Ecological and molecular mechanisms in plants to cope with environmental heterogeneity: a case study in *Viola elatior*

Dissertation

zur Erlangung des Doktorgrades

– Dr. rer. nat. –

der Naturwissenschaftlichen Fakultät
der Justus-Liebig-Universität Gießen

vorgelegt von

Dipl.-Biol. Benjamin Schulz

Gießen, September 2015

First referee:

Prof. Dr. R. Lutz Eckstein
Department of Environmental and Life Sciences – Biology
Karlstad University
SE-651 88 Karlstad, Sweden

Second referee:

Prof. Dr. Thomas Wilke
Department of Animal Ecology and Systematics
Justus-Liebig-University Gießen
Heinrich-Buff-Ring 26–32 IFZ, 35392 Gießen

This thesis is based on the following two publications and one manuscript:

1) Schulz B., Eckstein R.L., Durka W. (2013) Scoring and analysis of methylation sensitive amplification polymorphisms (MSAP) for epigenetic population studies. *Molecular Ecology Resources*, 13, 642–653.*
2) Schulz B., Eckstein R.L., Durka W. (2014) Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Molecular Ecology*, 23, 3523–3537.*
3) Schulz, B., Durka, W., Danihelka, J., Eckstein, R.L. Can persistent seed banks buffer genetic effects of declining population size and selection? (manuscript)

For all three papers I conducted most of the field and laboratory work, and had the main responsibility for data analysis and writing. Both, Prof. Dr. R.L. Eckstein and Dr. W. Durka were involved in study design. All co-authors contributed constructive suggestions and valuable comments.

The research reported in this thesis was carried out at the Institute of Landscape Ecology and Resource Management, Research Centre for Biosystems, Land Use and Nutrition (IFZ), Justus-Liebig-University Giessen, Germany.

* Reprinted with kind permission of Wiley-Blackwell

Contents

Chapter 1	Synthesis	1
Chapter 2	Scoring and analysis of methylation sensitive amplification polymorphisms for epigenetic population studies	21
Chapter 3	Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb	45
Chapter 4	Can persistent seed banks buffer genetic effects of declining population size and selection?	75
	Summary	105
	Zusammenfassung	107
	Acknowledgements	110
	List of publications	111

CHAPTER 1

ECOLOGICAL AND MOLECULAR MECHANISMS IN PLANTS TO COPE WITH ENVIRONMENTAL HETEROGENEITY: A SYNTHESIS

Introduction

Organisms can be exposed to strong fluctuations in environmental conditions. Unlike animals that might migrate to escape unfavorable conditions, plants due to their sessile nature directly need to face biotic or abiotic change. In the course of evolution plants therefore have developed a variety of adaptations and risk reduction mechanisms, enhancing either their own chance of survival or that of their offspring. As complex as the interdependencies within ecosystems are as diverse and complex might be also the mechanisms to cope with environmental change and its unpredictability, ranging from the molecular to the organismic and population level. This thesis focuses on two of these mechanisms, namely (1) the environmental adjustment by epigenetic variation and (2) the dispersal of offspring in time by the formation of persistent soil seed banks.

Epigenetic variation

The diversity of phenotypic traits within species or natural populations plays an integral role for their ability to cope with environmental heterogeneity. Both, phenotypic diversity as well as plasticity (i.e. the potential amplitude of traits within individuals) are foundations of important environment related processes such as local adaptation and range expansion and may be a decisive factor for population persistence or extinction. Whereas for many years ecologists and population biologists focused on the underlying genetic diversity to explain natural variation and microevolution, recent research strongly implies that also epigenetic variation may play a significant role (Bossdorf *et al.* 2008).

Epigenetics is the study of heritable changes in gene expression and function that cannot be explained by changes in the DNA sequence (Richards 2006; Bossdorf *et al.* 2008). An array of complex interacting epigenetic mechanisms that either activates or silences gene expression has been identified in the past years, including chemical modifica-

tions of DNA and histones, position effects and interference by small non-coding RNAs (Berger 2007). Probably the most extensively studied and best understood epigenetic mechanism is the reversible methylation of cytosine residues in the DNA (Feng *et al.* 2010). Whereas in vertebrates cytosine methylation is almost exclusively found in the symmetric CG-sequence context and in some model species like the nematode *Caenorhabditis elegans* even is completely absent (Feng *et al.* 2010), in plants it is generally more extensive and affects a wider sequence diversity than in animals (Vanyushin and Ashapkin 2011). Here, cytosine methylations are found throughout the genome in symmetric CG- and CHG-sites (H = A, C, T) and asymmetric CHH-sites. Methylations in the three sequence contexts differ in their predominant genomic location (genes vs. transposable elements) and are regulated and maintained by different DNA methyltransferases, some of which have no analogs in animals (Cokus *et al.* 2008; Vanyushin and Ashapkin 2011). It is tempting to speculate that the more complex patterns of DNA methylation in plants may be related to their sessile nature and thus to their need for more plastic responses and a higher phenotypic adaptability.

Epigenetic silencing or activation of protein coding genes can be inherited through meiosis over several generations (reviewed in Jablonka and Raz 2009) and gives rise to so called epialleles (Schmitz *et al.* 2011). Spectacular is the case of a naturally occurring flower mutant of *Linaria vulgaris* that was described already 250 years ago by Carl Linnaeus and still persists today, only occasionally reverting to the wild type. Recently it was shown that the change in flower symmetry from bilateral to radial in this mutant is the result of extensive methylation leading to the transcriptionally silencing of a flower morphology control gene (Cubas *et al.* 1999). Hence epigenetic mutations may be transmitted for hundreds of generation (Cubas *et al.* 1999).

Besides that, various studies have shown that the amount and pattern of DNA methylation in plants is sensitive to biotic and abiotic stressors such as pathogens (Wadra *et al.* 2004), herbivores (Herrera and Bazaga 2013), drought (Labra *et al.* 2002), extreme temperatures (Boyko *et al.* 2010) or nutrient availability (Boyko *et al.* 2010; Kou *et al.* 2011). Moreover, some studies even have proved that environmental sensitive methylation changes, too, are stably transmitted over generations (Verhoeven *et al.* 2010; Kou *et al.* 2011) and can be directly correlated with adaptive plant responses (Boyko *et al.* 2010; Correia *et al.* 2013). In contrast to the classical Darwinian view, this implies that epigenetic variation may allow for the direct transgenerational transmission of acquired traits that can increase the fitness of future generations. Epigenetic variation thus may serve as an alternative to comparative slow genetic changes through mutation, drift or selection and may fill the gap between random genetic and environmental variation by allowing for the rapid conversion of environmental heterogeneity into phenotypic differences (Richards 2006; Flores *et al.* 2013).

The emerging field of epigenetic ecology that studies epigenetic processes in an ecological and evolutionary context (Bossdorf *et al.* 2008) achieved considerable progress in the last few years and greatly extended our knowledge on many aspects of population biology. It could be shown that epigenetic variation is indeed involved in diverse ecological important processes such as inbreeding-depression (Vergeer *et al.* 2012), invasion (Rich-

ards *et al.* 2012) or plant-animal interaction (Herrera and Bazaga 2013) and that methylation variation can predict regional and intraspecific functional diversity (Latzel *et al.* 2013; Medrano *et al.* 2014). Overall, epigenetic diversity may have similar functional consequences as other levels of natural biodiversity and hence needs to be incorporated into basic ecological research (Latzel *et al.* 2013).

Soil seed banks

A further mechanism in plants to cope with variation of environmental conditions and to reduce the risk of failure is the dispersal of seeds in space and time. While the spread of seeds across multiple localities may be achieved by dispersal through wind, water or animals, the spread of offspring through time requires seed dormancy or the absence of appropriate germination cues (e.g. light or water). Seed dormancy is a block to the completion of germination of an intact viable seed under favorable conditions (Finch-Savage and Leubner-Metzger 2006) and across species a variety of dormancy mechanisms have evolved that may be determined by both, morphological and physiological seed properties (Baskin and Baskin 2001).

Non-germinating seeds over time may accumulate in the ground and build up a reservoir of viable seeds. Such soil seed banks are common in plants across a wide range of life history types, habitats and climate zones (Leck *et al.* 1989; Baskin and Baskin 2001) and are classified according to their seed longevity. Transient seed banks contain seeds that persist in the soil for less than one year, seeds in short-term persistent seed banks persist for at least one year but less than five years, and seeds of long-term persistent seed banks survive for at least five years and in some species may be stored viable in the ground for many decades or even hundreds of years (Leck *et al.* 1989; Thompson 2000). Generally, soil seed banks can play an integral role in diverse ecological contexts both at the community level, affecting the composition and dynamics of species, and at the population level, counteracting the consequences of environmental or demographic stochasticities. Particularly for rare and isolated species or for species from highly dynamic or disturbed habitats, persistent seed banks might guarantee long-term survival and population stability, by enabling the replacement of aboveground individuals after bottlenecks or extinction events (Hölzel and Otte 2004; Honnay *et al.* 2008).

Additionally, theoretical models predict that seed banks have an important impact on population genetic processes as they can consist of progeny produced in the course of several generations and probably under varying selection regimes (Templeton and Levin 1979; Tonsor 1993). Consequently, with increasing longevity of seeds, the genetic diversity within the seed bank might exceed that of aboveground populations (Templeton and Levin 1979). This may enable gene flow from past generations and hence could preserve genes within populations that are selected against at the aboveground level (Tonsor *et al.* 1993). Accordingly, seed banks might dampen deleterious effects of fragmentation or decreasing population size and may buffer against genetic drift and population differentiation in the course of environmental change (McCue and Holtsford 1998). Moreover, they

could have the potential to slow down microevolutionary processes by compensating effects of directional selection.

While empirical research clearly could show that persistent soil seed banks can increase effective population size, both in annuals (Lundemo *et al.* 2009; Hanin *et al.* 2013) and perennials (Falahati-Anbaran *et al.* 2011), so far it was impossible to conclusively confirm their potential to accumulate genetic diversity (Honnay *et al.* 2008; Mandák *et al.* 2012). Even though some studies detected higher seed bank than aboveground genetic diversity (McCue and Holtsford 1998; Morris *et al.* 2002), a meta-analysis of seed bank genetic studies could not generalize these findings (Honney *et al.* 2008). Honney *et al.* (2008) concluded from their results that persistent seed banks may protect species against genetic drift and can buffer the differentiation of population but that there are no substantial differences in genetic diversity between seed bank and aboveground individuals. However, their finding may be related to the fact that most of the analyzed studies relied on comparatively small data sets from only one or few populations with no information about history and age. Therefore, the authors recommended not to continue surveying the genetic diversity of the two groups, unless this is performed under different selection regimes, in order to compare the outcome of the selection process (Honnay *et al.* 2008).

Study species

The perennial hemicryptophyte *Viola elatior* Fries belongs to the genus *Viola,* subsection *Rostratae* within the Violaceae (Eckstein *et al.* 2006a). The genus consists of 525–600 species that are found mostly in temperate habitats of the Northern Hemisphere but are also present in higher elevations of mountain systems near the equator and in the Southern Hemisphere (Ballard *et al.* 1999). The subsection *Rostratae* comprises about 35 species from Europe, Asia and North America that are characterized by leafy stems and hooked, rostrate styli. The subsection consists of tetraploid, octoploid, and (sub-)dodecaploid species (Eckstein *et al.* 2006a). *Viola elatior* has an octoploid genome (2n = 40).

The distribution of *V. elatior* roughly covers the submeridional and temperate zone of western Eurasia ranging from the Parisian basin in the west to southern Siberia as far as to Lake Baikal in the east, in the southeast reaching the Chinese province of Xinjiang (Meusel *et al.* 1978). Whereas in its core area with summer-warm continental climates the species is found in steppe and forest-steppe vegetation, in Central Europe towards the western border of its distribution (Figure 1.1), *V. elatior* is strictly confined to alluvial habitats within large river corridors that are characterized by strong fluctuations of the groundwater level and a high variability in soil water potential (Eckstein *et al.* 2006a; Danihelka *et al.* 2009). Here it becomes increasingly rare and occurs in a range of floodplain habitats along a successional gradient, extending from late mown or abandoned oligotrophic to mesotrophic Molinion caeruleae meadows to nitrophilous tall forb communities of the class Artemisietea vulgaris within or along hedges and alluvial hardwood forests (Eckstein and Otte 2005; Eckstein *et al.* 2006a). As the ability of *V. elatior* to compete for light is low (Moora *et al.* 2003), with increasing succession to closed forests, population sizes gradually decline and the species finally disappears from the above-

ground vegetation (Eckstein *et al.* 2006a). Overall, population sizes can vary between tens and hundreds of individuals.

Like many other species within the genus, *V. elatior* has a mixed mating system with potentially cross-pollinated chasmogamous (CH) and obligatory self-pollinated cleistogamous (CL) flowers. Whereas the open CH flowers with light-blue petals appear from early May to late June, the closed CL flowers that generally do not develop petals start to emerge after cessation of CH flowering (Eckstein and Otte, 2005) and may be produced from June to October. Both flower types produce approximately equal numbers of seeds ranging between 20 and 40 per capsule (own observations). Nonetheless, seed production through CL flowers is dominating, resulting in very high selfing rates. In common garden experiments only around 4% of total capsule production consisted of CH capsules (Eckstein and Otte, 2005). Overall, the cleistogamous breeding system may be interpreted as a “fail-safe” or “bethedging” strategy that optimizes the reproductive output in fluctuating environments either through potentially variable (CH) or relatively invariable (CL) offspring (Matilla and Salonen, 1995; Eckstein *et al.* 2006a). Seed dispersal is primarily achieved by a ballistic mechanism that ejects the seeds up to 3 m away from the capsules when the valves dry out (Eckstein *et al.* 2006a). Furthermore the seeds of *V. elatior* bear a small elaiosom that may lead to secondary dispersal by ants.

Viola elatior shows a seasonal germination pattern with a germination peak in April and May and a strong dormancy during the summer time. The species builds up persistent soil seed banks that may lead to high seed accumulation in the ground. Hölzel and Otte (2004) found maximum seed densities of up to 2660 germinable seeds/m^2 under a densely populated floodplain meadow, with more than 80% of all seeds concentrated in the upper 5 cm of the soil layer. Especially under strongly fluctuating conditions of floodplain habitats, the seed bank seems to be an important part of the species’ life strategy, which is illustrated by various reports about sudden emergence of plants in the course of disturbance events after long-term absence from the aboveground vegetation (Eckstein *et al.* 2006a, and reference therein).

The study species is red-listed in Germany (category 2: highly endangered) and across Central Europe can be considered as endangered (IUCN category: EN). An increased knowledge on population genetic and epigenetic responses to habitat change may help to improve the conservation of this rare species.

Objectives and study questions

The main objective of this thesis was to study ecological and molecular mechanisms that are related to the adjustment, adaptation and persistence of plant populations in dynamic environments. Surveying populations of *V. elatior* along a successional gradient, the project overall intended to gain a deeper understanding of genetic and epigenetic processes in the course of changing selection regimes.

Specifically the project focused on the following objectives:

1. To develop and evaluate a data analysis approach for methylation-sensitive amplification polymorphisms (MSAP) that allows for more detailed conclusions in the context of epigenetic population studies.
2. To survey and compare the impact of genetic and epigenetic variation in regard to habitat-related population differentiation.
3. To test for differences of DNA methylation variation in the CG- and CHG-sequence context.
4. To investigate if persistent soil seed banks can maintain genetic diversity during environmental change and under decreasing population size.
5. To study if contrasting habitat types have an impact on the small-scale spatial genetic structure of seed bank and/or aboveground individuals?

Study area

For genetic and epigenetic population studies, plant samples were collected in the Upper Rhine floodplain south-west of Frankfurt am Main, Germany. For seed bank genetic analysis (Chapter 4) sampling was additionally conducted in the Thaya/Morava floodplain around Břeclav, Czech Republic.

The Upper Rhine Valley represents the western border of the distribution range of *V. elatior* (Figure 1.1) and is one of the species' strongholds in Central Europe (Hölzel 2003). Here, sampling took place in the nature reserves 'Kühkopf-Knoblochsaue' (49°49'N, 8°26'E) and 'Lampertheimer Altrhein' (49°36'N, 8°26'E) that are separated by approximately 25 km. The region is characterized by comparatively warm and dry climatic conditions with a mean temperature of 10.3 °C and a mean annual precipitation of 580 mm (Müller-Westermeier 1990). Fine grained, calcareous alluvial soils (> 60% clay) on top of sandy sediments are predominant (Böger 1991). Associated with the water level of the Rhine River there exist strong seasonal and interannual fluctuations of the groundwater level, generally leading to very wet conditions during winter and spring, and notably dry conditions during summer (Donath *et al.* 2003). The area is densely populated and strongly fragmented through settlements, roads and arable land. In consequence, the study species today is only found in protected nature reserves that, due to continuous management by mowing or grazing, provide rather high proportions of early- and mid-successional habitats.

In contrast, large parts of the Thaya/Morava region in the Czech Republic are less influenced by settlements and intense land-use, and the landscape is characterized by a higher percentage of near-natural alluvial forests and non-intensively managed patches of floodplain meadows. Here, populations of *V. elatior* are much wider scattered and mostly

occur in late-successional habitats within forest stands or along forest fringes (Eckstein *et al.* 2006b). Sampling took place in a larger area than in the Upper Rhine Valley with a maximum distance of around 70 km between populations. However, most study sites were situated within a radius of 15 km around the city of Břeclav (48°45'N, 16°53'E).

