

Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies

BENJAMIN SCHULZ,* R. LUTZ ECKSTEIN* and WALTER DURKA†

*Institute of Landscape Ecology and Resource Management, Research Centre for BioSystems, Land Use and Nutrition (IFZ), Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, D-35393 Giessen, Germany, †Department of Community Ecology (BZF), Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Straße 4, D-06120 Halle (Saale), Germany

Abstract

DNA methylation is an important, heritable epigenetic modification in most eukaryotic organisms that is connected with numerous biological processes. To study the impact of natural epigenetic variation in an ecological or evolutionary context, epigenetic studies are increasingly using methylation-sensitive amplification polymorphism (MSAP) for surveys at the population or species level. However, no consensus exists on how to interpret and score the multi-state information obtained from the MSAP banding patterns. Here, we review the previously used scoring approaches for population epigenetic studies and develop new alternatives. To assess effects of the different approaches on parameters of epigenetic diversity and differentiation, we applied eight scoring schemes to a case study of three populations of the plant species *Viola elatior*. For a total number of 168 detected polymorphic MSAP fragments, the number of ultimately scored polymorphic epiloci ranged between 78 and 286 depending on the particular scoring scheme. Both, estimates of epigenetic diversity and differentiation varied strongly between scoring approaches. However, linear regression and PCoA revealed qualitatively similar patterns, suggesting that the scoring approaches are largely consistent. For single-locus analyses of MSAP data, for example the search for loci under selection, we advocate a new scoring approach that separately takes into account different methylation types and thus seems appropriate for drawing more detailed conclusions in ecological or evolutionary contexts. An R script (MSAP_score.r) for scoring and basic data analysis is provided.

Keywords: data scoring, ecological epigenetics, methylation, methylation-sensitive amplification polymorphisms, MSAP, *Viola*

Received 21 December 2012; revision received 5 March 2013; accepted 6 March 2013

Introduction

DNA methylation is one of the most extensively studied epigenetic modifications in eukaryotic organisms and has been connected with numerous biological processes, extending from the level of single cells to the influence on ecological traits and microevolution (Bossdorf *et al.* 2008; Fujimoto *et al.* 2012). Together with changes of DNA-associated molecules, such as modifications of DNA- and histone proteins or changes of chromatin structure and small noncoding RNAs (Chatterjee & Vinson 2012), DNA methylation is part of a complex interacting epigenetic network that, without changing the underlying genetic code, modulates and controls gene expression.

In higher eukaryotes, DNA methylation almost exclusively occurs at the 5th carbon position of cytosine residues (Ratel *et al.* 2006) and is primarily found in the CG

dinucleotide context. While in mammals, non-CG methylations are abundant only in embryonic stem cells and rarely occur in somatic cells (Ramsahoye *et al.* 2000; Lister *et al.* 2009), plants harbour cytosine methylations at CHG and CHH sites (H = A, C, T) throughout their genomes. Methylated CG sites often occur in promoter regions and are generally linked to transcriptional repression, however, in both animals and plants CG methylation to some extent can also be associated with gene activation (Chatterjee & Vinson 2012; Grativol *et al.* 2012; Saze *et al.* 2012). Silencing or activation of protein-coding genes can be inherited through meiosis over several generations (reviewed in Jablonka & Raz 2009) and is giving rise to so-called epialleles (Schmitz *et al.* 2011).

Our knowledge about the functional role of DNA methylation and its impact on regulatory processes has dramatically increased through the development of new molecular and analysis tools (Grant-Downton & Dickinson 2005, 2006; Bock 2012). For model species, genome-wide profiles of DNA methylation are available at high resolu-

Correspondence: Benjamin Schulz, Fax: +49 (0)641 99 37 169; E-mail: benjamin.schulz@umwelt.uni-giessen.de

Box 1 Methylation-sensitive amplification polymorphisms—MSAP

The MSAP approach is technically a modification of the amplified fragment length polymorphism (Vos *et al.* 1995) using the same rare cutter *EcoRI* and substituting the frequent cutter *MseI* in two parallel runs by the more or less methylation-sensitive restriction enzymes *HpaII* and *MspI*. The two isoschizomers recognize and cleave the same tetranucleotide sequence 5'-CCGG, but differ in their sensitivity to the methylation state of cytosine.

Recent literature is somewhat inconsistent concerning the methylation sensitivity of the two enzymes (e.g. Salmon *et al.* 2008; Herrera & Bazaga 2010; Lira-Medeiros *et al.* 2010; Paun *et al.* 2010; Richards *et al.* 2012). According to the actual specifications of the restriction enzyme database REBASE (<http://rebase.neb.com/rebase/rebase.html>, accessed 20.02.2013), *HpaII* only recognizes sites that are hemi-methylated at the external cytosine (^{HMe}CCG), while *MspI* only recognizes sites being hemi- or fully methylated at the internal cytosine (^{HMe}CG or ^{Me}CG). Sites that are fully methylated at the external cytosine (^{Me}CCG) or hemi- or fully methylated at both, internal and external cytosines (^{HMe}C^{HMe}CG or ^{Me}C^{Me}CG) are not cut by either enzyme. However, CCGG-sequences being free of any methylation are digested by both. Some authors also reported an impaired cleavage of ^{Me}CCG-sites by *HpaII* (Korch & Hagblom 1986; Butkus *et al.* 1987) and ^{HMe}CCG-sites by *MspI* (Butkus *et al.* 1987), but as this is contradictory with later studies (see McClelland *et al.* 1994), it is not considered here.

The comparison of the resulting *EcoRI/HpaII* and *EcoRI/MspI* fragment profiles allows the detection of particular methylation states of the restriction sites (Fig. 1). In total, four conditions can be distinguished for a particular fragment: (i) condition I = fragments are present in both profiles indicating an unmethylated state, (ii) condition II = fragments are present only in *EcoRI/MspI* profiles indicating ^{HMe}CG- or ^{Me}CG-sites, (iii) condition III = fragments are present only in *EcoRI/HpaII* profiles indicating ^{HMe}CCG-sites and (iv) condition IV = complete absence of fragments in both profiles. The latter represents an uninformative state as absence of fragments can have multiple and equivocal reasons such as ^{Me}CCG-, ^{HMe}C^{HMe}CG- or ^{Me}C^{Me}CG-sites or a real fragment absence due to restriction site polymorphism.

tion using microarray technologies and next generation sequencing of bisulphite converted DNA (Bock 2012; Fujimoto *et al.* 2012). However, due to high costs and resource intensities, these methods usually are not suited for nonmodel species or for studies at the population level.

