigreed wild plants: tailoring genomic scan for within-population use. *Molecular Ecology* 18: 2602–2614.

Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology* 84: 131–176.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P *et al.* 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics* 5: e1000530, doi:10.1371/journal.pgen.1000530.

Kalisz S, Purugganan MD. 2004. Epialleles via DNA methylation: consequences for plant evolution. *Trends in Ecology and Evolution* 19: 309–314.

Keyte AL, Percifield R, Liu B, Wendel JF. 2006. Intraspecific DNA methylation polymorphism in cotton (*Gossypium hirsutum* L.). *Journal of Heredity* 97: 444–450.

Li YD, Shan XH, Liu XM, Hu LJ, Guo WL, Liu B. 2008. Utility of the methylation-sensitive amplified polymorphism (MSAP) marker for detection of DNA methylation polymorphism and epigenetic population structure in a wild barley species (*Hordeum brevisubulatum*). *Ecological Research* 23: 927–930.

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

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genetics* 38: 948–952.

Marfil CF, Camadro EL, Masuelli RW 2009. Phenotypic instability and epigenetic variability in a diploid potato of hybrid origin, *Solanum ruiz-lealii. BMC Plant Biology* 9, doi: 10.1186/1471-2229-9-21.

McClelland M, Nelson M, Raschke E. 1994. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Research* 22: 3640–3659.

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

Portis E, Acquadro A, Comino C, Lanteri S. 2004. Analysis of DNA methylation during germination of pepper (*Capsicum annuum* L.) seeds using methylation-sensitive amplification polymorphism (MSAP). *Plant Science* 166: 169–178.

Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytologist* 168: 81–91.

Reyna-López GE, Simpson J, Ruiz-Herrera J. 1997. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Molecular and General Genetics* 253: 703–710.

Richards EJ. 2006. Inherited epigenetic variation – revisiting soft inheritance. *Nature Reviews Genetics* 7: 395–401.

Riddle NC, Richards EJ. 2002. The control of natural variation in cytosine methylation in *Arabidopsis*. *Genetics* **162**: 355–363.

Salmon A, Clotault J, Jenczewski E, Chable V, Manzanares-Dauleux MJ. 2008. *Brassica oleracea* displays a high level of DNA methylation polymorphism. *Plant Science* **174**: 61–70.

Sha AH, Lin XH, Huang JB, Zhang DP. 2005. Analysis of DNA methylation related to rice adult plant resistance to bacterial blight based on methylation-sensitive AFLP (MSAP) analysis. *Molecular Genetics and Genomics* **273**: 484–490.

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

Vanyushin BF. 2006. DNA methylation in plants. In: Doerfler W, Böhm P, eds. *DNA methylation: basic mechanisms*. Berlin, Germany: Springer-Verlag, 67–122.

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

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A *et al.* 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biology* **5**: 1617–1629.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist* **185**: 1108–1118.

Xiong LZ, Xu CG, Maroof MAS, Zhang QF. 1999. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Molecular and General Genetics* **261**: 439–446.

Zhang XY. 2008. The epigenetic landscape of plants. *Science* **320**: 489–492.

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

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39**: 61–69.

About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at **www.newphytologist.org**.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.

- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £151 in Europe/$279 in the USA & Canada for the online edition (click on 'Subscribe' at the website).

- If you have any questions, do get in touch with Central Office (**newphytol@lancaster.ac.uk**; tel +44 1524 594691) or, for a local contact in North America, the US Office (**newphytol@ornl.gov**; tel +1 865 576 5261).