Due to its location close to the Austrian border the area was behind the "Iron-curtain" for many years and today is one of the most preserved and most extensive regions of floodplain communities in Central Europe (Maděra *et al.* 2011). Notwithstanding, the natural dynamics of the floodplain were significantly affected in the 1970s and 1980s by river channel modification and the construction of several large water reservoirs that resulted in a marked reduction of flooding events (Horsák *et al.* 2009). The region is characterized by a dry and warm climate and is an extension of the Pannonian Basin, hosting many thermophilous and highly endangered species (Horsák *et al.* 2009). Annual precipitation averages 490 mm and average daily temperature is 9.3 °C (Šebesta *et al.* 2012). The area is situated in the transition zone between the typical river corridor distribution of *V. elatior* (Figure 1.1) and a more continental distribution not connected with large rivers (Danihelka *et al.* 2009).

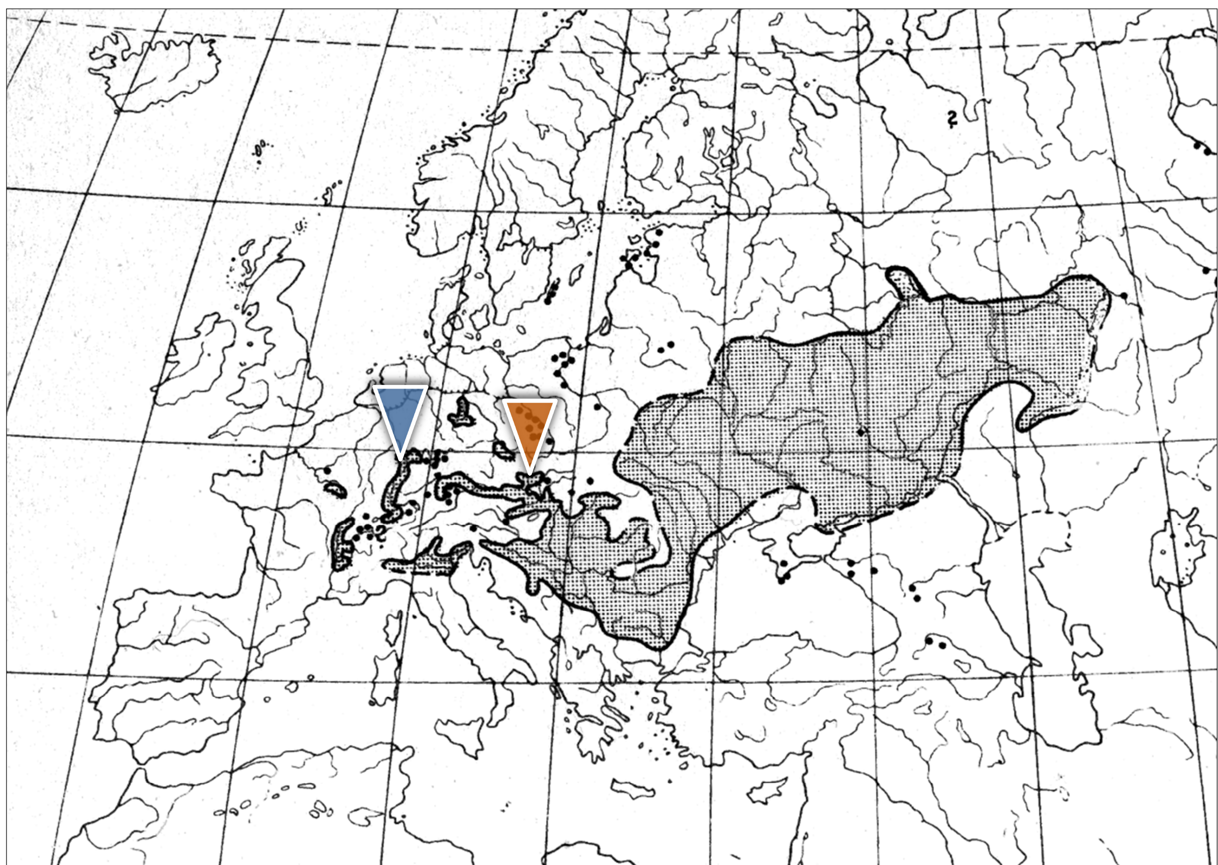


Figure 1.1 Distribution range of *Viola elatior* in Europe adapted from Meusel *et al.* (1978). Triangles indicate the two surveyed study regions: the Upper Rhine floodplain south of Frankfurt am Main, Germany (blue color) and the Thaya/Morava floodplain around Břeclav, Czech Republic (brown color).

Chapter outline

This thesis comprises three manuscripts that are presented in a chronological as well as logical order. Two of them have been published in peer-reviewed international scientific journals, whereas the third one has been submitted and is currently under review. This section gives a short outline of the contents of the manuscripts and provides a brief overview of the applied methods. The main findings and conclusions are presented in the following section.

Chapter 2: *Scoring and analysis of methylation sensitive amplification polymorphisms for epigenetic population studies*

Methylation sensitive amplification polymorphism (MSAP) analysis is a powerful tool to study DNA methylation variation in non-model species without detailed DNA sequences information. Moreover, as the method is comparatively low priced it allows to survey large numbers of individuals and thus enables to study ecological epigenetics at the population level. However, the scoring and interpretation of multistate MSAP information is complex and in the last years many different data scoring approaches have been employed in the literature. In this manuscript I reviewed the previously used MSAP scoring approaches and developed new alternatives. I then assessed effects of different scorings on parameters of epigenetic diversity and differentiation by testing them on a dataset of three populations of *V. elatior*. Overall, the objective of this study was to develop and justify a common and comparable framework for MSAP analysis that subsequently could be used in the epigenetic population study presented in Chapter 3.

Chapter 3: *Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb*

To gain a deeper understanding of the interplay and impact of genetic and epigenetic variation in the course of environmental changes, I compared six populations of *V. elatior* with amplified fragment length polymorphism (AFLP) and methylation sensitive amplification polymorphism (MSAP) markers. Three populations each were sampled in the Upper Rhine floodplains at the two extremes of a successional gradient, i.e. sunny floodplain meadows and shady alluvial woodland fringes. To obtain estimates of the light availability within populations, mean transmitted photosynthetic active radiation (PAR) per site was assessed with hemispherical photography. For MSAP markers I applied a newly developed scoring approach to separate the information of unmethylated and methylated fragments and to investigate the particular impact of methylations in two different sequence contexts. I then tested for differences in diversity and differentiation at the genetic and epigenetic level. To evaluate the relationships between geography and habitat type and genetic and epigenetic differentiation, I conducted pairwise and partial Mantel tests, applied analyses of molecular variance (AMOVA) and depicted the data with principal coordinate analysis (PCoA). Furthermore, correlation based genome scan analyses were applied to identify genetic or epigenetic markers that are correlated with site specific PAR

estimates. Finally, to obtain information on the degree of positive selection at the genetic level, I used two complementary differentiation based genome scan approaches for the AFLP data.

Chapter 4: *Can persistent seed banks buffer genetic effects of declining population size and selection?*

In this manuscript I explored the potential of persistent soil seed banks to buffer plant populations against the detrimental genetic effects that might be associated with environmental changes. I therefore compared genetic variation of aboveground and seed bank derived individuals (hereafter called cohorts) in populations of *V. elatior* along a successional gradient. Again, samples were collected at the two extremes of the species' environmental range in sunny floodplain meadows and shady alluvial woodland fringes. To allow for general conclusions that are independent of geographic location and to improve the statistical power, seven populations were sampled in the Upper Rhine region and eight in the Thaya/Morava floodplains. Genetic variation was assessed with AFLP markers. Due to unequal sample size of cohorts, ranging from 12 to 23 individuals, diversity estimates were calculated with a rarefaction approach. Genetic differentiation of populations was investigated with an AMOVA as well as with principal component analysis (PCA). Moreover, the small-scale spatial genetic structure within habitat types and aboveground and seed bank cohorts was examined using spatial autocorrelation methods.

Main results and conclusions

Evaluation of MSAP data scoring approaches

MSAP analysis is a modification of AFLP analysis, basically relying on the use of restriction enzymes that cut DNA in fragments of different length. Comparing the resulting fragment patterns among individuals subsequently allows to estimate genome-wide diversity and differentiation values.

MSAP uses the same rare cutter *Eco*RI as AFLP analysis, but substitutes the frequent cutter *Mse*I in two parallel runs by the isoschizomers *Hpa*II and *Msp*I. These two enzymes both cleave 5'-CCGG sequences, but differ in their sensitivity to the cytosine methylation status. Whereas both cut when the restriction site is unmethylated, *Hpa*II only cuts when the outer cytosine is hemimethylated and *Msp*I only cuts when the inner cytosine is hemi- or fully-methylated. Comparing *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I fragment profiles thus allows to detect four particular methylation conditions of the restriction site: (I) methylation absence, (II) hemi- or full-methylation in the CG-context, (III) hemimethylation in the CHG-context and (IV) any other possible methylation or fragment absence due to genetic restriction site polymorphism (Salmon *et al.* 2008). Condition IV represents an uninformative state and thus generally is excluded from the analysis.

In a final step, to enable the computation of descriptive indices, the multistate MSAP data needs to be scored and transformed into binary data (i.e. score '1' and score '0'). Sur-

veying the available literature revealed that for epigenetic population studies at least 5 different scoring approaches have been applied that can be categorized into three major groups: *Methylation Scoring*, *Non-Methylation Scoring* and *Mixed Scoring*. Briefly, in *Methylation Scoring* condition II and III are both scored as ‘1’, in *Non-Methylation Scoring* only condition I is scored as ‘1’ and in *Mixed Scoring*, between one and three markers are created for each epilocus of the multistate raw data matrix, scoring methylated and unmethylated fragments (condition I, II and III) separately. Furthermore, for each of the three main MSAP scoring approaches, additional criteria can be used to score the different types of methylation variation (for details see Chapter 2, Table 2.1). Together with two newly developed *Mixed Scoring* variants, overall eight different scoring schemes were applied to the data of three *V. elatior* populations.

MSAP analysis of this test data set revealed a total of 168 polymorphic loci in the multistate epigenetic raw data matrix. After transformation according to the eight scoring schemes, the number of final polymorphic markers widely differed, ranging from 78 to 286. Overall, estimated epigenetic diversity across populations showed the highest and lowest values for *Methylation* and *Non-Methylation Scoring*, ranging for percentages of polymorphic loci from 8.0% to 12.7% and for Shannon’s diversity index (H'_{epi}) from 0.22 to 0.36. This suggests that the large differences in epigenetic diversity that were found in some of the original publications using *Methylation* (e.g. Herrera and Bazaga 2010, mean H'_{epi} = 0.45) and *Non-Methylation Scoring* (e.g. Lira-Medeiros *et al.* 2010, mean H'_{epi} = 0.05) might at least partly be attributed to the scoring procedures used.

As depicted by PCoA and AMOVA also population differentiation differed between scoring approaches. Generally, clustering of individuals within populations was less condensed for *Methylation Scorings* than for other approaches and in all but one *Methylation Scoring* approach a greater amount of molecular variation was accounted for by differences among populations than by differences within populations. Overall, ϕ_{ST} values were very high ranging from 0.47 to 0.74. 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.

Related solely to the results of the case study, there seems to be not one best scoring approach for multilocus analyses. Notwithstanding, for single-locus analyses like genome scans or locus-by-locus AMOVA, both, information on methylated as well as unmethylated fragments seem to be important and probably could give different insides into population epigenetic processes. Moreover, pure *Methylation Scoring* assembles condition II and III fragments into one score, neglecting the fact that methylation in the CG- and CHG-context is catalyzed by different enzymes (Furner and Matzke 2011) and hence underlies different regulation. Thus, also the two methylation types might potentially account for different, probably counteracting epigenetic effects and a combined scoring could blur the real, effective epigenetic pattern. In conclusion, *Mixed Scoring* approaches, like the newly developed “Mixed Scoring 2” that generates the final epigenetic data matrix by transforming the three discernible methylation states at each multistate epilocus into separate binary subepiloci, overall seem to be favorable. Only in this way, the ambiguous

functional role of differently methylated and unmethylated fragments can be assessed and thus will allow for more detailed conclusions.

Indeed, the first preceding studies that have used "Mixed Scoring 2" for their MSAP analysis clearly could show functional and ecological important differences for unmethylated as well as for CG- and CHG-methylated fragments, both in plants (e.g. Schulz *et al.* 2014; Medrano *et al.* 2014) and animals (Wenzel and Piertney 2014). Moreover, also the results of a recent study that compared MSAP markers and global cytosine methylation supported the new scoring strategy (Alonso *et al.* 2015) and hence further illustrate the need for differentiated MSAP scoring.

Population epigenetics in changing environments

Epigenetic variation has been hypothesized to allow for rapid responses of plant populations to biotic or abiotic alterations (Boyko and Kovalchuk 2011) and thus to serve as an alternative to adaptations at the genetic level. However, the few available studies that investigated epigenetic variation in natural populations generally found a correlation between epigenetic and genetic variation suggesting that both types of variation at least partly depend on the same driving forces (Herrera and Bazaga 2010; Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012; Wu *et al.* 2013). But still, some of these studies also revealed a more close alignment of epigenetic differentiation to environmental differences than genetic differentiation (Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012) and thus indeed pinpoint to an important role of epigenetics in habitat adjustment.

So far the relationship between environmental differences and epigenetic variation has been only investigated in temporally rather stable habitats that persisted for long periods of time (e.g. habitats with differences in salinity, heavy metal content or altitude) and thus could have allowed for extensive adaptations at the genetic level. This complicates the study of independent epigenetic mechanisms. Hence, epigenetic population studies surveying dynamic and fast changing systems may allow for more detailed conclusions. However, directly testing epigenetic adaptation in natural populations is generally challenging, as the adaptive value of epigenetic variation is not easy to prove in the presence of genetic variation. Therefore, here an indirect approach was chosen, comparing the population structure of *V. elatior* at the genetic and epigenetic level and applying genome scan approaches to assess the putative contribution of genetic and epigenetic variation to environmental adaptation.

Overall, AFLP and MSAP analyses revealed comparatively low levels of genetic (H'_{gen} = 0.19) and epigenetic (H'_{epi} = 0.23) diversity and high genetic (ϕ_{ST} = 0.72) and epigenetic (ϕ_{ST} = 0.51) population differentiation in the surveyed populations. This comparatively low genetic diversity is consistent with earlier studies on *V. elatior* (Eckstein *et al.* 2006b) and with findings in other plants with predominant CL seed production (Durka *et al.* 2012). Besides very high inbreeding rates, also factors like spatial isolation and past population bottlenecks may have reinforced the observed pattern.

Even though epigenetic diversity generally tended to be higher than genetic diversity, both estimates were significantly correlated. Other studies surveying nonclonal species

likewise observed equal or higher epigenetic than genetic diversity (Herrera and Bazaga 2010; Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012; Wu *et al.* 2013), suggesting that this is a common pattern in genetically diverse plant species.

The two surveyed habitat types, floodplain meadows and alluvial woodland, had no profound effect on genetic or epigenetic diversity and only very small population size in one case markedly reduced diversity. However, although there was a positive correlation between epigenetic and genetic distances, epigenetic population differentiation was markedly reduced as compared to genetic differentiation. In addition, hierarchical AMOVA and partial Mantel tests revealed that epigenetic differentiation was overall more closely related to habitat conditions. This indicates that environmentally induced changes in methylation patterns lead to a convergence of populations experiencing similar habitat conditions and thus may counteract effects of historical demographic processes.

The use of the newly developed "Mixed Scoring 2" approach for MSAP analysis allowed to test for the particular impact of methylation in the CG- and CHG-sequence context. Strikingly, as depicted by PCoA, CG-methylation information separated the habitat types, whereas information of CHG-hemimethylation revealed hardly any population structure. This implies that methylation in the CG-context plays a more important role for habitat adjustment than changes of hemimethylation in the CHG-context.

Correlating genetic and epigenetic markers with site specific light availability using a Spatial Analysis Method (Joost *et al.* 2007) resulted in comparable percentages of light related outlier markers for genetic (17.0%) and epigenetic (14.2%) data. This principally suggests that both levels of molecular variation may have an important role for habitat adaptation. Notwithstanding, as revealed by differentiation based genome scan approaches for the genetic data, only very few light-related AFLP outliers (2 of 19) actually seemed to be under positive selection. Moreover, the percentage of genetic outliers was strongly reduced with differentiation based genome scans, ranging between 0 and 4.5% and thus generally questioning a large impact of genetic selection in *V. elatior*. On the contrary, depicting and comparing light-related AFLP and MSAP outliers with PCoA strongly supported the hypothesis that methylation variation plays the major role in response to habitat conditions. Whereas for neutral markers, genetic and epigenetic population structure were almost identical, for outlier markers only epigenetic variation led to a close clustering of individuals from the same habitat type.