An alternative technique allowing extensive analyses of epigenetic variation for a high number of individuals is methylation-sensitive amplification polymorphism (MSAP), based on the use of the isoschizomers *HpaII* and *MspI*. These two restriction enzymes differ in their sensitivity to the methylation state of their recognition site 5'-CCGG and allow the comparison of large amounts of anonymous, methylation-sensitive CCGG regions across the genome (see Box 1), thereby covering the most frequent methylation types in the CG and CHG sequence context. The MSAP approach was first described by Reyna-Lopez *et al.* (1997) in a study on fungi and later modified for the use in plant species by Xiong *et al.* (1999). Ever since, the approach was adopted in more than 100 publications, focusing mainly on developmental biology (e.g. Portis *et al.* 2004; Hanai *et al.* 2010; Moran & Perez-Figueroa 2011; Meng *et al.* 2012), hybridization and polyploidization (e.g. Salmon *et al.* 2005; Zhao *et al.* 2008; Hegarty *et al.* 2011; Rodriguez *et al.* 2012) and plant breeding (e.g. Takata *et al.* 2005; Zhang *et al.* 2007; Salmon *et al.* 2009; Long *et al.* 2011).

In the past 5 years, MSAP analyses also became an important tool to answer questions in the emerging field

of 'ecological epigenetics', studying epigenetic processes in an ecological context (Bossdorf *et al.* 2008). To gain a deeper knowledge about natural genomic methylation and the impact of epialleles for processes like phenotypic plasticity and ecological adaption, several studies have used the MSAP technique for population epigenetic studies with plants (e.g. Li *et al.* 2008; Gao *et al.* 2010; Herrera & Bazaga 2010, 2011; Richards *et al.* 2012), vertebrates (e.g. Taylor *et al.* 2010; Massicotte & Angers 2012; Schrey *et al.* 2012) and a flower-inhabiting yeast (Herrera *et al.* 2011). Despite the rising importance of MSAP analyses for ecological studies and the increasing number of publications using the approach for population epigenetic analyses (for an overview see Appendix S1), the appropriate scoring of the resulting multistate data is still a challenge. In recent studies, at least five different scoring methods have been employed to assess epigenetic variation and some authors have also used MSAP data to gain information about genetic variation (Herrera & Bazaga 2010; Lira-Medeiros *et al.* 2010). Thus, there is a strong need for a common and comparable framework for the analysis of MSAP data and the necessity for a common and uniform scoring method. Therefore, the main aim of the present work is to describe and compare existing scoring approaches. In particular, using data from a case study, we will show how different scoring methods affect the resulting patterns and parameters of epigenetic variation within and among plant populations.

Finally, we will propose a scoring approach for future MSAP analyses in the context of ecological epigenetics.

Scoring MSAP raw data—how to get two out of four?

For population epigenetic studies using MSAP, the multistate raw data matrix resulting from the *EcoRI/HpaII* and *EcoRI/MspI* profiles (Box 1) needs to be transformed into a binary data matrix, allowing statistical analyses and computation of descriptive indices such as epigenetic diversity or differentiation. To extract binary epigenetic information from combined MSAP profiles, basically three main groups of scoring approaches exist (Fig. 2). The most widely used approach, hereafter called Methylation Scoring, considers only methylated fragments (condition II and III) as relevant and scores these as presence of information (score: '1'), whereas unmethylated fragments (condition I) are scored as absence (score '0') and fragment absence (condition IV) is scored either as absence (score: '0') or as missing data (score: 'NA'). The second scoring approach, hereafter called Non-Methylation Scoring, conversely scores only those fragments that stayed unmethylated. The third approach, hereafter called Mixed Scoring, combines both and considers all three types of MSAP fragments (conditions I–III) as potentially important. In the Mixed Scoring, between one and three subloci are created for each locus

of the raw data matrix, scoring existing methylated and unmethylated fragments separately. The multistate information of MSAP is thus split into separate epiloci (e.g. loci yielding three different fragment types result in three subepiloci and loci with only one fragment type result in only one epilocus).

Furthermore, for the three main MSAP scoring approaches, additional criteria are used to score the four types of methylation variation, resulting in eight different scoring methods (Table 1).

Methylation Scoring

Mainly three types of Methylation Scoring have been applied in recent population epigenetic studies: (1) Salmon *et al.* (2008) scored the methylated conditions II and III as '1' and the conditions I and IV as '0'; (2) Vergeer *et al.* (2012) differentiated condition I from IV by scoring unmethylated fragments as '0' and the absence of fragments as missing data, thus accounting for the uninformative state of condition IV; (3) Herrera & Bazaga (2010) applied the same scoring criteria as Vergeer *et al.* (2012), but in addition used a locus-specific threshold to classify individual loci as either 'methylation-susceptible' or 'unmethylated' before transforming them into a binary data matrix. This methylation threshold is estimated for each primer combination separately as $e_{Hpa} + e_{Msp} - 2e_{Hpa}e_{Msp}$









	Methylation status	HpaII	MspI	Type of information
	No methylation	+	+	Condition I
	Full-methylation of internal cytosine	–	+	Condition II
	Hemi-methylation of internal cytosine	–	+	Condition II
	Hemi-methylation of external cytosine	+	(–)	Condition III
	Full-methylation of external cytosine	(–)	–	Condition IV
	Full-methylation of both cytosines	–	–	Condition IV
	Hemi-methylation of both cytosines	–	–	Condition IV
	Unknown	–	–	Condition IV

Fig. 1 Sensitivity of the isoschizomers *HpaII* and *MspI* to different types of methylation of the 5'-CCGG restriction site ('+' = enzyme cuts; '–' = enzyme does not cut; brackets indicate contradictory reports of impaired cleavage). Methylated cytosines at the *HpaII/MspI* restriction site are indicated by orange colour.

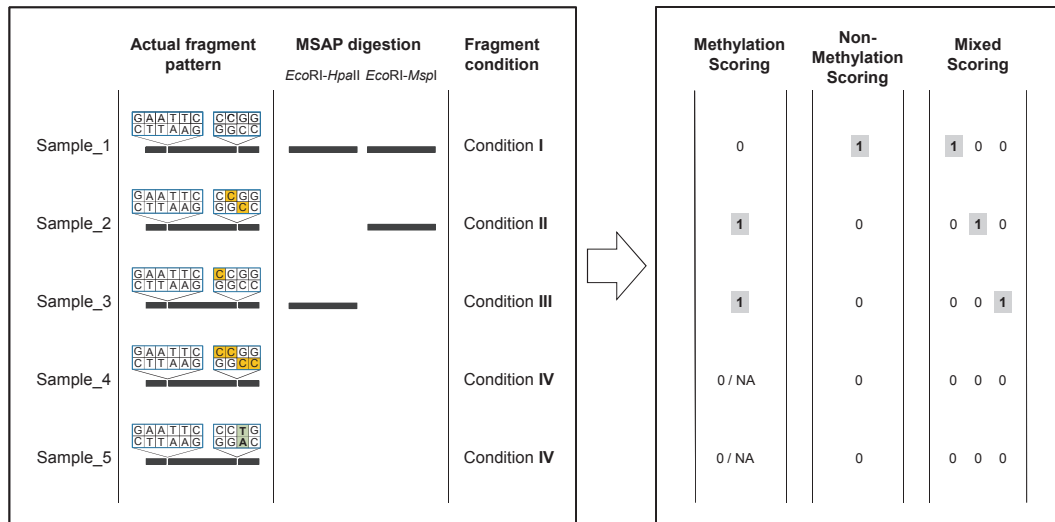


Fig. 2 Main groups of methylation-sensitive amplification polymorphisms (MSAP) scoring approaches: Methylation Scoring, Non-Methylation Scoring and Mixed Scoring. The chart depicts which fragments and methylation states at a certain locus are detected by MSAP analysis and are subsequently considered in the respective scoring approaches. *EcoRI* and *HpaII*/*MspI* restriction sites are accentuated. Methylated cytosines at the *HpaII*/*MspI* restriction site are indicated by orange colour. Restriction site mutation is indicated by green colour. NA denotes missing data.

(e_{Hpa} = error rate of *HpaII* profile, e_{Msp} = error rate of *MspI* profile, $e_{Hpa}e_{Msp}$ = error rate of combined profiles). All loci with observed proportions of discordant *HpaII*/*MspI* scores suggestive of methylation (i.e. number of individuals with contrasting *HpaII*/*MspI* scores divided by total number of sampled individuals) exceeding the threshold are classified as 'methylation-susceptible'. In their study on epigenetic differentiation of a violet species, Herrera & Bazaga (2010) used the remaining 'unmethylated' loci to assess the genetic diversity, treating them as dominant binary AFLP markers, scoring condition I as '1' and condition IV as '0'. For the purpose of this study, we only used the 'methylation-susceptible' loci obtained with a fixed threshold of 5% as performed by Moran & Perez-Figueroa (2011).

Non-Methylation Scoring

To our knowledge, Non-Methylation Scoring has been applied solely in the study of Lira-Medeiros *et al.* (2010). Here (4), all loci that contain condition III fragments are excluded from the data set as this type of fragments is determined not to be inherited over generations (Lira-Medeiros *et al.* 2010). Then, the *EcoRI*/*HpaII* and *EcoRI*/*MspI* profiles are analysed separately to assess the epigenetic and genetic structure, respectively. In essence, only the reduced *EcoRI*/*HpaII* profiles are scored which in fact represent unmethylated condition I fragments. To test the impact of the whole set of unmethylated fragments, we additionally applied a modification of Lira-Medeiros *et al.* (2010), hereafter called (5) Lira-Medei-

ros + *HpaII*, in which condition III fragments are not excluded but scored as '0'. In both (4) and (5), condition IV is scored '0'.

Mixed Scoring

Paun *et al.* (2010) separated the information provided in the MSAP raw data matrix into three marker types (6). For unmethylated markers, only condition I is scored as '1', for markers with ^{HMe}CG- or ^{Me}CG-sites, both condition I and II are scored as '1' and for markers with ^{HMe}C CG-sites condition I and III are scored as '1'. Thus, the methylated markers also include the information of the unmethylated condition I, respectively.

To separate the effective unmethylated and effective methylated fragments and to test for the particular impact of the methylated condition II and III, we suggest two new scoring variants. In (7) 'Mixed Scoring 1' for unmethylated markers only condition I is scored as '1' and for methylated markers both condition II and III are scored as '1'. In (8) 'Mixed Scoring 2', additionally the methylated markers are separated into markers with ^{HMe}CG- or ^{Me}CG-sites and markers with ^{HMe}C CG-sites, scoring either condition II or III as '1', respectively. In (6) to (8), condition IV is scored '0'.

Case study: epigenetic diversity of a perennial violet

The eight scoring approaches are partly complementary but differ considerably with respect to the amount of

Table 1 Scoring schemes of the eight scoring approaches used in the case study (NA denotes the treatment of condition IV as missing data)

HpaII/MspI banding pattern	1/1	0/1	1/0	0/0
Type of information	Condition I	Condition II	Condition III	Condition IV
Methylation status	Nonmethylated	^{HMe} CG & ^{Me} CG	^{HMe} CCG	No information
<i>METHYLATION SCORING</i>				
1 Salmon <i>et al.</i> (2008)	0	1	1	0
2 Vergeer <i>et al.</i> (2012)	0	1	1	NA
3 Herrera & Bazaga (2010)*	0	1*	1*	NA
<i>NON-METHYLATION SCORING</i>				
4 Lira-Medeiros <i>et al.</i> (2010)†	1	0	Loci excluded	0
5 Lira-Medeiros + HpaII	1	0	0	0
<i>MIXED SCORING</i>				
6 Paun <i>et al.</i> (2010)				
Unmethylated	1	0	0	0
^{HMe} CG & ^{Me} CG	1	1	0	0
^{HMe} CCG	1	0	1	0
7 Mixed Scoring 1				
Unmethylated	1	0	0	0
Methylated	0	1	1	0
8 Mixed Scoring 2				
Unmethylated	1	0	0	0
^{HMe} CG & ^{Me} CG	0	1	0	0
^{HMe} CCG	0	0	1	0

*Only loci exceeding a specific methylation threshold are scored.

†Loci containing type 3 fragments are not included in the scoring.

information extracted from MSAP profiles. To assess the impact of the MSAP scoring on descriptive parameters of epigenetic variation, we performed a case study with three populations of *Viola elatior* from contrasting habitats.

Plant material

Viola elatior (Violaceae) is a rare perennial iteroparous hemicryptophyte, which occurs in Central Europe along a successional gradient from floodplain meadows to alluvial woodland fringes (Eckstein & Otte 2005; Eckstein *et al.* 2006). The species has an octoploid genome ($2n = 40$) and exhibits a mixed mating system with potentially cross-pollinated chasmogamous and obli-

tory self-pollinated cleistogamous flowers. However, most seeds are produced by cleistogamous flowers (Eckstein & Otte 2005).