Interestingly, for light-related MSAP markers most unmethylated outliers were related to low light and most CG-methylated outliers to high light, indicating a directional epigenetic relay mechanism that activates stress-related genes by demethylation under low light conditions and downregulates them by methylation under high light conditions. Indeed, it is well known that biotic and abiotic stressors can induce selective demethylation processes and transcriptional activation of stress-related genes (Wada *et al.* 2004; Choi and Sano 2007) and that global hypomethylation may be a response to environmental stressors or different habitat conditions (Lira-Medeiros *et al.* 2010; Wu *et al.* 2013). Moreover, also signals of foliage shade have been correlated with genome hypomethylation and were shown to be a crucial factor for stem elongation, probably triggering shade-avoidance responses (Tatra *et al.* 2000).

Aboveground and seed bank genetics in changing environments

Genetic analyses of aboveground (AG) and seed bank (SB) cohorts of *V. elatior*, generally corroborated the results of the smaller data set in the former study and revealed very low levels of within-population genetic diversity, both in Germany and the Czech Republic. Mean values over populations and cohorts ranged for band richness from 1.17 (AG) to 1.15 (SB), and for percentage of polymorphic loci from 20% (AG) to 18% (SB).

Comparing genetic diversity between regions or between habitat types revealed no significant differences. Specifically, the absence of any differences in diversity between the two contrasting habitat types was surprising, as it is assumed that decreasing population sizes along successional gradients may lead to a loss of genetic variation through effects of increased random genetic drift, higher inbreeding rates and the accumulation of deleterious mutation (e.g. van Treuren *et al.* 1991; Young *et al.* 1996). Moreover, the changing conditions could result in an increased probability of local extinction of certain genotypes due to selection and thus may further aggravate the loss of genetic variation (Raffl *et al.* 2006). However, while there were no differences in AG or SB genetic diversity, the direct relationship between AG and SB cohorts clearly differed between the two habitat types. Whereas populations from floodplain meadows overall exhibited significantly higher AG than SB diversity, no differences could be detected in populations from alluvial woodland. This strongly implies that the relative SB genetic diversity (i.e. compared to AG cohorts) increases with ongoing succession and despite decreasing population size. Moreover, in three of eight woodland populations the SG genetic diversity even exceeded AG diversity.

But what is driving this change and how can the relative seed bank genetic diversity increase towards late successional stages whereas population size gradually declines? The most likely explanation seems to be related to the mixed mating system of *V. elatior*. Accordingly, the contribution of outcrossed CH seeds to reproduction might increase from early to late successional stages. Hence, under favorable early successional conditions with high plant densities, populations would maintain their approved genotypes mainly through selfed CL seeds, resulting in a depletion of SB genetic diversity. Contrary, under more unfavorable conditions an increase in outcrossing would increase relative SB genetic diversity and in turn might compensate the detrimental effects of small population size and keep AG genetic diversity at a constant level.

Several non-mutually exclusive and linked processes could account for an increased outcrossing in woodland habitats. First, late successional conditions might favor higher CH/CL capsule ratios. Indeed, in other cleistogamous species increasing CH/CL ratios were correlated with decreasing plant density (Cheplick 2007), decreasing light availability (Mattila and Salonen 1995; Cheplick 2007) and increasing soil water availability (Brown 1952). Second, as CH capsules mature in the hottest period of the year (i.e. June–July) lower water availability in meadow habitats may result in lower CH seed quality or higher CH seed abortion rates. Third, also anthropogenic effects could have decreased CH seed contribution in floodplain meadows, as most of them are regularly managed by a one-time mowing in the time of CH capsule ripening in early June.

An alternative and/or complementary explanation to a changed CH/CL ratio might be that seed longevity generally is higher in woodland habitats. As soil parameters like moisture and temperature are more constant and balanced than in grassland, woodland seed banks might be assembled from more seed generations and in consequence would exhibit a higher relative genetic diversity.

In agreement with the very low within-population genetic diversity, genetic differentiation in *V. elatior* was very high, with 80.1% (AG) and 83.5% (SB) of genetic variation residing among populations. Overall, differentiation was virtually identical among SB and AG samples at all levels of comparison (i.e. among regions and among habitats). This is in contrast to some previous seed bank genetic studies that found lower differentiation among SB than among AG cohorts (McCue and Holtsford 1998; Zhaghloul *et al.* 2013). It was hypothesized that by chance or selection, AG cohorts can become more differentiated than their potentially multigenerational and hence overall more homogeneous seed pools (McCue and Holtsford 1998). Even though this generally seems to be not the case in *V. elatior*, some populations indeed showed significant differentiation between SB and AG cohorts. Interestingly, cohort differentiation was more frequent in woodland than in meadow habitats (4/8 vs. 2/7, respectively), further corroborating changes in the interplay of seed bank and aboveground individuals along the gradient.

To test for differences in outcrossing rates between the two habitats, besides the overall population structure additionally the small-scale spatial genetic structure (SGS) within each population was analyzed. SGS arises due to spatially restricted gene dispersal and is mainly related to the amount of gene flow by seeds and pollen (Zeng *et al.* 2011). Thus, as dispersal distance of seeds should be rather comparable between habitats, any difference in SGS can be expected to reflect differences in pollen dispersal and hence in outcrossing rates. Strikingly, SGS analysis corroborated the postulated increase of outcrossing towards woodland habitats and revealed significant differences along the successional gradient. Whereas meadow populations showed comparative high SGS that even exceeded values reported for other predominantly selfing species, SGS in woodland populations was markedly lower and tended to reflect more those for mixed mating species (Vekemans and Hardy 2004).

Even though SGS was generally lower in SB than AG cohorts, overall there were no significant differences between both groups. Similarly, most of the few other available studies on that topic also detected significant SGS in both cohorts (Shimono *et al.* 2006; Ottewell *et al.* 2011). This suggests that any past SGS within the seed bank appears to break down relatively fast and is not stored over longer periods of time. Otherwise much stronger SGS should have been present in SB cohorts of woodland populations. Thus, the absence of significant differences between cohorts indicates that in *V. elatior* the longevity of soil seeds and adult plants largely overlaps and that most seeds do not persist for more than few adult generations (Tonsor *et al.* 1993). Notwithstanding, in the case of the study species this time span seems to be sufficient to have a sustainable impact on aboveground population genetics.

Conclusions & Outlook

Overall, the results of this study could show that both, epigenetic variation and the presence of persistent soil seed banks may play a decisive role in the adjustment and adaptation of plants to environmental heterogeneity.

Under fast changing conditions the environmental shaping of the epigenome seems to be a stronger force than selection changing the genome. This suggests that epigenetic diversity might be more important for short-term responses to environmental fluctuation than genetic diversity. Provided that genetically independent adaptive epigenetic variation is heritable, 'soft inheritance' thus indeed could represent an alternative system to classical 'hard inheritance'. Especially for rare and inbreeding species like *V. elatior* that suffer from spatial isolation and small population sizes, methylation variation could facilitate long-term population survival even in the absence of extensive genetic diversity.

Likewise, also persistent seed banks seem to counteract the detrimental effects that can be related to environmental change. The observed increase in relative seed bank diversity from meadow to forest habitats implies that an increased genetic buffer capacity of the seed bank dampens the decrease of genetic diversity towards late successional habitats. Ultimately, this will also improve population recovery after extinction events that are much more likely in the course of increasing canopy closure than under favorable early successional conditions.

Taken together, both mechanisms might play a complemental role in enhancing the chances for population persistence under fluctuating conditions. Whereas epigenetic variation allows populations to respond on rather short time scales, persistent seed banks may keep population genetic diversity on a constant level, and thus, over the long term also enable for more slow responses at the genetic level. As epigenetic variation in turn at least partly is relying on the underlying genetic variation, both levels of diversity seem to interact in a complex manner, illustrating once more the convoluted interdependencies in biological systems.

As it is often the case, the findings of this study could answer some questions but in turn raised many new ones. Do the identified differences in CG- and CHG-methylation variation represent a general epigenetic pattern under contrasting environmental conditions? Do genetically more diverse species, too, show a stronger epigenetic than genetic correlation to habitat differentiation? Is an increased relative seed bank diversity towards late successional habitats a specific situation in cleistogamous plants?

Hence, further studies are needed that use mixed scoring approaches for MSAP analysis and survey also species that exhibiting other mating systems and a higher overall genetic diversity. Moreover, also combining both study fields would be very interesting, investigating if persistent seed banks could accumulate epigenetic variation, too. In addition, raising seed bank individuals from different habitats under a common greenhouse environment would allow to test if specific habitat related epigenetic changes are indeed transmitted over generations or contrary only represent a transient states that solely adjust to the prevailing conditions.

References

Abratowska A, Wasowicz P, Bednarek PT, Telka J, Wierzbicka M (2012) Morphological and genetic distinctiveness of metallicolous and non-metallicolous populations of *Armeria maritima s.l.* (Plumbaginaceae) in Poland. *Plant Biology*, 14, 586–595.

Alonso C, Pérez R, Bazaga P, Medrano M, Herrera CM (2015) MSAP markers and global cytosine methylation in plants: a literature survey and comparative analysis for a wild-growing species. *Molecular Ecology Resources*, early view DOI:10.1111/1755-0998.12426

Baskin CC, Baskin JM (2001) *Seeds: ecology, biogeography, and evolution of dormancy and germination*. Academic Press, San Diego, USA.

Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature*, 447, 407–412.

Böger K (1991) *Grünlandvegetation im Hessischen Ried - pflanzensoziologische Verhältnisse und Naturschutzkonzeption, Botanik und Naturschutz in Hessen*, Beiheft 3, Botanische Vereinigung für Naturschutz in Hessen e.V. (BVNH), Frankfurt a. M.

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

Boyko A, Blevins T, Yao Y *et al.* (2010) Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-Like proteins. *PLoS ONE*, 5, e9514.

Boyko A, Kovalchuk I (2011) Genome instability and epigenetic modification—heritable responses to environmental stress? *Current Opinion in Plant Biology*, 14, 260–266.

Brown WV (1952) The relation of soil moisture to cleistogamy in Stipa leucotricha. *Botanical Gazette*, 113, 438–444.

Cheplick GP (2007) Plasticity of chasmogamous and cleistogamous reproductive allocation in grasses. *Aliso: A Journal of Systematic and Evolutionary Botany*, 23, 286–294.

Choi C-S, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Molecular Genetics and Genomics*, 277, 589–600.

Cokus SJ, Feng S, Zhang X *et al.* (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452, 215–219.

Correia B, Valledor L, Meijón M *et al.* (2013) Is the interplay between epigenetic markers related to the acclimation of cork oak plants to high temperatures? *PLoS ONE*, 8, e53543.

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

Danihelka J, Niklfeld H, Šípošová H (2009) *Viola elatior*, *V. pumila* and *V. stagnina* in Austria, Czechia and Slovakia: a story of decline. *Preslia*, 81, 151–171.

Donath TW, Hölzel N, Otte A, Marrs R (2003) The impact of site conditions and seed dispersal on restoration success in alluvial meadows. *Applied vegetation science*, 6, 13–22.

Durka W, Nossol C, Welk E *et al.* (2012) Extreme genetic depauperation and differentiation of both populations and species in Eurasian feather grasses (Stipa). *Plant Systematics and Evolution*, 299, 259–269.

Eckstein RL, Hölzel N, Danihelka J (2006a) Biological Flora of Central Europe: *Viola elatior*, *V. pumila* and *V. stagnina*. *Perspectives in Plant Ecology, Evolution and Systematics*, 8, 45–66.

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

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.

Falahati-Anbaran M, Lundemo S, Agren J, Stenoien HK (2011) Genetic consequences of seed banks in the perennial herb *Arabidopsis lyrata* subsp. *petraea* (Brassicaceae). *American Journal of Botany*, 98, 1475–1485.

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.

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytologist*, 171, 501–523.

Flores KB, Wolschin F, Amdam GV (2013) The role of methylation of DNA in environmental adaptation. *Integrative and Comparative Biology*, 53, 359–372.

Furner IJ, Matzke M (2011) Methylation and demethylation of the *Arabidopsis* genome. *Current opinion in plant biology*, 14, 137–141.

Hanin N, Quaye M, Westberg E, Barazani O (2013) Soil seed bank and among-years genetic diversity in arid populations of *Eruca sativa* Miller (Brassicaceae). *Journal of Arid Environments*, 91, 151–154.

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 (2013) Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Botanical Journal of the Linnean Society*, 171, 441–452.

Hölzel N (2003) Re-assessing the ecology of rare flood-meadow violets (*Viola elatior*, *V. pumila* and *V. persicifolia*) with large phytosociological data sets. *Folia Geobotanica*, 38, 281–298.

Hölzel N, Otte A (2004) Assessing soil seed bank persistence in flood-meadows: The search for reliable traits. *Journal of Vegetation Science*, 15, 93–100.

Honnay O, Bossuyt B, Jacquemyn H, Shimono A, Uchiyama K (2008) Can a seed bank maintain the genetic variation in the above ground plant population? *Oikos*, 117, 1–5.

Horsák M, Bojková J, Zahrádková S, Omesová M, Helešic J (2009) Impact of reservoirs and channelization on lowland river macroinvertebrates: A case study from Central Europe. *Limnologica-Ecology and Management of Inland Waters*, 39, 140–151.

Jablonka 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.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16, 3955–3969.

Kou HP, Li Y, Song XX *et al.* (2011) Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* L.). *Journal of Plant Physiology*, 168, 1685–1693.

Labra M, Ghiani A, Citterio S *et al.* (2002) Analysis of cytosine methylation pattern in response to water deficit in pea root tips. *Plant biology*, 4, 694–699.

Latzel V, Allan E, Bortolini Silveira A *et al.* (2013) Epigenetic diversity increases the productivity and stability of plant populations. *Nature Communications*, 4:2875, doi: 10.1038/ncomms3875

Leck MA, Parker VT, Simpson RL (Eds.) (1989) *Ecology of soil seed banks*. Academic Press, San Diego, USA.

Lira-Medeiros CF, Parisod C, Fernandes RA *et al.* (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE*, 5, e10326.

Lundemo S, Falahati-Anbaran M, Stenoien HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Molecular Ecology*, 18, 2798–2811.

Maděra P, Šebesta J, Řepka R, Klimánek M (2011) Vascular plants distribution as a tool for adaptive forest management of floodplain forests in the Dyje river basin. *Journal of Landscape Ecology*, 4, 18–34.

Mandák B, Zákravský P, Mahelka V, Plačková I (2012) Can soil seed banks serve as genetic memory? A study of three species with contrasting life history strategies *PLoS ONE*, 7, e49471.

Mattila T, Salonen V (1995) Reproduction of *Viola mirabilis* in relation to light and nutrient availability. *Canadian Journal of Botany*, 73, 1917–1924.

McCue KA, Holtsford TP (1998) Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). *American Journal of Botany*, 85, 30–36.

Medrano M, Herrera CM, Bazaga P (2014) Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb. *Molecular Ecology*, 23, 4926–4938.

Meusel H, Jager E, Rauschert S, Weinert E (1978) *Vergleichende Chorologie der zentraleuropäischen Flora*. Gustav Fischer Verlag, Jena, Germany.

Moora M (2003) Responses of a rare (*Viola elatior*) and a common (*Viola mirabilis*) congeneric species to different management conditions in grassland—is different light competition ability responsible for different abundances? *Acta Oecologica*, 24, 169–174.

Morris AB, Baucom RS, Cruzan MB (2002) Stratified analysis of the soil seed bank in the cedar glade endemic *Astragalus bibullatus*: evidence for historical changes in genetic structure. *American Journal of Botany*, 89, 29–36.

Müller-Westermeier G (1990) *Klimadaten der Bundesrepublik Deutschland*. Deutscher Wetterdienst, Offenbach.

Ottewell KM, Bickerton D, Lowe AJ (2011) Can a seed bank provide demographic and genetic rescue in a declining population of the endangered shrub Acacia pinguifolia? *Conservation Genetics*, 12, 669–678.

Raffl C, Schönswetter P, Erschbamer B (2006) "Sax-sess"—genetics of primary succession in a pioneer species on two parallel glacier forelands. *Molecular Ecology*, 15, 2433–2440.

Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395–401.

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.

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.

Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334, 369–373.

Schulz B, Eckstein RL, Durka W (2014) Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Molecular Ecology*, 23, 3523–3537.

Šebesta O, Gelbič I, Minář J (2012) Mosquitoes (Diptera: Culicidae) of the lower Dyje river basin (Podyjí) at the Czech-Austrian border. *Open Life Sciences*, 7, 288–298.

Shimono A, Ueno S, Tsumura Y, Washitani I (2006) Spatial genetic structure links between soil seed banks and above-ground populations of *Primula modesta* in subalpine grassland. *Journal of Ecology*, 94, 77–86.

Tatra GS, Miranda J, Chinnappa CC, Reid DM (2000) Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of *Stellaria longipes*. *Physiologia Plantarum*, 109, 313–321.

Templeton AR, Levin DA (1979) Evolutionary consequences of seed pools. *American Naturalist*, 114, 232–249.

Thompson K (2000) The functional ecology of soil seed banks. In: *Seeds: the ecology of regeneration in plant communities* (ed Fenner M), pp. 215–235. CAB International, Wallingford, UK.

Tonsor SJ, Kalisz S, Fisher J, Holtsford TP (1993) A life-history based study of population genetic structure: seed bank to adults in *Plantago lanceolata*. *Evolution*, 47, 833–843.

Van Treuren R, Bijlsma R, Van Delden W, Ouborg N (1991) The significance of genetic erosion in the process of extinction. I. Genetic differentiation in *Salvia pratensis* and *Scabiosa columbaria* in relation to population size. *Heredity*, 66, 181–189.

Vanyushin BF, Ashapkin VV (2011) DNA methylation in higher plants: past, present and future. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1809, 360–368.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, 13, 921–935.

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.

Wada Y, Miyamoto K, Kusano T, Sano H (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. *Molecular Genetics and Genomics*, 271, 658–666.

Wenzel MA, Piertney SB (2014) Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, 23, 4256–4273.

Wu W-Q, Yi MR, Wang X-F *et al.* (2013) Genetic and epigenetic differentiation between natural *Betula ermanii* (Betulaceae) populations inhabiting contrasting habitats. *Tree Genetics and Genomes*, 9, 1321–1328.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution*, 11, 413–418.

Zeng X, Michalski SG, Fischer M, Durka W (2011) Species diversity and population density affect genetic structure and gene dispersal in a subtropical understory shrub. *Journal of Plant Ecology*, 5, 270–278.

CHAPTER 2

SCORING AND ANALYSIS OF METHYLATION SENSITIVE AMPLIFICATION POLYMORPHISMS FOR EPIGENETIC POPULATION STUDIES

Benjamin Schulz, R. Lutz Eckstein and Walter Durka
Molecular Ecology Resources 13 (2013) 642–653

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 multistate 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, e.g. 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.

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 and 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.* 2010; Lister *et al.* 2009), plants harbor 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 (Grativol *et al.* 2011; Chatterjee and Vinson 2012; Saze *et al.* 2012). Silencing or activation of protein-coding genes can be inherited through meiosis over several generations (reviewed in Jablonka and 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 and Dickinson 2005, 2006; Bock 2012). For model species, genome-wide profiles of DNA methylation are available at high resolution using microarray technologies and next generation sequencing of bisulfite converted DNA (Fujimoto *et al.* 2012; Bock 2012). However, due to high costs and resource intensities these methods usually are not suited for non-model 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 *Hpa*II and *Msp*I. 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 2.1), 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 and Perez-Figuera 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.* 2012).

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 and Bazaga 2010, 2011; Richards *et al.* 2012), vertebrates (e.g. Taylor *et al.* 2010; Massicote and Angers 2011; 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 2.1), 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 and Bazaga 2010; Lira-Medeiros *et al.* 2010).

Box 2.1 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 *Eco*RI and substituting the frequent cutter *Mse*I in two parallel runs by the more or less methylation sensitive restriction enzymes *Hpa*II and *Msp*I. 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; Lira-Medeiros *et al.* 2010; Paun *et al.* 2010, Herrera & Bazaga 2010; Richards *et al.* 2012). According to the actual specifications of the restriction enzyme database REBASE (http://rebase.neb.com, accessed 20.02.2013), *Hpa*II only recognizes sites that are hemimethylated at the external cytosine (^{HMe}CCG), while *Msp*I 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 *Hpa*II (Korch & Hagblom 1986; Butkus *et al.* 1987) and ^{HMe}CCG-sites by *Msp*I (Butkus *et al.* 1987), but since this is contradictory with later studies (see McClelland *et al.* 1994) it is not considered here.

The comparison of the resulting *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I fragment profiles allows the detection of particular methylation states of the restriction sites (Figure 2.1). In total, 4 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 *Eco*RI/*Msp*I profiles indicating ^{HMe}CG- or ^{Me}CG-sites, (iii) condition III = fragments are present only in *Eco*RI/*Hpa*II profiles indicating ^{HMe}CCG-sites and (iv) condition IV = complete absence of fragments in both profiles. The latter represents an uninformative state since 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.

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 paper 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.

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

Figure 2.1 Sensitivity of the isoschizomers *Hpa*II and *Msp*I 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 *Hpa*II/*Msp*I restriction site are indicated by orange color.

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 *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I profiles (Box 2.1) 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 (Figure 2.2). 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, hereaf-

ter 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 sub-loci 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 sub-epiloci 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 2.1).

Figure 2.2 Main groups of 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. *Eco*RI and *Hpa*II/*Msp*I restriction sites are accentuated. Methylated cytosines at the *Hpa*II/*Msp*I restriction site are indicated by orange color. Restriction site mutation is indicated by green color. NA denotes missing data.

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 and 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 $e_{Hpa} + e_{Msp} - 2e_{Hpa}e_{Msp}$ (e_{Hpa} = error rate of *Hpa*II profile, e_{Msp} = error rate of *Msp*I profile, $e_{Hpa}e_{Msp}$ = error rate of combined profiles). All loci with observed proportions of discordant *Hpa*II/*Msp*I scores suggestive of methylation (i.e. number of individuals with contrasting *Hpa*II/*Msp*I 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 and 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 done by Moran and 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 *Eco*RI/*Hpa*II and *Eco*RI/*Msp*II profiles are analyzed separately to assess the epigenetic and genetic structure, respectively. In essence, only the reduced *Eco*RI/*Hpa*II 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-Medeiros + *Hpa*II, 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}CCG-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}CCG-sites, scoring either condition II or III as "1", respectively. In (6) to (8), condition IV is scored "0".

Table 2.1 Scoring schemes of the eight scoring approaches used in the case study (NA denotes the treatment of condition IV as missing data)

*Hpa*II/*Msp*I 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 and ^{Me}CG	^{HMe}CCG	No information
METHYLATION SCORING				
1 Salmon *et al.* 2008	0	1	1	0
2 Vergeer *et al.* 2012	0	1	1	NA
3 Herrera and Bazaga 2010[a]	0	1*	1*	NA
NON-METHYLATION SCORING				
4 Lira-Medeiros *et al.* 2010#	1	0	Loci excluded	0
5 Lira-Medeiros + *Hpa*II	1	0	0	0
MIXED SCORING				
6 Paun *et al.* 2010				
unmethylated	1	0	0	0
^{HMe}CG and ^{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 and ^{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.

Case Study: Epigenetic diversity of a perennial violet

The eight scoring approaches are partly complementary but differ considerably with respect to the amount of 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 and 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 obligatory self-pollinated cleistogamous flowers. However, most seeds are produced by cleistogamous flowers (Eckstein and 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, i.e. only flowering individuals were sampled.

MSAP epi-genotyping

MSAP analysis was performed according to the protocol provided in Appendix 2.2 using eight selective primer combinations. The 67 samples were scored in one batch for presence "1" or absence "0" of fragments obtained with *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I resulting in a primary MSAP data matrix of 134 lines (Appendix 2.3). 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 *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I 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 2.1). Fourth, epiloci that turned out monomorphic after trans-

formation were excluded. Data transformation and selection of polymorphic loci was performed using the R script MSAP_calc (Appendix 2.4), which allows to perform the different scorings (see Pérez-Figueroa 2013 for an automated Methylation Scoring). These data matrices are also included in Appendix 2.3.

Data analysis

Data analysis of the binary epigenetic data matrices was performed using a marker based strategy, i.e. 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

$$H'_{epi} = -\sum p_i \log_2 p_i$$

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 and 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 and 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 and 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

MSAP 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 $^{\mathrm{HMe}}$CG- or $^{\mathrm{Me}}$CG-sites and 53 included fragments with $^{\mathrm{HMe}}$CCG-sites.

Table 2.2 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 + *Hpa*II	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.

After transformation to the 8 binary epigenetic data matrices the number of polymorphic epiloci ranged from 78 for the *Herrera-scoring* to 286 for the "*Mixed Scoring 2*" (Table 2.2). 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 2.2), 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 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 was highest for "site3" (Table 2.2). 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} (Figure 2.3) 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.

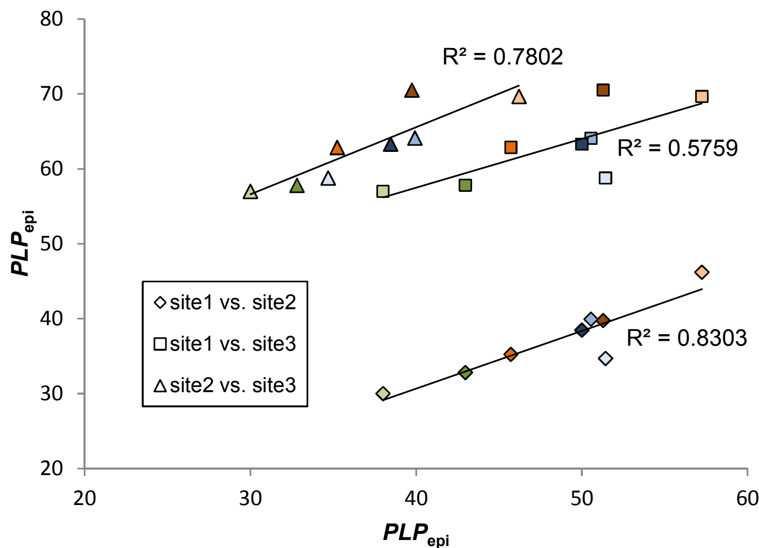


Figure 2.3 Linear regressions of percentage polymorphic epiloci (PLP_{epi}) obtained with different MSAP scoring approaches for three populations of *Viola elatior*. Colors refer to different scoring approaches (see Table 2.1).

Principal coordinates analyses (PCoA) of epigenetic distances in all cases clearly separated the three populations, forming well-defined clusters (Figure 2.4). Variation explained by 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.

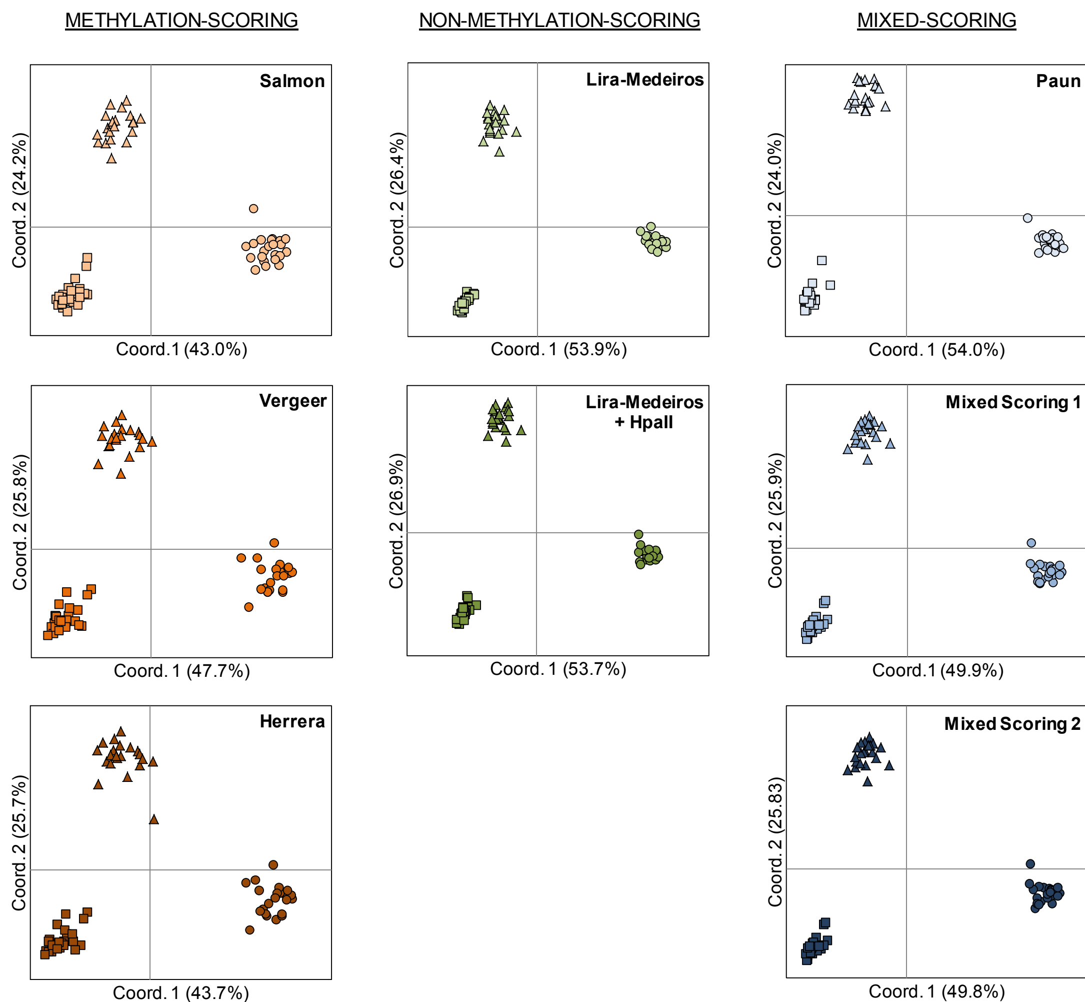


Figure 2.4 Principal Coordinates Analyses (PCoA) of square root transformed Nei and Li distances of epigenetic binary data matrices obtained with different MSAP scoring approaches (see Table 2.1). Populations of *Viola elatior* are indicated by symbols (○ = site1, □ = site2, Δ = site 3).

Strong epigenetic population differentiation was corroborated by the AMOVA (Table 2.3). 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 2.3), ranged between 50.3% for the *Salmon-scoring* and 66% for the *Lira-Medeiros-scoring*. In general, Mixed Scoring approaches revealed the highest number, but lowest percentage of significantly differentiated epiloci compared to Methylation and Non-Methylation Scoring. As expected, epigenetic differentiation was negatively correlated with epigenetic diversity (Figure 2.5) 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.

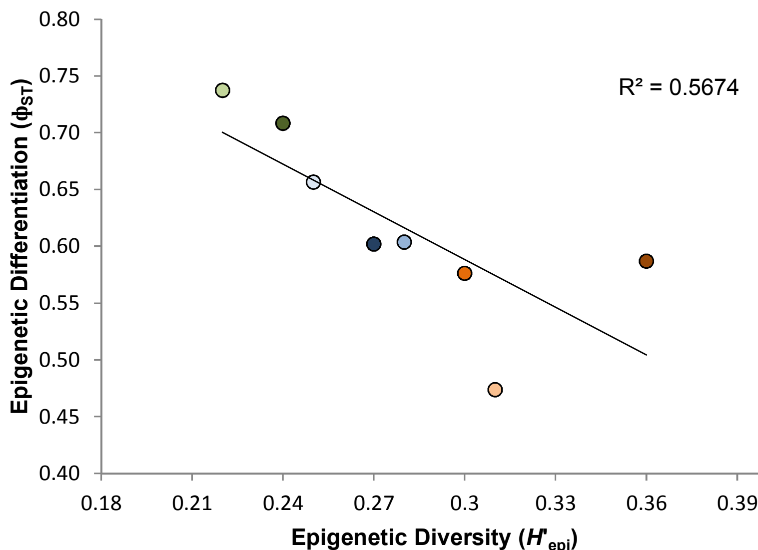


Figure 2.5 Linear regression of epigenetic diversity (H'_{epi}) and epigenetic phenotypic differentiation (ϕ_{ST}) obtained with different MSAP scoring approaches for three populations of *Viola elatior*. Colors refer to different scoring approaches (see Table 2.1).

Table 2.3 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 + *Hpa*II	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								
Differentiated loci	73	49	44	66	82	136	154	153
Differentiated loci (%)	50.3	60.5	65.7	66	64.1	55.5	56.4	53.7

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 ^{HMe}CG, ^{Me}CG and ^{HMe}CCG. All other possible methylation types which are not distinguishable from real fragment absence (condition IV) are ignored. For animal species, cytosine methylation in the CHG-sequence context are 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 and Ba-

zaga 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 are catalyzed by different enzymes (Matzke and Furner 2010) and thus underlie 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 and 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, since 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 favorable as it allows to test for both, the potential contribution of unmethylated and methylated fragments. Since 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, ^{HMe}CG- or ^{Me}CG markers and ^{HMe}CCG-markers). However, in the *Paun-scoring* unmethylated fragments are incorporated into the two methylated marker types which likely 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 and Bazaga 2010; Lira-Medeiros *et al.* 2010) have applied certain MSAP scoring variants to assess genetic variation. Whereas Herrera and 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 *Msp*I 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 methyla-

tion 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 by using the two isoschizomers *Acc65*I and *Kpn*I 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 *Kpn*I 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 assess 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}, i.e. between the *Herrera-scoring* (H'_{epi} = 0.36) and the *Lira-Medeiros-scoring* (H'_{epi} = 0.22). Thus, the large difference found in epigenetic variation in the original publications between *Viola cazorlensis* (H'_{epi} = 0.45; Herrera and Bazaga 2010) and *Laguncularia racemosa* (H'_{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 ϕ_{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

MSAP 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 functionalrole of differently methylated and unmethylated fragments that might have both, activating or downregulating 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, 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 C, Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters*, 11, 106–115.

Butkus V, Petrauskiene L, Maneliene Z *et al.* (1987) Cleavage of methylated CCCGGG sequences containing either N4-methylcytosine or 5-methytcytosine with Mspl, Hpall, Smal, Xmal and Cfr9I restriction endonudeases. *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, 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.

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.

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 and 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.

Grativol 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 and 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.

Jablonka 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.

Liu S, Sun K, Jiang T *et al.* (2012) Natural epigenetic variation in the female great roundleaf bat (*Hipposideros armiger*) populations. *Molecular Genetics and Genomics*, 287, 643–650.