For the purpose of this comparative analysis, we used three populations of *V. elatior* from the Upper Rhine Valley located in the nature reserve 'Lampertheimer Altrhein' north of Mannheim, Germany. We selected three sites differing strongly in light availability: site1 (80.5% mean transmitted photosynthetic active radiation (PAR); 49°36'8.19"N; 8°26'50.15"E), site2 (12.5% PAR; 49°35'50.56"N; 8°26'48.69"E) and site3 (16.5% PAR, 49°35'44.70"N; 8°25'55.13"E). Distances between populations ranged between 500 and 1300 m. At each site, young and undamaged leaves from 21 to 24 randomly chosen reproductive plants were collected and immediately cooled to

below 10 °C. Samples were stored at –25 °C and freeze-dried for 48 h. To assure that any developmentally related variation in DNA methylation would not confound methylation patterns, care was taken to collect plants from all sites on the same date and at the same phenological stage, that is, only flowering individuals were sampled.

MSAP epigenotyping

Methylation-sensitive amplification polymorphisms analysis was performed according to the protocol provided as Appendix S2 using eight selective primer combinations. The 67 samples were scored in one batch for presence '1' or absence '0' of fragments obtained with *EcoRI/HpaII* and *EcoRI/MspI* resulting in a primary MSAP data matrix of 134 lines (Appendix S3). Error rate estimation was based on 18 replicate samples (27%), starting from the same DNA extracts. The overall error rate was 2.4%.

We generated epigenetic data matrices by comparison of the two digestion profiles. First, a multistate epigenetic raw data matrix of 67 lines containing condition I, II, III and IV was generated. Second, all epiloci that showed a monomorphic pattern or a deviation between *EcoRI/HpaII* and *EcoRI/MspI* in only one individual were excluded from the data set to prevent biased parameter estimates (Bonin *et al.* 2004). Third, the epigenetic raw data matrix was transformed to binary epigenetic data matrices according to the eight scoring approaches (Table 1). Fourth, epiloci that turned out monomorphic after transformation were excluded. Data transformation and selection of polymorphic loci were performed using the R script *MSAP_calc* (Appendix S4), which allows to perform the different scorings (see Pérez-Figueroa 2013 for an automated Methylation Scoring). These data matrices are also included in Appendix S3.

Data analysis

Data analysis of the binary epigenetic data matrices was performed using a marker based strategy, that is, we did not calculate allele frequencies (Bonin *et al.* 2007). Epigenetic diversity within populations was quantified using the R script *MSAP_calc* as (i) number and (ii) percentage of polymorphic epiloci (PLP_{epi}) and (iii) as mean Shannon's information index $H'_{\text{epi}} = -\sum p_i \log_2 p_i$ where p_i is the frequency of the epigenetic marker score '1' within the population. Patterns of individual and population differentiation were depicted by principal coordinates analyses (PCoA) with *GENALEX* 6.41 (Peakall & Smouse 2006) using the option covariance-standardized. The PCoA was based on a matrix of Nei and Li

distances (synonymous to Soerensen and Dice distance) calculated with *DistAFLP* (Mougel *et al.* 2002). This distance measure does not treat shared band absence as identical and thus excludes the uninformative state of absent MSAP fragments. As the Nei and Li distance is a semimetric and non-Euclidean measure, we applied square root transformation to the distance matrices to meet the assumptions of PCoA analyses (Legendre & Legendre 1998). The partitioning of epigenetic variance within and among populations was estimated as epigenetic phenotypic differentiation (Φ_{ST}) with an analysis of molecular variance (AMOVA) using *ARLEQUIN* 3.5.1.2. (Excoffier & Lischer 2010). The distance matrix for the AMOVA was based on the default option for haplotypic data computed with pairwise differences and a gamma a value of 0. Additionally, using the same software, we conducted a locus-by-locus AMOVA to characterize the epigenetic phenotypic differentiation at each locus.

Results

Methylation-sensitive amplification polymorphisms analysis of the 67 individuals with eight primer combinations revealed a total of 168 polymorphic loci in the epigenetic raw data matrix. The number of polymorphic loci for each primer combination ranged between 12 and 29. Overall, 7116 MSAP fragments were detected across the 168 loci and 67 individuals, consisting of 51.4% condition I, 40.6% condition II and 8.0% condition III fragments. Whereas most polymorphic loci showed more than one type of MSAP fragments, some loci displayed either only condition I fragments (14%), condition II fragments (9%) or condition III fragments (14%). In total, 128 loci included unmethylated fragments, 105 included fragments with ^{HMe}CG- or ^{Me}CG-sites and 53 included fragments with ^{HMe}CCG-sites.

After transformation to the eight binary epigenetic data matrices, the number of polymorphic epiloci ranged from 78 for the *Herrera-scoring* to 286 for the '*Mixed Scoring 2*' (Table 2). The Methylation and Non-Methylation Scoring approaches obtained similar numbers of around 100 (78–145) polymorphic epiloci, whereas the Mixed Scoring approaches revealed around 250 (245–286) loci. At population level, the mean number of polymorphic epiloci ranged between 57 for the *Herrera-scoring* to 189 for the '*Mixed Scoring 2*'.

Comparing the scoring approaches revealed different mean levels of epigenetic variation across populations (Table 2), with percentage of bands ranging between 65% (*Vergeer*) and 74% (*Paun*), percentage of private bands ranging between 8.0% (*Paun*) and 12.7% (*Vergeer*), percentage polymorphic loci ranging between 42% (*Lira-Medeiros*) and 58% (*Salmon*) and Shannon

Table 2 Measures of epigenetic diversity within three populations of *Viola elatior* obtained with different MSAP scoring approaches.