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 (*Dactylorhiza*: 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 annuum* 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, Biniszkiewicz 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 and 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, Quarin 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, Bellis H, Chable V, Manzanares-Dauleux MJ (2009) Identification of differentially expressed genes related to aberrant phenotypes in *Brassica oleracea* var. *botrytis*. *Plant Breeding*, 128, 631–639.

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.

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, 4407–14.

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.

Appendix

Appendix 2.2 MSAP genotyping protocol

Total genomic DNA was extracted from dried leaf tissue using the DNeasy 96 Plant extraction kit (QIAGEN). The MSAP-analysis was based on a standard AFLP protocol (Vos *et al.* 1996), replacing the *Mse*I enzyme in two separate runs with the methylation-sensitive enzymes *Hpa*II and *Msp*I using appropriate adaptors and primers (Supplementary Table S2.1). For restriction and ligation (RL) 5.2 µl genomic DNA (ca. 130 ng) were combined with 5.8 µl RL reaction mix containing 0.55 µl BSA (1 mg/ml; New England Biolabs, NEB), 1.1 µl 0.5 M NaCl, 5 u *Eco*RI (NEB), 5 u *Hpa*II or *Msp*I (Fermentas) in parallel reactions, 67 u T4 DNA ligase (NEB), 1.1 µl T4 DNA ligase buffer (NEB), 1 µl *Eco*RI adapter (5 pmol) and 1 µl *Hpa*II/*Msp*I adapter (50 pmol). The reaction was incubated overnight at 25 °C and diluted 1:2. For the preselective amplification (PCR1), 4 µl RL product were combined with 16 µl PCR1 reaction mix containing 1.5 ng/µl *Eco*RI- and *Hpa*II/*Msp*I preselective primers each, 200 µM dNTPs (Roth), 2 µl 10 x Dream Tag buffer (QIAGEN), 0.8 u Dream Tag polymerase (QIAGEN) and 9.84 µl H_2O. The thermocycler protocol was 72.0 °C (2 min) followed by 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min) and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler gradient. The PCR1 product was diluted 1:5. For the selective amplification (PCR2), 1 µl PCR1 product was combined with 3.4 µl PCR2 reaction mix containing 2.2 µl Multiplex PCR kit (QIAGEN) and 0.6 µl fluorescent labeled *Eco*RI primer (1 pmol/µl) and 0.6 µl *Hpa*II/*Msp*I (5 pmol/µl) selective primers each. The thermocycler protocol was 94.0 °C (2 min) followed by 10 cycles of 94.0 °C (20 s), 66.0 °C (30 s, decreasing 1 °C per cycle) and 72.0 °C (2 min) and 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min), and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler pro 384.

After an initial screening of 32 primer combinations, eight selective primer combinations were chosen for the MSAP analyses (see beneath). Separation and visualization of the MSAP fragments was done on a ABI 3130 capillary sequencer (Applied Biosystems, Foster City, USA) with Genescan 500(-250) LIZ internal size standard (Applied Biosystems). We used GeneMapper version 3.7 (Applied Biosystems) to analyze the MSAP profiles. Binning of MSAP fragments was done using a peak height threshold of 10 rfu. All samples were binned manually in the same batch, including both *Msp*I and *Hpa*II. Peak height data were exported and for each fragment a specific peak height threshold was manually determined based on the peak height distribution which allowed scoring presence (1) and absence (0) of fragments. Error rate estimation was based on 18 replicate samples (27%) that were repeated, starting from the same DNA extracts. To prevent underestimation of the error rate only polymorphic loci were considered. The overall error rate was 2.4%.

Supplementary Table S2.1 Adaptor- and primer sequences used for MSAP amplification

Primer	Sequence
Adaptors	
*Eco*RI-adapter top	5'-CTCGTAGACTGCGTACC-3'
*Eco*RI-adapter bottom	5'-AATTGGTACGCAGTCTAC-3'
*Hpa*II/*Msp*I-adapter top	5'-GATCATGAGTCCTGCT-3'
*Hpa*II/*Msp*I -adapter bottom	5'-CGAGCAGGACTCATGA-3'
Preselective primers	
*Eco*RI + A	5'-GACTGCGTACCAATTCA-3'
*Hpa*II/*Msp*I	5'-ATCATGAGTCCTGCTCGG-3'
Selective primers	
*Eco*RI + AAC-FAM[1,2]	5'-GACTGCGTACCAATTCAAC-3'
*Eco*RI + ACA-VIC[3,4]	5'-GACTGCGTACCAATTCACA-3'
*Eco*RI + AAG-NED[5,6]	5'-GACTGCGTACCAATTCAAG-3'
*Eco*RI + AGG-PET[7,8]	5'-GACTGCGTACCAATTCAGG-3'
*Hpa*II/*Msp*I + TCA[1,3]	5'-ATCATGAGTCCTGCTCGGTCA-3'
*Hpa*II/*Msp*I + TCC[2,5]	5'-ATCATGAGTCCTGCTCGGTCC-3'
*Hpa*II/*Msp*I + TCG[6]	5'-ATCATGAGTCCTGCTCGGTCG-3'
*Hpa*II/*Msp*I + TCCA[4,7]	5'-ATCATGAGTCCTGCTCGGTCCA-3'
*Hpa*II/*Msp*I + TCTA[8]	5'-ATCATGAGTCCTGCTCGGTCTA-3'

Superscript numbers indicate primer combinations used for the selective amplification.

Additional supporting information may be found in the attached CD-ROM or in the online version of the published article:

Appendix 2.1 Overview of publications using MSAP or metAFLP for population epigenetic studies

Appendix 2.3 Primary MSAP data, multistate epigenetic raw data and resulting binary data for eight scoring approaches

Appendix 2.4 R script (MSAP_score.r) for automated scoring and basic data analysis

CHAPTER 3

EPIGENETIC VARIATION REFLECTS DYNAMIC HABITAT CONDITIONS IN A RARE FLOODPLAIN HERB

Benjamin Schulz, R. Lutz Eckstein and Walter Durka
Molecular Ecology 23 (2014) 3523–3537

Abstract

Variation of DNA-methylation is thought to play an important role for rapid adjustments of plant populations to dynamic environmental conditions, thus compensating for the relatively slow response time of genetic adaptations. However, genetic and epigenetic variation of wild plant populations has not yet been directly compared in fast changing environments. Here we surveyed populations of *Viola elatior* from two adjacent habitat types along a successional gradient characterized by strong differences in light availability. Using AFLP and MSAP analyses we found relatively low levels of genetic (H'_{gen} = 0.19) and epigenetic (H'_{epi} = 0.23) diversity and high genetic (ϕ_{ST} = 0.72) and epigenetic (ϕ_{ST} = 0.51) population differentiation. Diversity and, respectively, differentiation were significantly correlated, suggesting that epigenetic variation partly depends on the same driving forces as genetic variation. Correlation based genome scans detected comparable levels of genetic (17.0%) and epigenetic (14.2%) outlier markers associated with site specific light availability. However, as revealed by separate differentiation based genome scans for AFLP, only few genetic markers seemed to be actually under positive selection (0–4.5%). Moreover, PCoA and Mantel tests showed that overall epigenetic variation was more closely related to habitat conditions, indicating that environmentally-induced methylation changes may lead to convergence of populations experiencing similar habitat conditions and may thus play a major role for the transient and/or heritable adjustment to changing environments. Additionally, using a new MSAP scoring approach, we found that mainly the unmethylated (ϕ_{ST} = 0.60) and CG-methylated states (ϕ_{ST} = 0.46) of epiloci contributed to population differentiation and putative habitat related adaptation, whereas CHG-hemimethylated states (ϕ_{ST} = 0.21) only played a marginal role.

Introduction

To cope with spatial heterogeneity and temporal dynamics of their habitats, plants as sessile organisms constantly need to respond to varying environmental conditions. While slow adaptation of morphology, physiology or development can be attained by genetic changes through mutation, drift and selection, rapid responses to biotic or abiotic alterations necessitate alternative regulations (Boyko and Kovalchuk 2011). As an alternative to 'hard-' or 'Mendelian inheritance', 'soft inheritance' through epigenetic mechanisms has been proposed to fill this gap between random genetic and environmental variation (Richards 2006). Epigenetic silencing or activation of protein coding genes, which is sensitive to environmental variation, may convert environmental heterogeneity into phenotypic differences (Flores *et al.* 2013). These epigenetically induced phenotypic changes can then be inherited through meiosis over several generations (reviewed in Jablonka and Raz 2009), giving rise to so-called epialleles. Furthermore, due to its semi-stable nature, inherited epigenetic variation is suited to plastically react to environmental fluctuations that last for only few generations. Especially in long-lived perennials and species that mainly reproduce vegetatively or through self-pollination, such fast adjustment through epigenetic processes seems essential to ensure continued survival of populations under episodic stress (Castonguay and Angers 2012; Bräutigam et al 2013).

An array of epigenetic mechanisms has been identified, including chemical modification of DNA and histones, position effects, and interference by small non-coding RNAs (Berger et. al 2007). Without changing the underlying DNA sequence these mechanisms can rapidly modulate existing genetic information through the control of gene expression or the reorganization of chromatin structure (Sahu *et al.* 2013). The most extensively studied and best understood epigenetic mechanisms is the potentially reversible methylation of DNA, which in higher eukaryotes mainly involves the methylation of cytosine residues (Feng *et al.* 2010). In plants, cytosine methylations are found throughout the genome in three different sequence contexts, symmetric CG- and CHG-sites (H = A, C, T) and asymmetric CHH-sites. Methylations in all three contexts are predominantly found in transposable elements and repetitive sequences where they generally are related to transcriptional repression (Saze *et al.* 2012).

Various studies have shown that the amount and pattern of DNA-methylation in plants is sensitive to biotic and abiotic stressors such as pathogens (Wada *et al.* 2004), herbivores (Herrera and Bazaga 2013), drought (Labra *et al.* 2002), extreme temperatures (Boyko *et al.* 2010) or nutrient availability (Boyko *et al.* 2010; Kou *et al.* 2011). Depending on the species, genotype and environmental cue, stress treatments induced hypo- or hypermethylation or balanced shifts in global methylation patterns that in some cases were stably transmitted to the next generation (e.g. Verhoeven *et al.* 2010; Kou *et al.* 2011). Recent studies even directly correlated stress induced inherited changes of DNA methylation with adaptive plant responses to salt stress (Boyko *et al.* 2010) or high-temperature (Correia *et al.* 2013). Taken together, the inheritance and the adaptive value of stress-induced

alterations in DNA methylations strongly support the idea of an epigenetically regulated system of "soft inheritance" (Zhang *et al.* 2013).

However, there is still a dearth of studies that explore the extent of epigenetic variation and its adaptive potential in natural populations (Angers *et al.* 2010; Latzel *et al.* 2013). Moreover, whereas previous surveys mostly focused on species with very little or no genetic variation (e.g. Richards *et al.* 2012; Yu *et al.* 2013), improving the ability to attribute epigenetic variation to specific environmental conditions (Ledon-Rettig 2013), only few studies compared the impact of genetic and epigenetic variation in genetically diverse species (Herrera and Bazaga 2010; Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012; Wu *et al.* 2013). However, most plant populations are genetically diverse, implying a complex relationship between genetic, epigenetic and environmental variation. Indeed, the available studies found a certain correlation between genetic and epigenetic variation, but also revealed that epigenetic differentiation was more closely aligned with the environment than genetic differentiation (Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012), suggesting that heritable epigenetic changes might play an important role for local adaptation. However, all of the investigated habitats represented temporally stable conditions that persisted for long periods of time (e.g. differences in salinity, heavy metal content or altitude), potentially resulting in extensive genetic adaptation. Hence, this complicates the study of epigenetically based 'soft inheritance' as a mechanism for rapid and independent environmental adjustment. Consequently, there is a need for studies that focus on more dynamic, faster changing systems that are less likely to allow for genetic adaptation.

One such fast changing environmental parameter is the decreasing light availability along successional gradients. Quality and quantity of light may both have effects on DNA methylation (Greco *et al.* 2013; Tatra *et al.* 2000). Moreover, as light is a strong environmental cue, influencing an array of other biotic and abiotic parameters (e.g. competition, water availability or temperature) a high selection pressure is expected under changing light conditions, making successional gradients an ideal study system to compare genetic and epigenetic variation in response to fast changing environments.

To assess methylation variation of large numbers of individuals, population studies mostly use methylation sensitive amplification polymorphisms (MSAP) that essentially can differentiate unmethylated, CG-methylated and CHG-hemimethylated states of particular epiloci (Salmon *et al.* 2008). However, for statistical analyses most previous studies combined the two discernible methylation types into one common state, ignoring regulatory and potential functional differences. Moreover, unmethylated states are generally not considered, thus missing further resolution and information about the putatively active state of the epigenome (Schulz *et al.* 2013). Consequently, there is a need for studies that consider all three MSAP fragment types and thus potentially enable more detailed insights into epigenetic processes.

Here, we investigated genetic and epigenetic variation of *Viola elatior* in two adjacent habitat types along a successional gradient characterized by strong differences in light availability. Using amplified fragment length polymorphisms (AFLP) and MSAP and applying genome scan approaches to detect putatively adaptive genetic and epigenetic outlier loci we asked the following questions: (1) Do genetic and epigenetic variation

contribute differentially to habitat related differentiation? (2) Are epigenetic differences between habitat types consistent among regions and thus potentially decoupled from genetic variation? (3) How do different methylation types (i.e. unmethylated, CG-methylated or CHG-hemimethylated fragments) contribute to epigenetic population structuring?

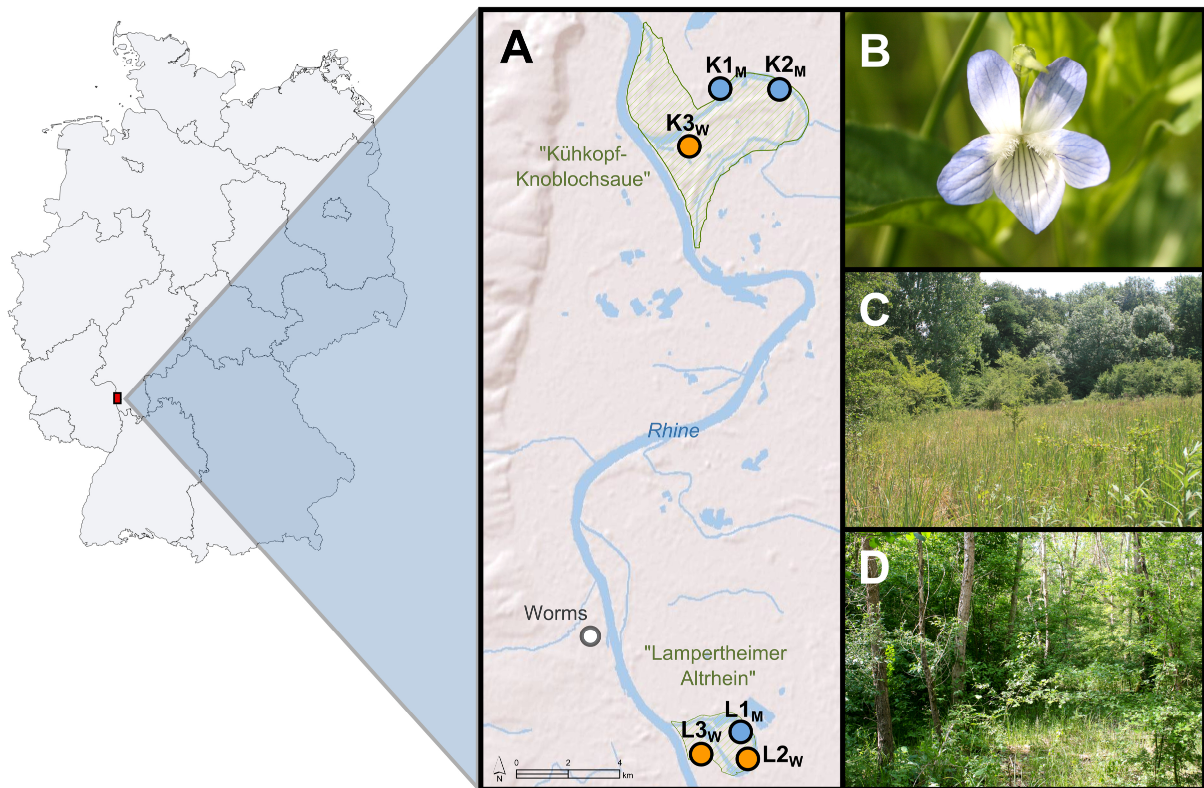


Figure 3.1 Location and study sites. (A) Section of the Upper Rhine Valley south of Frankfurt/Main, Germany, with the six sampled populations of *Viola elatior* in the nature reserves Kühkopf-Knoblochsaue and Lampertheimer Altrhein. Symbol colors indicate the habitat type (blue = meadow, brown = woodland). (B) *Viola elatior* flower. (C) and (D) Surveyed habitat types with strong differences in light availability: (C) floodplain meadow, (D) alluvial woodland fringe.