	Methylation Scoring			Non-Methylation Scoring		Mixed Scoring		
	Salmon	Vergeer	Herrera	Lira-Medeiros	Lira-Medeiros + HpaII	Paun	Mixed Scoring 1	Mixed Scoring 2
Informative epiloci	145	105	78	100	128	245	273	286
Bands per population								
site1	98	65	56	59	81	174	179	184
site2	88	61	50	63	83	175	171	172
site3	107	78	65	80	99	195	206	212
mean	97.7	68.0	57.0	67.3	87.7	181.3	185.3	189.3
Bands per population (%)								
site1	67.6	61.9	71.8	59	63.3	71.0	65.6	64.3
site2	60.7	58.1	64.1	63	64.8	71.4	62.6	60.1
site3	73.8	74.3	83.3	80	77.3	79.6	75.5	74.1
mean	67.4	64.8	73.1	67.3	68.5	74.0	67.9	66.2
Private bands per population (%)								
site1	14.5	13.3	10.3	13.0	13.3	12.2	13.9	15.4
site2	5.5	6.7	2.6	4.0	4.7	3.3	5.1	5.2
site3	14.5	18.1	11.5	15.0	12.5	8.6	13.6	15.4
mean	11.5	12.7	8.1	10.7	10.2	8.0	10.9	12.0
PLP_{epi}								
site1	57.2	45.7	51.3	38.0	43.0	51.4	50.5	50.0
site2	46.2	35.2	39.7	30.0	32.8	34.7	39.9	38.5
site3	69.7	62.9	70.5	57.0	57.8	58.8	64.1	63.3
mean	57.7	47.9	53.8	41.7	44.5	48.3	51.5	50.6
H'_{epi}								
site1	0.31	0.28	0.35	0.20	0.24	0.27	0.28	0.27
site2	0.23	0.21	0.25	0.15	0.16	0.17	0.20	0.19
site3	0.38	0.40	0.48	0.31	0.32	0.31	0.35	0.34
mean	0.31	0.30	0.36	0.22	0.24	0.25	0.28	0.27
overall	0.50	0.58	0.62	0.61	0.61	0.56	0.55	0.53

PLP_{epi} —percentage of polymorphic epiloci; H'_{epi} —Shannon's information index; MSAP, methylation-sensitive amplification polymorphisms.

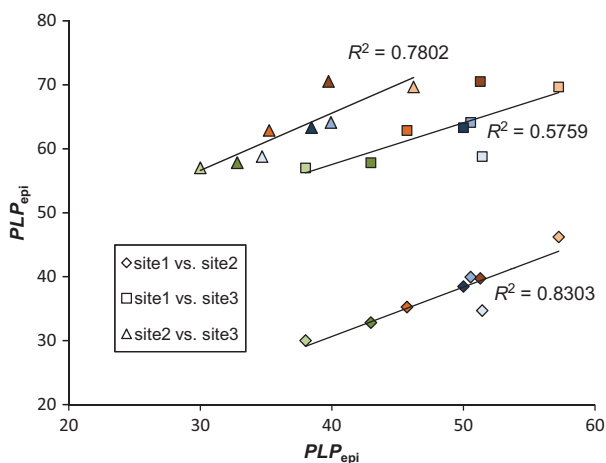


Fig. 3 Linear regressions of percentage polymorphic epiloci (PLP_{epi}) obtained with different methylation-sensitive amplification polymorphisms (MSAP) scoring approaches for three populations of *Viola elatior*. Colours refer to different scoring approaches (see Table 1).

diversity index ranging between 0.22 (*Lira-Medeiros*) and 0.36 (*Herrera*).

Comparing diversity parameters between the three populations revealed that for all scoring approaches, the number and percentage of bands per population were highest for 'site3' (Table 2). In contrast, 'site1' and 'site2' differed in rank depending on scoring, with 'site1' showing the lowest numbers in the Non-Methylation Scoring approaches and the *Paun*-scoring and 'site2' showing the lowest numbers in the other scoring schemes. The percentage of private bands per population was consistently lowest in site2, whereas site1 and site3 had the highest values depending on the scoring approach. For PLP_{epi} (Fig. 3) and Shannon's information index H'_{epi} , site2 and site3 consistently had the lowest and highest levels of epigenetic variation, respectively, irrespective of the scoring approach.

Principal coordinates analyses of epigenetic distances in all cases clearly separated the three populations, forming well-defined clusters (Fig. 4). Variation explained by

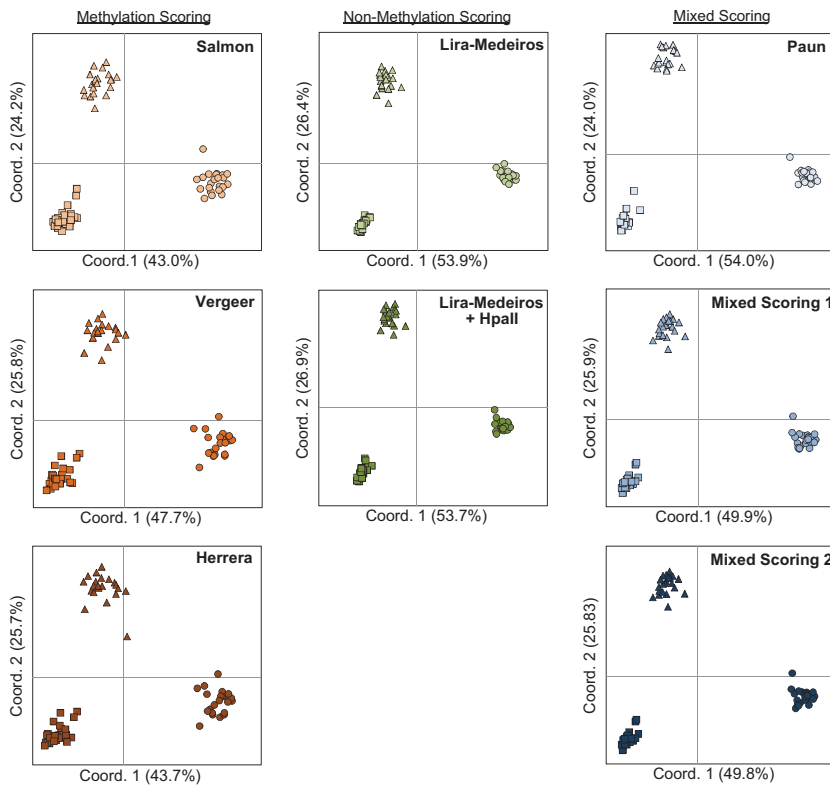


Fig. 4 Principal Coordinates Analyses (PCoA) of square root transformed Nei and Li distances of epigenetic binary data matrices obtained with different methylation-sensitive amplification polymorphisms (MSAP) scoring approaches (see Table 1). Populations of *Viola elatior* are indicated by symbols (\circ = site1, \square = site2, Δ = site 3).