Material and Methods

Study species

Viola elatior (Violaceae) is a perennial iteroparous hemicryptophyte. In Central Europe the species is restricted to different floodplain habitats ranging from floodplain meadows to alluvial woodland fringes, characterized by changing environmental conditions due to disturbance and succession (Eckstein and Otte 2005; Eckstein *et al.* 2006). These environmental changes may proceed rather fast, encompassing a time span of only few plant generations. Although *V. elatior* survives under a wide range of light conditions, with in-

creasing succession to closed alluvial forests, population sizes decline and the species finally disappears (Eckstein *et al.* 2006). However, due to its persistent seed bank, after major disturbance events populations may recover even after years of absence from the above-ground vegetation (Eckstein *et al.* 2006). In Germany, the species is rare and endangered with populations of varying sizes from tens to several hundred individuals. The species is octoploid (2n = 40) and exhibits a mixed mating system with potentially cross-pollinated chasmogamous (CH) and obligatory self-pollinated cleistogamous (CL) flowers (Eckstein *et al.* 2006). Nonetheless, seed production through CL flowers is dominating. In common garden experiments only around 4% of capsules were produced by CH flowers (Eckstein and Otte 2005; Schulz *et al.* unpublished), suggesting a high selfing rate also in natural populations. Additionally, vegetative reproduction by disintegration of old pleiocorms or formation of root buds occurs (Eckstein *et al.* 2006).

Sampling Design

The study was conducted with six stands (hereafter called populations) from the Upper Rhine Valley located in the nature reserves "Kühkopf-Knoblochsaue" (K) and "Lampertheimer Altrhein" (L) near Frankfurt, Germany (Figure 3.1). First, all known populations of *V. elatior* in the two regions were visually classified according to their light environment. Then, populations for the two extremes of the species' environmental range, i.e. sunny floodplain meadows (M) and shady alluvial woodland fringes (W), were selected. Due to limited numbers of appropriate populations in the different habitat types with more than 20 individuals, we chose two M- and one W-site in region K and one M- and two W-sites in region L (Figure 3.1, Table 3.1). Distance between regions is 25 km and within regions, distances between populations range between 0.5 and 4 km.

After mapping *V. elatior* at each site on a 1 m grid, 20 to 25 populated grid cells were randomly selected and young undamaged leaves from one reproductive plant per cell were collected. Samples were immediately cooled to below 10 °C, stored at -25 °C and freeze-dried for 48 h.

Light availability

Mean transmitted photosynthetic active radiation (PAR) per site was assessed with hemispherical photography. At each site, 20 digital hemispherical photographs were taken in randomly selected populated grid cells (i.e. independent from plant sampling) with a Nikon CoolPix 4500 camera and a Fish-eye converter lens FC-E8 (Nikon, Tokyo, Japan) mounted on a tripod 70 cm above ground. Aperture and exposure time were fitted following Zhang *et al.* (2005). Images were converted to black-and-white bitmaps using automatic threshold detection algorithms with SIDELOOK 1.1.01 (Nobis and Hunziker, 2005) and then analyzed using GAP LIGHT ANALYZER (GLA) 2.0 (Frazer *et al.* 1999). Light availability for each site was subsequently quantified as mean daily percentage of transmitted total PAR for the period from 1 May to 30 September averaged across the 20 images (Table 3.1).

Table 3.1 Overview of sampled populations of *Viola elatior*

Population ID	Region	Latitude	Longitude	Samples	Grid cells (m^2)	Habitat Type	Mean Transmitted PAR ± SD (%)
$K1_M$	Kühkopf-Knoblochsaue	49°50'01"N	8°25'32"E	20	~1400	meadow	87.8 ± 6.5
$K2_M$	Kühkopf-Knoblochsaue	49°49'49"N	8°28'03"E	20	159	meadow	73.8 ± 8.3
$K3_W$	Kühkopf-Knoblochsaue	49°48'50"N	8°24'57"E	25	158	wood-land	23.6 ± 14.2
$L1_M$	Lamperthei-mer Altrhein	49°36'08"N	8°26'50"E	22	170	meadow	80.5 ± 14.7
$L2_W$	Lamperthei-mer Altrhein	49°35'49"N	8°26'50"E	24	31	wood-land	12.5 ± 3.3
$L3_W$	Lamperthei-mer Altrhein	49°35'44"N	8°25'55"E	21	144	wood-land	16.5 ± 9.4

AFLP genotyping and MSAP epi-genotyping

We investigated a total of 132 individuals for genetic and epigenetic variation with AFLP and MSAP. For each individual, the same DNA sample was used for both approaches. Total genomic DNA was extracted from dried leaf tissue using the DNeasy 96 Plant extraction kit (QIAGEN). AFLP and MSAP methodology followed the protocol as described in Appendix 3.1, using eight selective primer combinations, respectively (Supplementary Table S3.1).

For MSAP analyses comparison of the banding patterns of *Eco*RI/*Hpa*II and *Eco*RI/*Msp*I reactions results in four conditions of a particular fragment: I = fragments present in both profiles (1/1), indicating an unmethylated state; II = fragments present only in *Eco*RI/*Msp*I profiles (0/1), indicating hemi- or fully-methylated CG-sites; III = fragments present only in *Eco*RI/*Hpa*II profiles (1/0), indicating hemimethylated CHG-sites; IV = absence of fragments in both profiles (0/0), representing an uninformative state caused either by different types of methylation or due to restriction site polymorphism. To separate unmethylated and methylated fragments and to test for the particular impact of the methylated conditions II and III, we used the "Mixed-Scoring 2" approach (Schulz *et al.* 2013). This approach generates the final epigenetic data matrix by transforming the three discernible methylation states at each multistate epilocus into separate binary subepiloci, thereby keeping all available information and allowing for fragment specific analyses. Thus, for a particular epilocus up to three subepiloci are generated, coding the non-methylated (n-subepiloci), the CG-methylated (m-subepiloci) and the

CHG-hemimethylated state (h-subepiloci). For a detailed description of the steps of epigenotyping, see Schulz *et al.* (2013). Overall error rate for AFLP and MSAP data was 3.3 and 2.1%, respectively, based on 33 replicate samples.

Data analysis

AFLP and MSAP binary data sets were analyzed following the same framework using a band or marker based strategy, i.e without calculating allele frequencies (Bonin *et al.* 2007). Genetic and epigenetic diversity within populations were quantified using the R script MSAP_calc (Schulz *et al.* 2013) as (i) percentage of total and private bands and (ii) percentage of polymorphic loci (PLP_{gen}) and subepiloci (PLP_{epi}) and (iii) mean Shannon's information index H'_{gen} and H'_{epi}.

Patterns of individual relationships were depicted by principal coordinates analyses (PCoA) with GENALEX 6.41 (Peakall and Smouse 2006) based on a matrix of Nei and Li distances calculated with DistAFLP (Mougel *et al.* 2002). Distance matrices were square root transformed to meet the assumptions of PCoA analyses (Legendre and Legendre 1998). Genetic and epigenetic variation among groups of populations (ϕ_{CT}), among populations within groups (ϕ_{SC}) and within populations (ϕ_{ST}) was partitioned with hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010). Additionally, pairwise ϕ_{ST} values were estimated among populations. Significance levels were determined after 9999 permutations. To evaluate the relationships between geography and habitat type and genetic and epigenetic differentiation we conducted pairwise and partial Mantel tests using the 'vegan' library in R (Oksanen *et al.* 2007) with 999 permutations. We used Euclidian geographic distances and for habitat types we constructed a matrix coding pairs of meadow-woodland populations by '1' and pairs of equal habitats by '0'.

To detect signatures of light related adaptation in the genetic and epigenetic data sets we used the Spatial Analysis Method (SAM) (Joost *et al.* 2007), which uses multiple univariate logistic regression. In our case, the site specific mean values of percentage of transmitted PAR (Table 3.1) were assigned to genetic and epigenetic marker data of all individuals from one population. SAM uses the individual as reference unit, functions independently of any presumed population structure and is largely assumption free (Joost *et al.* 2007; Paun *et al.* 2010). Only if the two statistical tests implemented in SAM (likelihood ratio G, and Wald test) reject the null hypothesis, a model is considered significant (Joost *et al.* 2007). Bonferroni correction of the significance level for multiple comparisons was applied corresponding to a 95% and 99% confidence interval (CI). To depict effects of putatively adaptive markers on population structure, AFLP and MSAP data sets were partitioned into subsets of neutral and outlier markers (95% CI) and were then visualized with PCoA as described above. For MSAP data corresponding n-, m- or h-subepiloci (i.e. resulting from one epilocus) were jointly excluded from the neutral data set when at least one of them was classified as an outlier.

To obtain further information on the degree of positive selection at the genetic level, we used two differentiation based genome scan approaches on the AFLP data: DFDIST/FDIST (Beaumont and Balding 2004) as included in the workbench MCHEZA (Antao and Beaumont 2011) and BAYESCAN 2.1 (Foll and Gaggiotti 2008). DFDIST/FDIST analyses in MCHEZA were conducted with 50 000 simulations, using the combined "Neutral mean F_{ST}" and "Force mean F_{ST}" algorithms and two detection levels at 95% and 99% confidence interval (CI). Results were corrected for multiple comparisons by setting false discovery rate to 0.1. BAYESCAN analyses were run with a burn-in of 50 000 iterations, a sample size of 10 000 and a thinning interval of 50, resulting in a total of 550 000 iterations. An additional burn-in was carried out by 20 short pilot runs of 5000 iterations. Only loci exceeding a "strong" detection level (Bayesian factor BF > 10) were considered as putative outliers.

Results

Genetic and epigenetic diversity

AFLP analysis of the 132 individuals resulted in 428 scorable loci of which 112 (26%) were polymorphic. Across all individuals, 106 unique AFLP phenotypes were found, i.e. 20% of individuals had shared AFLP phenotypes, ranging from 10% in $K1_M$ to 42% in $L2_W$. Within populations between one and four AFLP phenotypes were shared by more than one individual (Table 3.2). As the mean distance between individuals with shared AFLP phenotypes was 14.1 m and only 6 of 49 pairs were separated by less than 3 m, we assumed that the majority of pairs do not represent vegetative clones but closely related, inbred individuals. Thus, all individuals were retained in the subsequent analyses. Assessment of genetic diversity across populations (Table 3.2) revealed mean values of 67.7% for percentage of bands present, 2.7% for private bands, 37.5% for PLP_{gen} and 0.19 for H'_{gen}.

MSAP analysis resulted in 555 scorable epiloci of which 275 (49.5%) were polymorphic. These polymorphic epiloci yielded 444 polymorphic subepiloci, consisting of 207 n-, 157 m- and 80 h-subepiloci. 42.9% of all epiloci resulted in only one subepilocus type (19.3, 11.3 and 12.4% for n-, m- and h-subepiloci, respectively), 40.4% in n- and m-subepiloci, 11.2% in n- and h-subepiloci, 1.1% in m- and h-subepiloci and 4.4% yielded all three types of subepiloci. All 132 individuals had unique MSAP phenotypes. Values of diversity for the different subepiloci across populations are given in Table 3.2. Population level genetic and epigenetic diversity was significantly correlated (Spearman rank-correlation: rho = 0.83, $p = 0.042$ for *PLP*; rho = 0.89, $p = 0.019$ for H') and epigenetic diversity was or tended to be higher than genetic diversity (paired t-test: $t(5) = -3.20$, $p = 0.024$ for *PLP*; $t(5) = -2.43$, $p = 0.060$ for H').

Table 3.2 Measures of genetic and epigenetic diversity within six populations of *Viola elatior*

	AFLP	MSAP all	MSAP n-subepiloci	MSAP m-subepiloci	MSAP h-subepiloci
Polymorphic loci	112	444	207	157	80
Phenotypes per population*					
$K1_M$ (n = 20)	18 (2)	20	20	20	20
$K2_M$ (n = 20)	14 (4)	20	20	20	20
$K3_W$ (n = 25)	22 (1)	25	25	25	25
$L1_M$ (n = 22)	19 (2)	22	22	22	22
$L2_W$ (n = 24)	14 (4)	24	24	24	24
$L3_W$ (n = 21)	18 (1)	21	21	21	21
Bands per population (%)					
$K1_M$	62.5	61.7	61.4	72.6	41.3
$K2_M$	75.9	70.9	74.4	72.6	58.8
$K3_W$	67.0	63.7	61.8	68.8	58.8
$L1_M$	67.9	56.1	54.6	63.7	45.0
$L2_W$	61.6	52.3	51.7	61.1	36.3
$L3_W$	71.4	62.2	63.8	67.5	47.5
mean	*67.7*	*61.1*	*61.3*	*67.7*	*47.9*
Private bands per population (%)					
$K1_M$	2.7	1.6	1.4	1.9	1.3
$K2_M$	3.6	5.4	5.8	4.5	6.3
$K3_W$	0.0	2.9	2.9	1.3	6.3
$L1_M$	4.5	2.5	2.4	1.9	3.8
$L2_W$	1.8	0.9	1.0	0.6	1.3
$L3_W$	3.6	2.0	2.4	1.3	2.5
mean	*2.7*	*2.6*	*2.7*	*1.9*	*3.5*
PLP					
$K1_M$	37.5	45.1	43.0	51.6	37.5
$K2_M$	40.2	61.0	61.8	61.8	57.5
$K3_W$	39.3	49.8	43.0	55.4	56.3
$L1_M$	32.1	36.7	31.9	40.1	42.5
$L2_W$	17.9	27.7	22.7	30.6	35.0
$L3_W$	46.4	46.9	43.0	52.2	46.3
mean	*35.6*	*44.5*	*40.9*	*48.6*	*45.8*
H'					
$K1_M$	0.22	0.25	0.24	0.29	0.19
$K2_M$	0.26	0.34	0.35	0.34	0.29
$K3_W$	0.15	0.23	0.22	0.24	0.25
$L1_M$	0.19	0.19	0.17	0.20	0.24
$L2_W$	0.08	0.13	0.11	0.14	0.18
$L3_W$	0.25	0.24	0.22	0.28	0.22
mean	*0.19*	*0.23*	*0.22*	*0.25*	*0.23*
overall	*0.59*	*0.44*	*0.50*	*0.43*	*0.31*

* numbers in brackets indicate AFLP phenotypes that were shared by more than one individual.
PLP – percentage of polymorphic loci (PLP_{gen} for AFLP data) and epiloci (PLP_{epi} for MSAP data);
H' – Shannon's information index (H'_{gen} for AFLP data and H'_{epi} for MSAP data).

Genetic and epigenetic structure

Principal coordinates analysis of genetic distances separated K- and L-sites, both being further subdivided into two clusters (Figure 3.2). Whereas regions were mainly separated along the second coordinate, no grouping for habitat-type could be observed.

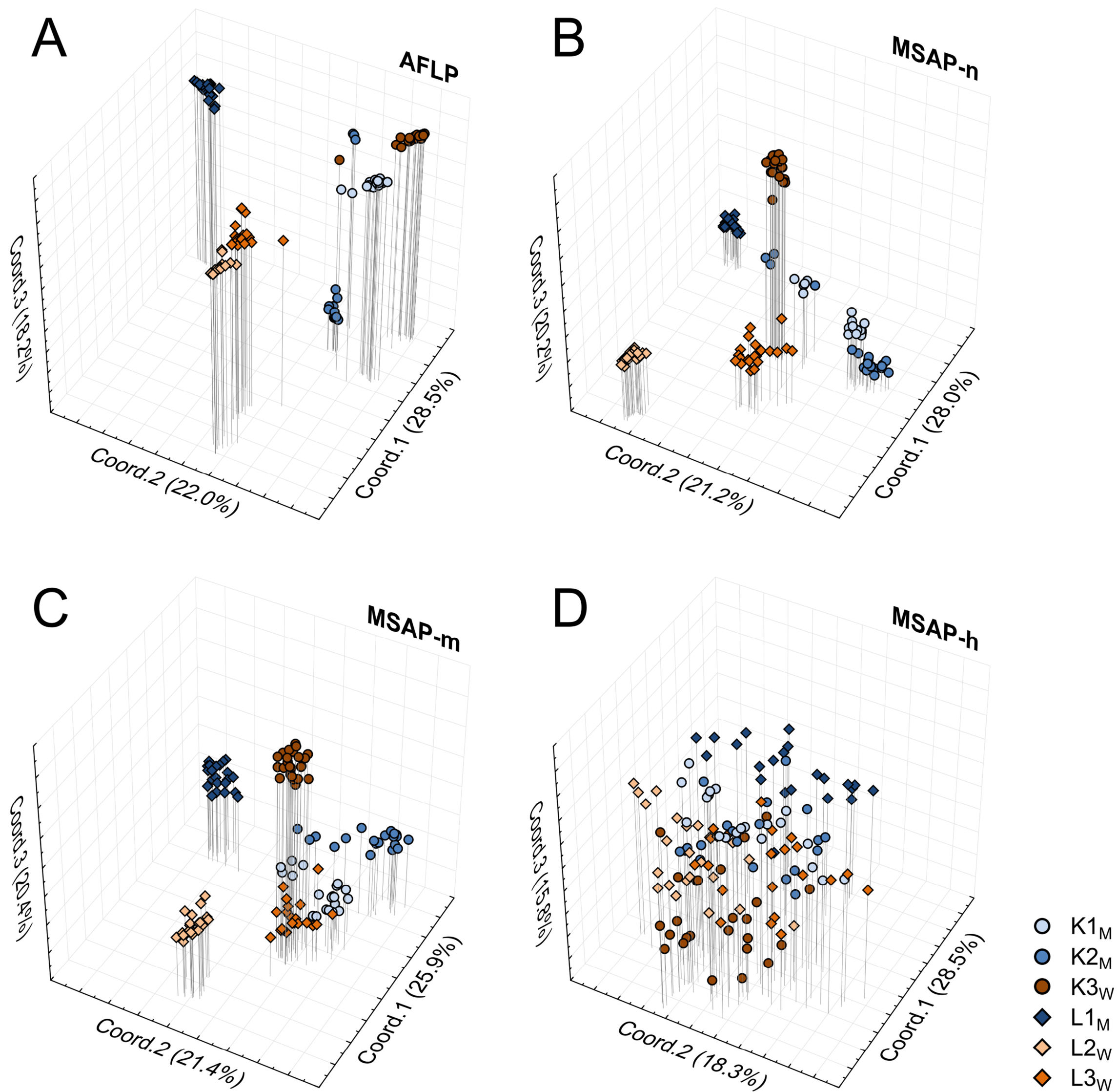


Figure 3.2 Principal Coordinates Analyses (PCoA) of genetic (A) and epigenetic distances (B–D) of six populations of Viola elatior. Epigenetic data was partitioned into the three discriminable methylation types, i.e. (B) unmethylated (n), (C) CG-methylated (m) and (D) CHG-hemimethylated (h) subepiloci. Regions are indicated by symbol type (circle = Kühkopf-Knoblochsaue, diamond = Lampertheimer Alt-rhein). Habitats are displayed by color (blue = meadow, brown = woodland).