Table 3 Analysis of Molecular Variance (AMOVA) of differently scored MSAP data for three populations of *Viola elatior*

	Methylation Scoring			Non-Methylation Scoring		Mixed Scoring		
	Salmon	Vergeer	Herrera	Lira-Medeiros	Lira-Medeiros + HpaII	Paun	Mixed Scoring 1	Mixed Scoring 2
AMOVA global								
Variation among populations (%)	47.4	57.6	58.7	73.7	70.9	65.7	60.4	60.2
Variation within populations (%)	52.6	42.4	41.3	26.3	29.1	34.3	39.6	39.8
Φ_{ST}	0.474	0.576	0.587	0.737	0.708	0.657	0.604	0.602
AMOVA locus-by-locus								
Significantly differentiated loci	73	49	44	66	82	136	154	153
Significantly differentiated loci (%)	50.3	60.5	65.7	66	64.1	55.5	56.4	53.7

MSAP, methylation-sensitive amplification polymorphisms.

the first three axes ranged between 77.0% and 78.1% for the Methylation Scoring, between 86.5% and 86.9% for the Non-Methylation Scoring and between 82.9% and 84.6% for Mixed Scoring approaches. Individuals were less clumped for the Methylation Scoring than for other scorings.

Strong epigenetic population differentiation was corroborated by the AMOVA (Table 3). In all scoring approaches, except the *Salmon-scoring*, a greater amount

of variation was accounted for by differences among populations than by variation within populations. However, Φ_{ST} values differed strongly between the scoring approaches, ranging from 0.474 for the *Salmon-scoring* to 0.737 for the *Lira-Medeiros-scoring*. Similarly, the percentage of significantly differentiated epiloci, as identified by locus-by-locus AMOVA (Table 3), ranged between 50.3% for the *Salmon-scoring* and 66% for the *Lira-Medeiros-scoring*. In general, Mixed Scoring approaches

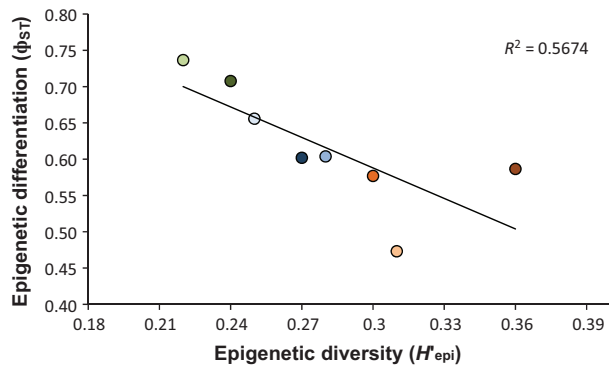


Fig. 5 Linear regression of epigenetic diversity (H'_{epi}) and epigenetic phenotypic differentiation (Φ_{ST}) obtained with different methylation-sensitive amplification polymorphisms (MSAP) scoring approaches for three populations of *Viola elatior*. Colours refer to different scoring approaches (see Table 1).

revealed the highest number, but lowest percentage of significantly differentiated epiloci compared with Methylation and Non-Methylation Scoring. As expected, epigenetic differentiation was negatively correlated with epigenetic diversity (Fig. 5) across scoring approaches. However, the *Herrera*- and *Salmon*-scoring accounted differently for within and among population variation as they showed the largest deviation from the regression line.

Discussion

Benefits and limitations of MSAP scoring approaches

Although Methylation-, Non-Methylation- and Mixed Scoring approaches have been used to assess epigenetic variation previously, the choice of scoring method has rarely been justified. The question arises which kind of epigenetic information is relevant in an ecological or evolutionary context.

Most MSAP studies used a kind of Methylation Scoring, concentrating on the effect of a subset of methylation types, namely $HMeCG$, $MeCG$ and $HMeCCG$. All other possible methylation types which are not distinguishable from real fragmentation absence (condition IV) are ignored. For animal species, cytosine methylation in the CHG sequence context is very rare (Feng *et al.* 2010; Zemach *et al.* 2010) and thus condition IV seems negligible. In contrast, plants frequently exhibit methylations in the CHG sequence context, which leads to an underestimation of genome-wide methylation levels when using Methylation Scoring. For the model species *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa*, Feng *et al.* (2010) observed overall methylation levels ranging between 23.3% and 59.4% in the CG context and between 5.92% and 20.9% in the CHG context, showing the potential impact of an

underestimation of methylation levels for plant species due to the uninformative state of condition IV. Although, consequently treating condition IV as missing data might be therefore more accurate when using Methylation Scoring (e.g. Herrera & Bazaga 2010; Vergeer *et al.* 2012), missing values will compromise the subsequent data analysis. A further drawback of Methylation Scoring is the assembly of condition II and III into one score, neglecting the fact that methylation in the CG and CHG context is catalysed by different enzymes (Furner & Matzke 2011) and thus underlies different regulating processes. Moreover, CG and CHG methylation might potentially account for different, probably counteracting epigenetic effects and thus a combined scoring might blur the real, effective epigenetic pattern.

Although earlier papers dealing with methylation-sensitive restriction enzymes suggested focusing on unmethylated fragments (Quint & Cedar 1981), most of the recent MSAP studies ignored this idea. However, especially ecologically relevant processes like adaption or phenotypic variation are potentially better represented by unmethylated fragments, as in most cases, demethylation rather than methylation seems to account for gene expression (Zemach *et al.* 2010; Jones 2012; Raynal *et al.* 2012) and thus actively contributes to phenotypic variation. Furthermore, when considering only unmethylated fragments, the uncertainty due to condition IV is avoided. However, a major disadvantage of the Non-Methylation Scoring is that epigenetic variation of condition II and III is not taken into account.

Mixed Scoring approaches represent a compromise between scoring either methylated or unmethylated fragments. By incorporating both groups of fragments, some drawbacks of the single scoring methods can be avoided and more of the underlying information is utilized. Especially with regard to locus-specific analyses, like identifying putative epiloci under selection or correlating epigenetic patterns with phenotypic variation, Mixed Scoring seems to be favourable as it allows to test for both, the potential contribution of unmethylated and methylated fragments. As natural selection directly targets phenotypic variation and either silencing or activation of genes can lead to phenotypic adaption to particular environmental conditions, separately considering condition I, II and III might therefore give the most comprehensive picture. Paun *et al.* (2010) identified several epiloci putatively under selection using mixed scoring and considering three marker types (i.e. unmethylated markers, $HMeCG$ - or $MeCG$ markers and $HMeCCG$ -markers). However, in the *Paun*-scoring, unmethylated fragments are incorporated into the two methylated marker types which probably introduces some bias as unmethylated condition I fragments may be rescored up to three times (e.g. when a locus contains all three

types of MSAP fragments). This disadvantage is circumvented in 'Mixed Scoring 2' which considers each condition only once and might therefore better represent the actual methylation patterns.