PCoA of epigenetic distances revealed varying population differentiation patterns among subepiloci (Figure 3.2). For n-subepiloci, in contrast to genetic data, $K3_W$ was separated from both K-meadow sites that formed two mixed clusters. All L-sites were clearly separated. For m-subepiloci individuals were less clumped but clustering overall resembled that of n-subepiloci. For both, n- and m-subepiloci habitat types tended to be separated along the first axis. In contrast, for h-subepiloci individuals were widely scattered and without clear population differentiation. When n-, m- and h-subepiloci were combined in one analysis, similar patterns of population differentiation were found as for n- and m-subepiloci (data not shown).

Analysis of molecular variance of genetic data revealed a global ϕ_{ST} of 0.72 and pairwise ϕ_{ST} between 0.52 and 0.84 (Supplementary Table S3.2), indicating very strong genetic population differentiation. Hierarchical AMOVA (Table 3.3) showed that 11.6% of genetic variance resided between regions, but only 2.1% between habitats. Most variation was partitioned among populations within geographic or habitat groups (62.1% and 70.5%, respectively).

For the combined epigenetic data set (Table 3.3) overall differentiation was lower than for genetic data but still high (global $\phi_{ST} = 0.51$), with pairwise population ϕ_{ST} ranging from 0.35 to 0.67 (Supplementary Table S3.2). Hierarchical AMOVA revealed that only 1.2% of epigenetic variance resided between regions, but 6.1% between habitats. Variation among populations within groups was similar for regions (49.7%) and habitats (45.9%). Considering n-, m- and h-subepiloci separately revealed global ϕ_{ST} values of 0.60, 0.46, and 0.21, respectively. In all MSAP data sets, stronger differentiation was observed among habitats than among regions, contrary to the AFLP data.

Mantel tests revealed a strong correlation between genetic and epigenetic differentiation ($r = 0.76$, $p < 0.005$). However, no simple isolation-by-distance or isolation-by-habitat pattern was found as no significant relationship existed between genetic or epigenetic differentiation and geographical distance ($p > 0.18$) or habitat similarity ($p > 0.1$) (Supplementary Table S3.3). However, when controlling for geographic distance, habitat similarity was significantly correlated with epigenetic ($p = 0.04$), but not with genetic differentiation ($p = 0.30$).

Table 3.3 Hierarchical AMOVA for AFLP and MSAP data performed by grouping populations according to regions of origin or habitat types

Loci/Groups	Variance components	V	% total	*p*	ϕstatistics
AFLP loci (112)					
All populations (n = 6)	Among all populations	12.09	72.36	< 0.0001	$\phi_{ST} = 0.72$
Regions (n = 2)	Among regions	2.03	11.56	0.20	$\phi_{CT} = 0.12$
	Among populations within regions	10.87	62.08	< 0.0001	$\phi_{SC} = 0.70$
	Within populations	4.62	26.36	< 0.0001	$\phi_{ST} = 0.74$
Habitats (n = 2)	Among habitats	0.36	2.12	0.39	$\phi_{CT} = 0.02$
	Among populations within habitats	11.88	70.48	< 0.0001	$\phi_{SC} = 0.72$
	Within populations	4.62	27.40	< 0.0001	$\phi_{ST} = 0.73$
MSAP all subepiloci (444)					
All populations (n = 6)	Among all populations	22.87	50.73	< 0.0001	$\phi_{ST} = 0.51$
Regions (n = 2)	Among regions	0.56	1.24	0.50	$\phi_{CT} = 0.01$
	Among populations within regions	22.53	49.73	< 0.0001	$\phi_{SC} = 0.50$
	Within populations	22.21	49.02	< 0.0001	$\phi_{ST} = 0.51$
Habitats (n = 2)	Among habitats	2.80	6.07	0.10	$\phi_{CT} = 0.06$
	Among populations within habitats	21.19	45.86	< 0.0001	$\phi_{SC} = 0.49$
	Within populations	22.21	48.07	< 0.0001	$\phi_{ST} = 0.52$
MSAP n-subepiloci (207)					
All populations (n = 6)	Among all populations	14.77	59.87	< 0.0001	$\phi_{ST} = 0.60$
Regions (n = 2)	Among regions	0.43	1.75	0.39	$\phi_{CT} = 0.02$
	Among populations within regions	14.51	58.41	< 0.0001	$\phi_{SC} = 0.59$
	Within populations	9.90	39.85	< 0.0001	$\phi_{ST} = 0.60$
Habitats (n = 2)	Among habitats	2.01	7.90	0.10	$\phi_{CT} = 0.08$
	Among populations within habitats	13.56	53.24	< 0.0001	$\phi_{SC} = 0.58$
	Within populations	9.90	38.85	< 0.0001	$\phi_{ST} = 0.61$
MSAP m-subepiloci (157)					
All populations (n = 6)	Among all populations	7.02	45.67	< 0.0001	$\phi_{ST} = 0.46$
Regions (n = 2)	Among regions	0.27	1.78	0.49	$\phi_{CT} = 0.02$
	Among populations within regions	6.85	44.28	< 0.0001	$\phi_{SC} = 0.45$
	Within populations	8.35	53.94	< 0.0001	$\phi_{ST} = 0.46$
Habitats (n = 2)	Among habitats	0.63	4.02	0.21	$\phi_{CT} = 0.04$
	Among populations within habitats	6.64	42.53	< 0.0001	$\phi_{SC} = 0.44$
	Within populations	8.35	53.45	< 0.0001	$\phi_{ST} = 0.47$
MSAP h-subepiloci (80)					
All populations (n = 6)	Among all populations	1.08	21.45	< 0.0001	$\phi_{ST} = 0.21$
Regions (n = 2)	Among regions	-0.15	-2.92	1.00	$\phi_{CT} = -0.03$
	Among populations within regions	1.17	23.45	< 0.0001	$\phi_{SC} = 0.23$
	Within populations	3.97	79.47	< 0.0001	$\phi_{ST} = 0.21$
Habitats (n = 2)	Among habitats	0.16	3.16	0.10	$\phi_{CT} = 0.03$
	Among populations within habitats	0.99	19.28	< 0.0001	$\phi_{SC} = 0.20$
	Within populations	3.97	77.55	< 0.0001	$\phi_{ST} = 0.22$

Outlier detection

For the AFLP data set SAM identified 19 (17.0%) and 17 (15.2%) outliers that were associated with light environment at 95% and 99% CI, respectively (Table 3.4, Supplementary Table S3.4). In contrast, with both differentiation based genome scans the number of genetic outliers was strongly reduced. Whereas DFDIST/FDIST detected 5 (4.5%) and 1 (0.9%) outliers at 95% and 99% CI (Supplementary Table S3.4), respectively, no outliers at all were identified by BAYESCAN. Two (1.8%) outliers were similarly detected by SAM and DFDIST/FDIST.

For the complete set of the 444 MSAP subepiloci SAM identified 50 (11.3%) and 30 (6.8%) outliers at 95% and 99% CI, respectively (Table 3.4, Supplementary Table S3.5). Related to the original 275 epiloci, 39 (14.2%) and 24 (8.7%) of all epiloci resulted in one or two outlier subepiloci at 95% and 99% CI, respectively (Supplementary Table S3.5). At 95% CI the three methylation types contributed differently to outliers as 15.0%, 10.2% and 3.8% of the n-, m- and h-subepiloci, respectively, revealed signatures of adaptation ($\chi^2 = 6.21$, $p = 0.045$).

Table 3.4 Numbers of AFLP and MSAP outlier (epi)loci for six populations of *Viola elatior* as assessed by SAM analyses for correlation with light environment

	AFLP	MSAP	MSAP-n	MSAP-m	MSAP-h
SAM 95% CI	19	50	31	16	3
SAM 99% CI	17	30	20	7	3

Comparing PCoA patterns of neutral and outlier markers separately revealed strong differences between marker groups (Figure 3.3). For neutral markers, overall population structuring was very similar between AFLP and MSAP data, dividing regions along a diagonal between the first and second axes. In contrast, for outlier markers, only epigenetic data clearly separated habitat types. Moreover, whereas for AFLP data all meadow populations formed individual clusters along the second axes and regions still tended to be separated, for MSAP data, meadow populations formed one larger cluster that only showed some variation along the third axes.

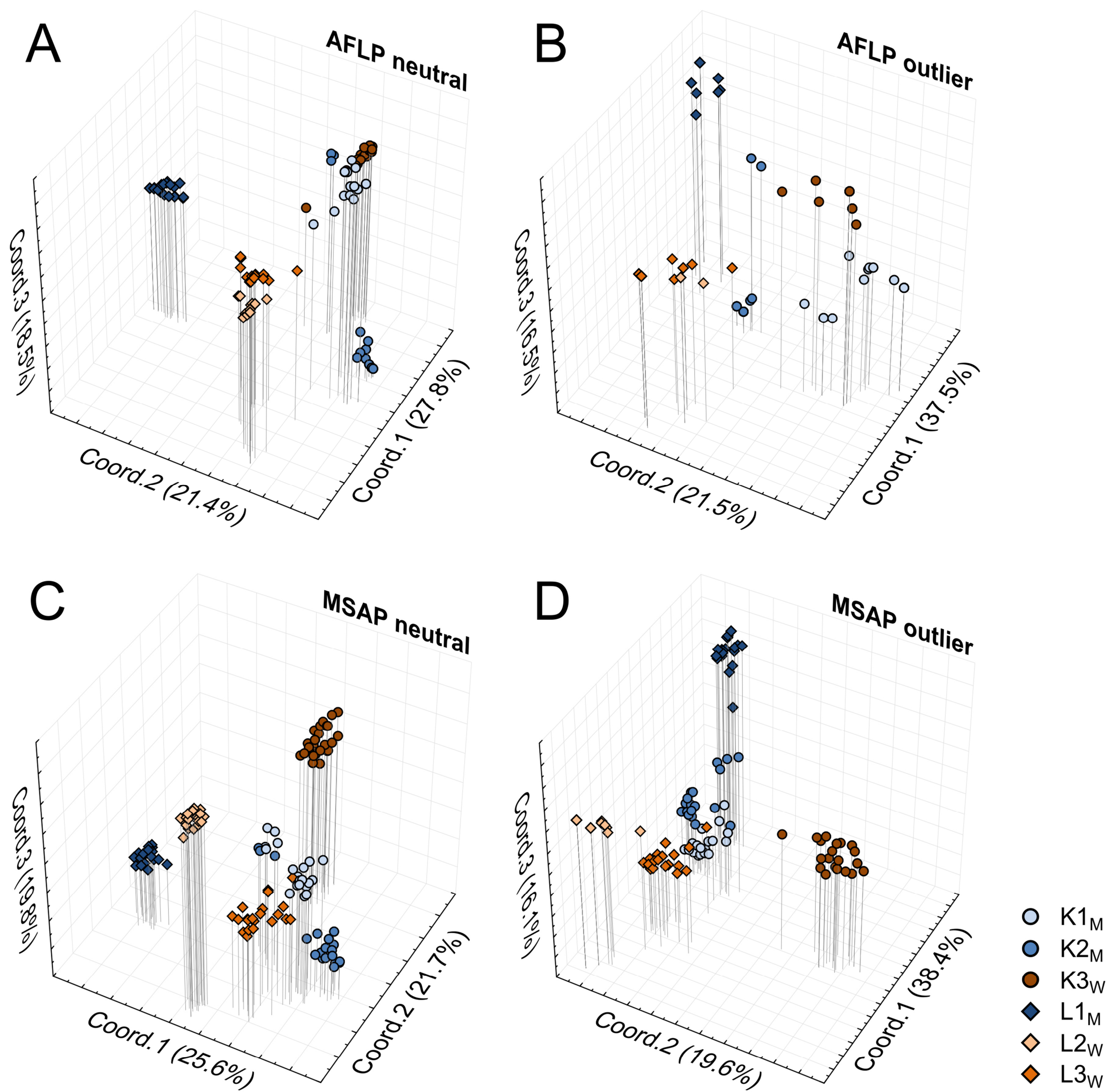


Figure 3.3 Principal Coordinates Analyses (PCoA) of genetic (A–B) and epigenetic (C–D) distances of neutral and outlier (subepi)loci as assessed by SAM for correlation with light availability. For legend, see Figure 3.2.

Behavior of subepilocus outliers

For most m-subepiloci identified as outliers by SAM also the corresponding n-subepiloci were identified as outliers (11/16 for 95% CI; 6/7 for 99% CI), i.e. both, the n- and m-subepilocus derived from the same MSAP epilocus were putatively adaptive (Supplementary Table S3.5). Interestingly, all putatively adaptive n-/m-subepilocus pairs identified by SAM had reciprocal epigenotypes at low and high light conditions. For example, epilocus ACA_TCCA_188 was always non-methylated in woodland sites, whereas in meadow sites it was mainly CG-methylated (Figure 3.4). Strikingly, almost all identified

light-related n-/m-pairs (9/11 for 95% CI; 5/6 for 99% CI) were similarly associated with woodland and meadow sites, respectively. Moreover, in the complete set of SAM outliers (99% CI) significantly more n-subepiloci were associated with low light availability than expected (17 of 20 for 99% CI; χ^2-test with Yates continuity correction: $\chi^2 = 4.10$, $p = 0.043$).

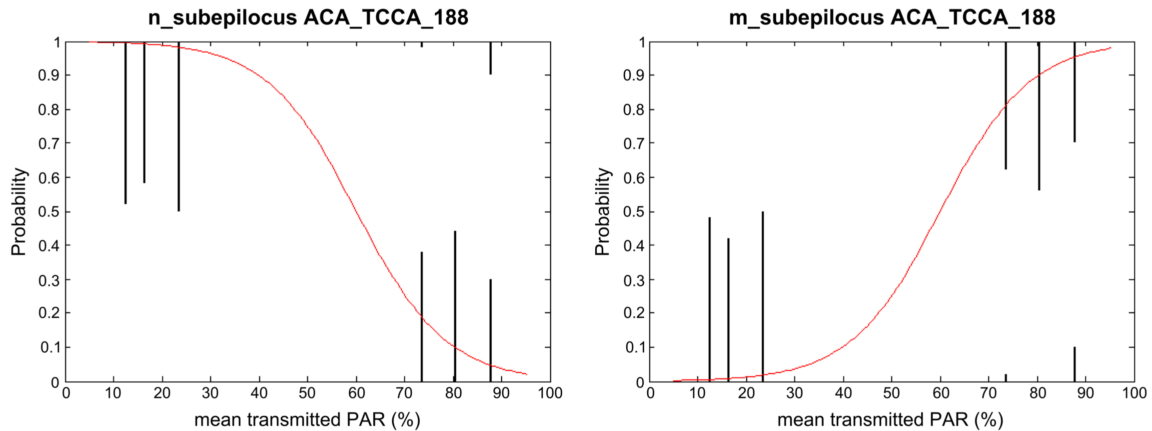


Figure 3.4 Example of two corresponding outlier subepiloci, illustrating a putatively adaptive epigenetic switch mechanism. The probability of band presence/absence of both, n- and m-subepilocus of the MSAP marker ACA_TCCA_188 are significantly associated with mean percentage of transmitted photosynthetic active radiation (PAR). The graphs show the logistic sigmoid function of marker presence versus the mean light availability at the sampling sites. At low light (woodland) the marker ACA_TCCA_188 is found to be completely unmethylated, while at high light (meadow) nearly all individuals show a CG-methylation. Bars represent the number of present (top) and absent (bottom) bands for each of the six surveyed populations.

Discussion

Epigenetic modifications, such as DNA methylation, may play a pivotal role in the response of plant populations to environmental change (Bossdorf *et al.* 2008; Sahu *et al.* 2013). Additionally, due to their potential inheritance, it was proposed that epigenetic marks can be under selection and even might impact evolution (Richards 2006; Angers *et al.* 2010, Flatscher *et al.* 2012). However, directly testing epigenetic adaptation in natural populations is challenging, as the adaptive value of epigenetic variation is hard to prove under complex and dynamic environmental conditions and in the presence of genetic variation. Here we used an indirect approach to assess the putative contribution of genetic and epigenetic variation in populations of *V. elatior* to environmental adaptation along a light gradient by comparing genetic and epigenetic population structure and applying a correlation based outlier approach.