A general drawback of the MSAP method is the unavailability of comparative information about genetic variation and thus the necessity for separate genetic analyses. Some authors (Herrera & Bazaga 2010; Lira-Medeiros *et al.* 2010) have applied certain MSAP scoring variants to assess genetic variation. Whereas Herrera & Bazaga (2010) treated those loci as genetic AFLP markers that did not exceed an estimated methylation threshold, Lira-Medeiros *et al.* (2010), after excluding all loci with condition III fragments, used the *MspI* profiles as genetic marker. Both scoring variants consequently treat condition IV as mutated restriction sites, ignoring methylation as cause. However, Zhang *et al.* (2007) showed that a large proportion of observed fragment absences actually represent methylation polymorphisms rather than sequence variation, thus questioning the use of MSAP profiles as genetic markers.

A modification of the original MSAP, called metAFLP (Bednarek *et al.* 2007), avoids the disadvantage of condition IV using the two isoschizomers *Acc65I* and *KpnI*, which are sensitive and insensitive to cytosine methylations of the recognition sequence 5'-GGTACC, respectively. Direct comparison of the digestion patterns allows for assessment of both, genetic and epigenetic variation. As *KpnI* cleaves when either the internal or external cytosine is methylated and shows an impaired cleavage (50%) when both cytosines are methylated (see REBASE specification), it is not possible to differentiate methylation in the CG, CHG or CHH contexts.

Case study

The different MSAP scoring approaches resulted in widely differing numbers of epiloci and revealed strong differences of diversity estimates for three populations of *V. elatior*. Consistent differences were found between Methylation, Non-Methylation and Mixed Scoring for PLP_{epi} and mean Shannon's information index H'_{epi} . We found a 64% difference between highest and lowest overall values for H'_{epi} , that is between the *Herrera-scoring* ($H'_{\text{epi}} = 0.36$) and the *Lira-Medeiros-scoring* ($H'_{\text{epi}} = 0.22$). Thus, the large difference found in epigenetic variation in the original publications between *Viola cazorlensis* ($H'_{\text{epi}} = 0.45$; Herrera & Bazaga 2010) and *Laguncularia racemosa* ($H'_{\text{epi}} = 0.05$; Lira-Medeiros *et al.* 2010) might at least partly be attributed to the scoring procedures used. Estimates of epigenetic differentiation were very high, ranging between $\Phi_{\text{ST}} = 0.474$ (*Salmon-scoring*) and 0.737 (*Lira-Medeiros-scoring*). However, although distinct differences could be found for certain scoring variants,

neither the type nor the number of epiloci seemed to strongly affect the estimates of epigenetic population structure.

Conclusions

Methylation-sensitive amplification polymorphisms is a powerful tool to study epigenetic variation by investigating cytosine methylation for a large number of loci across the genome. Although the eight scoring approaches revealed some distinct differences, the resulting patterns of epigenetic diversity within and among populations appeared to be quite robust. Therefore, there seems to be not one best scoring approach for multilocus analyses. However, for single-locus analyses like genome scans or locus-by-locus-AMOVA, all types of methylation polymorphisms and resulting fragments should be considered separately. Only then the ambiguous functional role of differently methylated and unmethylated fragments that might have both, activating or down-regulating effects can be assessed. For studies focusing on single-locus analyses, we therefore advocate the use of 'Mixed Scoring 2' as it allows evaluating the different methylation types and their role in ecological or evolutionary processes.

Acknowledgements

We want to thank Josef Scholz vom Hofe and Lena Föhr for help and logistic support during the field work, and Shawn Narum, Carlos M. Herrera and one anonymous reviewer for valuable comments on the manuscript. Financial support was obtained from the Deutsche Forschungsgemeinschaft (DFG, grant EC 209/7-1).

References

- Bednarek PT, Orłowska R, Koebner RMD, Zimny J (2007) Quantification of the tissue-culture induced variation in barley (*Hordeum vulgare* L.). *BMC Plant Biology*, **7**, 10.
- Bock C (2012) Analysing and interpreting DNA methylation data. *Nature Reviews Genetics*, **13**, 705–719.
- Bonin A, Bellemain E, Bronken Eidesen P *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.
- Bonin A, Ehrlich 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 C, Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters*, **11**, 106–115.
- Butkus V, Petrauskienė L, Maneliene Z *et al.* (1987) Cleavage of methylated CCCGGG sequences containing either N4-methylcytosine or 5-methylcytosine with *MspI*, *HpaII*, *SmaI*, *XmaI* and *Cfr9I* restriction endonucleases. *Nucleic Acids Research*, **15**, 7091–7102.
- Chatterjee R, Vinson C (2012) CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 763–770.

- Eckstein RL, Otte A (2005) Effects of cleistogamy and pollen source on seed production and offspring performance in three endangered violets. *Basic and Applied Ecology*, **6**, 339–350.
- Eckstein RL, Hölzel N, Danihelka J (2006) Biological Flora of Central Europe: *Viola elatior*, *V. pumila* and *V. stagnina*. *Perspectives in Plant Ecology, Evolution and Systematics*, **8**, 45–66.
- Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Feng S, Cokus SJ, Zhang X *et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, **107**, 8689–8694.
- Fujimoto R, Sasaki T, Ishikawa R *et al.* (2012) Molecular mechanisms of epigenetic variation in plants. *International Journal of Molecular Sciences*, **13**, 9900–9922.
- Furner IJ, Matzke M (2011) Methylation and demethylation of the *Arabidopsis* genome. *Current Opinion in Plant Biology*, **14**, 137–141.
- Gao L, Geng Y, Li B, Chen J, Yang J (2010) Genome-wide DNA methylation alterations of *Alternanthera philoxeroides* in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant, Cell & Environment*, **33**, 1820–1827.
- Grant-Downton RT, Dickinson HG (2005) Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Annals of Botany*, **96**, 1143–1164.
- Grant-Downton RT, Dickinson HG (2006) Epigenetics and its implications for plant biology 2. The “epigenetic epiphany”: epigenetics, evolution and beyond. *Annals of Botany*, **97**, 11–27.
- Gratival C, Hemerly AS, Ferreira PCG (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 176–185.
- Hanai LR, Floh EIS, Fungaro MHP *et al.* (2010) Methylation patterns revealed by MSAP profiling in genetically stable somatic embryogenic cultures of *Ocotea catharinensis* (Lauraceae). *In Vitro Cellular & Developmental Biology. Plant*, **46**, 368–377.
- Hegarty MJ, Batstone T, Barker GL *et al.* (2011) Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). *Molecular Ecology*, **20**, 105–113.
- Herrera CM, Bazaga P (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. *New Phytologist*, **187**, 867–876.
- Herrera CM, Bazaga P (2011) Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Molecular Ecology*, **20**, 1675–1688.
- Herrera CM, Pozo MI, Bazaga P (2011) Jack of all nectars, master of most: DNA methylation and the epigenetic basis of niche width in a flower-living yeast. *Molecular Ecology*, **21**, 2602–2616.
- Jablunka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology*, **84**, 131–176.
- Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**, 484–492.
- Korch C, Hagblom P (1986) In-vivo-modified gonococcal plasmid pJD1—a model system for analysis of restriction enzyme sensitivity to DNA modifications. *European Journal of Biochemistry*, **161**, 519–524.
- Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd edn. Elsevier Science, Amsterdam.
- Li Y, Shan X, Liu X *et al.* (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.
- Lira-Medeiros CF, Parisod C, Fernandes RA *et al.* (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE*, **5**, e10326.
- Lister R, Pelizzola M, Dowen RH *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
- Long Y, Xia W, Li R *et al.* (2011) Epigenetic QTL mapping in *Brassica napus*. *Genetics*, **189**, 1093–1102.
- Massicotte R, Angers B (2012) General-purpose genotype or how epigenetics extend the flexibility of a genotype. *Genetics Research International*, **2012**, 1–7.
- 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.
- Meng FR, Li YC, Yin J *et al.* (2012) Analysis of DNA methylation during the germination of wheat seeds. *Biologia Plantarum*, **56**, 269–275.
- Moran P, Perez-Figueroa A (2011) Methylation changes associated with early maturation stages in the Atlantic salmon. *BMC Genetics*, **12**, 86.
- Mougel C, Thioulouse J, Perri G, Nesme X (2002) A mathematical method for determining genome divergence and species delineation using AFLP. *International Journal of Systematic and Evolutionary Microbiology*, **52**, 573–586.
- Paun O, Bateman RM, Fay MF *et al.* (2010) Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylophiza*: Orchidaceae). *Molecular Biology and Evolution*, **27**, 2465–2473.
- Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
- Pérez-Figueroa A (2013) msap: a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. *Molecular Ecology Resources*, in press [doi: 10.1111/1755-0998.12064].
- Portis E, Acquadro A, Comino C, Lanteri S (2004) Analysis of DNA methylation during germination of pepper (*Capsicum annum* L.) seeds using methylation-sensitive amplification polymorphism (MSAP). *Plant Science*, **166**, 169–178.
- Quint A, Cedar H (1981) In vitro methylation of DNA with Hpa II methylase. *Nucleic Acids Research*, **9**, 633–646.
- Ramsahoye BH, Biniszkiwicz D, Lyko F *et al.* (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences*, **97**, 5237–5242.
- Ratel D, Ravanat JL, Berger F, Wion D (2006) N6-methyladenine: the other methylated base of DNA. *BioEssays*, **28**, 309–315.
- Raynal NJ-M, Si J, Taby RF *et al.* (2012) DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory. *Cancer Research*, **72**, 1170–1181.
- Reyna-Lopez 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 & General Genetics: MGG*, **253**, 703–710.
- Richards CL, Schrey AW, Pigliucci M (2012) Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. *Ecology Letters*, **15**, 1016–1025.
- Rodriguez MP, Cervigni GDL, Quarín CL, Ortiz JPA (2012) Frequencies and variation in cytosine methylation patterns in diploid and tetraploid cytotypes of *Paspalum notatum*. *Biologia Plantarum*, **56**, 276–282.
- Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology*, **14**, 1163–1175.
- Salmon A, Clotault J, Jenczewski E, Chable V, Manzaneres-Dauleux MJ (2008) *Brassica oleracea* displays a high level of DNA methylation polymorphism. *Plant Science*, **174**, 61–70.
- Salmon A, Bellis H, Chable V, Manzaneres-Dauleux MJ (2009) Identification of differentially expressed genes related to aberrant phenotypes in *Brassica oleracea* var. *botrytis*. *Plant Breeding*, **128**, 631–639.
- Saze H, Tsugane K, Kanno T, Nishimura T (2012) DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. *Plant and Cell Physiology*, **53**, 766–784.
- Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, **334**, 369–373.
- Schrey AW, Coon CAC, Grispo MT *et al.* (2012) Epigenetic variation may compensate for decreased genetic variation with introductions: a case

- study using house sparrows (*Passer domesticus*) on two continents. *Genetics Research International*, **2012**, 1–7.
- Takata M, Kishima Y, Sano Y (2005) DNA methylation polymorphisms in rice and wild rice strains: detection of epigenetic markers. *Breeding Science*, **55**, 57–63.
- Taylor E, Blouin MS, Thuillier V *et al.* (2010) No evidence for large differences in genomic methylation between wild and hatchery steelhead (*Oncorhynchus mykiss*). *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 217–224.
- Vergeer P, Wagemaker N, Ouborg NJ (2012) Evidence for an epigenetic role in inbreeding depression. *Biology Letters*, **8**, 798–801.
- Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- 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.
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **328**, 916–919.
- Zhang MS, Yan HY, Zhao N *et al.* (2007) Endosperm-specific hypomethylation, and meiotic inheritance and variation of DNA methylation level and pattern in sorghum (*Sorghum bicolor* L.) inter-strain hybrids. *Theoretical and Applied Genetics*, **115**, 195–207.
- Zhao Y, Yu S, Xing C, Fan S, Song M (2008) Analysis of DNA methylation in cotton hybrids and their parents. *Molecular Biology*, **42**, 169–178.

B.S., R.L.E. and W.D. designed research; B.S. performed lab work, analyzed data and wrote the manuscript under the supervision of R.L.E. and W.D.; W.D. developed the R-script; all authors read and approved the final manuscript.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Overview of publications using MSAP or metAFLP for population epigenetic studies.

Appendix S2 MSAP genotyping protocol and applied primer combinations.

Appendix S3 Primary MSAP data, multistate epigenetic raw data and resulting binary data for the eight scoring approaches.

Appendix S4 R script (MSAP_score.r) for automated scoring and basic data analysis.