Genetic and epigenetic diversity

AFLP analyses revealed low levels of genetic diversity both overall and within populations of *V. elatior*. This is consistent with findings for other mixed mating plant species with predominant cleistogamous seed production (Sun, 1999; Durka *et al.* 2013), showing strong inbreeding, little or no genetic variability within populations and strong population differentiation (Culley and Wolfe 2001). In the case of the rare *V. elatior* additional factors like spatial isolation and past population bottlenecks may have aggravated this pattern. Epigenetic diversity, although in most populations slightly higher, in general was comparable to genetic diversity among populations, resulting in a significant correlation of genetic and epigenetic diversity estimates. This implies that epigenetic diversity largely depends on genetic diversity and at least partially underlies the same driving forces. Other studies surveying non-clonal species also observed equal or higher epigenetic than genetic diversity (Herrera and Bazaga 2010; Abratowska *et al.* 2012, Lira-Medeiros *et al.* 2010; Wu *et al.* 2013), suggesting that this is a common pattern in genetically diverse plant species.

For *V. elatior* neither habitat nor region had a profound effect on genetic or epigenetic diversity and only very small population size ($L2_W$) markedly reduced diversity. In contrast, previous epigenetic studies that investigated populations from different environments (Lira-Medeiros *et al.* 2010; Abratowska *et al.* 2012; Wu *et al.* 2013) found significant differences between habitats. This might have two reasons: first, these studies compared spatially and temporally stable environments instead of relatively dynamic habitat types along a successional gradient. If a less favorable habitat is associated with elevated levels of environmental stress that may lead to genome instability and higher mutation rates, under constant conditions this could be the driving force for an increase in genetic diversity (Wu *et al.* 2013), whereas under dynamic conditions this effect would be outbalanced between habitats or would only act for relatively short periods of time. Second, *V. elatior* possesses a long lived persistent soil seed bank (Eckstein *et al.* 2006). Through continuous recruitment of genotypes from past generations, the seed bank may buffer temporal variation in genetic diversity (Honnay *et al.* 2008).

Genetic and epigenetic structure

At the genetic level *V. elatior* showed very strong population differentiation ($\phi_{ST} = 0.72$) reflecting high selfing rates and small population sizes resulting in genetic drift. This is further corroborated by a lack of correlation between genetic differentiation and geographic distances indicating that spatial isolation did not play the major role for population differentiation.

Although there was a positive correlation between epigenetic and genetic distances, epigenetic population differentiation compared to genetic differentiation was markedly reduced and more closely related to habitat conditions. This indicates that environmentally-induced changes in methylation patterns may lead to a convergence of populations experiencing similar habitat conditions and thus counteract effects of historical demographic processes. Moreover, as revealed by PCoA, the different subepiloci types contributed differently to population structure, suggesting that the two methylation types

distinguished (i.e. CG-methylation and CHG-hemimethylation), possess different functional roles. Whereas n- and m-subepiloci both similarly separated the habitat types, h-subepiloci revealed hardly any population structure. This indicates that regulation of gene function by methylation or demethylation in the CG-context plays a more important role for habitat adjustment than changes of hemimethylation in the CHG-context. A functional difference between the two methylation types is further supported by the observation that of all polymorphic epiloci only 5.5% included both fragment types, whereas 51.7% and 23.6% included methylations either only in the CG- or CHG-context, respectively. These findings clearly show the need for separate analyses of CG- and CHG-methylated states, which are rarely distinguished in epigenetic population studies (but see Salmon *et al.* 2008, Paun *et al.* 2010).

Outlier detection

Outlier detection is a widely used approach to analyze putative selection processes using anonymous genetic marker data. However, genome scan approaches have been rarely used for MSAP data (Paun *et al.* 2010; Schrey *et al.* 2012), which may reflect the uncertainty if the central underlying assumption is valid for methylation variation, namely that heritable marker loci give rise to phenotypes which are targets of natural selection leading to change in frequencies of adaptive markers. First, epigenetic variation may depend on genetic variation (i.e. obligatory and facilitated epigenetic variation *sensu* Richards 2006). Hence, it is impossible to distinguish whether a certain epigenetic state or any underlying causative genetic state is putatively under selection. Second, epigenetic variation can be either generated randomly (epimutation) or environmentally induced, which has different implications for selection processes (Shea *et al.* 2011; Verhoeven and Preite 2014) and thus their detectability. Third, epigenetic variation can have different degrees of transgenerational stability (Herman *et al.* 2014), ranging from transient changes, over short-term (i.e. across two generations) to long-term heritable changes (i.e. several generations). The traditional view is that adaptive traits are selected for by natural selection only if traits are heritable. Thus, the degree of epigenetic heritability will impact the selective outcome. However, if over a long time the same advantageous phenotype arises every generation by transient epigenetic modifications it will also be selected for. Recent theory (Herman *et al.* 2014) and some first evidence in natural populations (Herrera *et al.* 2014) suggest that epigenetic stability itself is a selective trait, thus introducing another layer of complexity to the system. Lastly, in outlier locus analyses AFLP loci are used as markers that can track differentiation based on selection at linked loci. In contrast, the epigenetic interpretation of MSAP outlier loci is different as it must be assumed that the MSAP locus itself is causal to function. Taken together, epigenetic variation does not easily fit into the hitherto existing framework of selection and adaptation, and thus may not easily be suited for classical differentiation based genome scan analyses. This notwithstanding, correlation based genome scan approaches like SAM, which are largely assumption free, still offer a good opportunity to study putatively adaptive methylation variation in relationship to environmental conditions.

SAM analyses of *V. elatior* populations resulted in comparable percentages of light related outlier markers for genetic (17.0%) and epigenetic (14.2%) data. Given the limited power to detect significant marker associations due to the relative small number of surveyed populations and the use of averaged light values for all individuals from one site, the results imply that both levels of variation may play an important role for habitat related adaptation. However, as revealed by differentiation based genome scan approaches, for AFLP data the percentage of genetic outliers was strongly reduced and only very few of the SAM outliers (2 of 19) actually seemed to be under positive selection. Moreover, with DFDIST/FDIST (4.5%) and BAYESCAN (0%) the percentage of genetic outliers was strongly reduced, thus questioning a large impact of genetic selection. In contrast, PCoA results for neutral and outlier MSAP markers strongly support the hypothesis that methylation variation plays a major role in response to habitat conditions. Whereas for neutral markers genetic and epigenetic population structure were almost identical, for outlier markers only epigenetic variation led to a close clustering of the genetically well differentiated meadow populations. Interestingly, such clustering was not observed for woodland populations, suggesting that this habitat is more heterogeneous and thus entailing higher differentiation in response to other environmental parameters than light. A stronger habitat related impact of epigenetic variation is furthermore corroborated by the results of partial Mantel tests, showing that when controlling for geographical distance, only epigenetic but not genetic variation was significantly correlated with habitat similarity.

There is rising evidence from experimental studies that large proportions of stress-induced changes in epigenetic variation may be heritable across generations (Verhoeven *et al.* 2010, Kou *et al.* 2011). Verhoeven at al. (2010) found that ecological stressors led to methylation changes of 15–30% of polymorphic epiloci in apomictic dandelions. Strikingly, the vast majority of methylation changes (74–92%) were faithfully transmitted to the unstressed progeny and only a small proportion reverted back to the original state. Comparably high percentages of inheritance of methylation polymorphisms (75%) and subsequent persistence of multilocus epigenetic differentiation occurred in populations of *Helloborus foetidus* (Herrera *et al.* 2013). Thus, it seems plausible that also in the case of *V. elatior* a considerable amount of the epigenetic outliers detected by SAM are heritable and hence might enable 'soft inheritance'. However, as we surveyed only one generation of plants, our study cannot distinguish whether the observed epigenetic correlation with habitats reflects epigenetically mediated within-generation phenotypic plasticity or indeed heritable adaptations.

Behavior of subepilocus outliers

Our scoring approach allowed us to differentiate individual MSAP fragment types. The finding that at 95% CI only 3.8% of the h-subepiloci but 10.2% of the m-subepiloci were detected as outliers further indicates functional differences between the two methylation types. Even more interesting was the observation that the vast majority of outliers were n-subepiloci (15% of all n-subepiloci), suggesting that primarily the demethylated and thus probably active state of an epilocus is correlated with specific environmental condi-

tions. This is corroborated by the fact that for almost all m-subepiloci outliers also the corresponding n-subepiloci were outliers.

Of the outliers identified by SAM most n-subepiloci were related to low light and most m-subepiloci to high light, indicating a directional epigenetic regulation mechanism. Especially the corresponding n- and m-subepiloci outliers suggest a switch mechanism that activates stress-related genes by demethylation under low-light conditions and down-regulates them by methylation under high-light conditions. Indeed, biotic and abiotic stressors can induce selective demethylation processes and transcriptional activation of stress-related genes (Wada *et al.* 2004; Choi and Sano 2007). Moreover, global hypomethylation may be a response to environmental stressors (e.g. Zhong *et al.* 2009; Wang *et al.* 2011) or different habitats (Lira-Medeiros *et al.* 2010; Wu *et al.* 2013). However, stress-induced DNA hypermethylation has been observed, too (Labra *et al.* 2002; Lu *et al.* 2007) and also the few studies specifically addressing the effects of light availability are equivocal: whereas reduction of light led to genome hypermethylation in seagrass (Greco *et al.* 2013), low R/FR ratios as a signal of foliage shade resulted in genome hypomethylation in *Stellaria longipes* (Tatra *et al.* 2000). Additionally, demethylation was a crucial factor controlling stem elongation (Tatra *et al.* 2000) and thus might trigger shade-avoidance responses. Overall these and our results indicate that while demethylation is a common response to stress, species-specific and stress-specific responses are to be expected.

Conclusions

The results of our study suggest that DNA methylation variation plays a decisive role in the response of *V. elatior* to changing light conditions. Epigenetic population differentiation was more strongly related to habitat types than was genetic differentiation and epigenetic outlier loci led to a stronger habitat related population clustering than genetic outlier loci. This implies that under dynamic conditions the environmental shaping of the epigenome is a stronger force than selection changing the genome. Especially for rare and inbreeding species like *V. elatior* suffering from spatial isolation and small population sizes, genetically independent adaptive epigenetic variation may facilitate long-term population survival in dynamic and strongly contrasting habitats. Thus, provided that epigenetic variation is heritable, 'soft inheritance', i.e. environmentally induced methylation changes, represents an alternative system to classical 'hard inheritance'.

The application of a novel MSAP scoring approach revealed new insights into the contribution of methylation types to population differentiation and potential habitat related adaptation. Using a pure methylation-scoring approach would have obscured the association between the environment and demethylated fragments, thus missing information about the putative transcriptional active part of the epigenome, as well as about corresponding outlier pairs that may represent an epigenetic switch mechanism. Future epigenetic population studies using mixed scoring approaches will give further insights into the

interplay of methylation types and will show if the differences observed in this study represent a general epigenetic pattern under contrasting environmentally conditions.

Acknowledgements

We want to thank Josef Scholz vom Hofe and Lena Föhr for help during field work, and Aurélie Bonin, Ovidiu Paun and two anonymous reviewers for valuable comments on the manuscript. Financial support was obtained from the Deutsche Forschungsgemeinschaft (DFG, grant EC209/7-1).

References

Abratowska A, Wasowicz P, Bednarek PT, Telka J, Wierzbicka M (2012) Morphological and genetic distinctiveness of metallicolous and non-metallicolous populations of *Armeria maritima s.l.* (Plumbaginaceae) in Poland. *Plant Biology*, 14, 586–595.

Angers B, Castonguay E, Massicotte R (2010) Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Molecular Ecology*, 19, 1283–1295.

Antao T, Beaumont MA (2011) Mcheza: a workbench to detect selection using dominant markers. *Bioinformatics*, 27, 1717–1718.

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

Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature*, 447, 407–412.

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.

Boyko A, Blevins T, Yao Y *et al.* (2010) Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-like proteins. *PLoS ONE*, 5, e9514.

Boyko A, Kovalchuk I (2011) Genome instability and epigenetic modification—heritable responses to environmental stress? *Current Opinion in Plant Biology*, 14, 260–266.

Bräutigam K, Vining KJ, Lafon-Placette C *et al.* (2013) Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecology and Evolution*, 3, 399–415.

Castonguay E, Angers B (2012) The key role of epigenetics in the persistence of asexual lineages. *Genetics Research International*, 2012, 1–9.

Choi C-S, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Molecular Genetics and Genomics*, 277, 589–600.

Correia B, Valledor L, Meijón M *et al.* (2013) Is the interplay between epigenetic markers related to the acclimation of cork oak plants to high temperatures? *PLoS ONE*, 8, e53543.

Culley TM, Wolfe AD (2001) Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. *Heredity*, 86, 545–556.

Durka W, Nossol C, Welk E *et al.* (2012) Extreme genetic depauperation and differentiation of both populations and species in Eurasian feather grasses (*Stipa*). *Plant Systematics and Evolution*, 299, 259–269.

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.

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.

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, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science*, 330, 622–627.

Flatscher R, Frajman B, Schönswetter P, Paun O (2012) Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? *Genetics Research International*, 2012, 1–9.

Flores KB, Wolschin F, Amdam GV (2013) The role of methylation of DNA in environmental adaptation. *Integrative and Comparative Biology*, 53, 359–372.

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.

Frazer GW, Canham CD, Lertzman KP (1999) Gap Light Analyzer (GLA), Version 2.0: Imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, users manual and program documentation. Burnaby, BC, Canada: Simon Fraser University and Millbrook, NY, USA: Institute of Ecosystem

Greco M, Chiappetta A, Bruno L, Bitonti M (2013) Effects of light deficiency on genome methylation in *Posidonia oceanica*. *Marine Ecology Progress Series*, 473, 103–114.

Herman JJ, Spencer HG, Donohue K, Sultan SE (2014) How stable "should" epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution*, 68, 632–643.

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 (2013) Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Botanical Journal of the Linnean Society*, 171, 441–452.

Herrera CM, Medrano M, Bazaga P (2013) Epigenetic differentiation persists after male gametogenesis in natural populations of the perennial herb *Helleborus foetidus* (Ranunculaceae). *PLoS ONE*, 8, e70730.

Herrera CM, Medrano M, Bazaga P (2014) Variation in DNA methylation transmissibility, genetic heterogeneity and fecundity-related traits in natural populations of the perennial herb *Helleborus foetidus*. *Molecular Ecology*, 23, 1085–1095.

Honnay O, Bossuyt B, Jacquemyn H, Shimono A, Uchiyama K (2008) Can a seed bank maintain the genetic variation in the above ground plant population? *Oikos*, 117, 1–5.

Jablonka 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.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16, 3955–3969.

Kou HP, Li Y, Song XX *et al.* (2011) Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* L.). *Journal of Plant Physiology*, 168, 1685–1693.

Labra M, Ghiani A, Citterio S *et al.* (2002) Analysis of cytosine methylation pattern in response to water deficit in pea root tips. *Plant biology*, 4, 694–699.

Latzel V, Allan E, Bortolini Silveira A *et al.* (2013) Epigenetic diversity increases the productivity and stability of plant populations. *Nature Communications*, 4:2875, doi: 10.1038/ncomms3875

Ledon-Rettig CC (2013) Ecological epigenetics: An introduction to the symposium. *Integrative and Comparative Biology*, 53, 307–318.

Legendre P, Legendre L (1998) *Numerical ecology*. Elsevier Science, Amsterdam.

LiraMedeiros CF, Parisod C, Fernandes RA *et al.* (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE*, 5, e10326.

Lu G, Wu X, Chen B, Gao G, Xu K (2007) Evaluation of genetic and epigenetic modification in rapeseed (*Brassica napus*) induced by salt stress. *Journal of Integrative Plant Biology*, 49, 1599–1607.

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.

Nobis M, Hunziker U (2005) Automatic thresholding for hemispherical canopy-photographs based on edge detection. *Agricultural and Forest Meteorology*, 128, 243–250.

Oksanen J, Kindt R, Legendre P *et al.* (2007) vegan: Community ecology package version 1.15-1. Available at http://www.r-project.org.

Paun O, Bateman RM, Fay MF *et al.* (2010) Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorhiza*: 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.

Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395–401.

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.

Sahu PP, Pandey G, Sharma N *et al.* (2013) Epigenetic mechanisms of plant stress responses and adaptation. *Plant Cell Reports*, 32, 1151–1159.

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.

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.

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.

Schulz B, Eckstein RL, Durka W (2013) Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies. *Molecular Ecology Resources*, 13, 642–653.

Shea N, Pen I, Uller T (2011) Three epigenetic information channels and their different roles in evolution: Epigenetic mechanisms and evolution. *Journal of Evolutionary Biology*, 24, 1178–1187.

Sun M (1999) Cleistogamy in *Scutellaria indica* (Labiatae): effective mating system and population genetic structure. *Molecular Ecology*, 8, 1285–1295.

Tatra GS, Miranda J, Chinnappa CC, Reid DM (2000) Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of *Stellaria longipes*. *Physiologia Plantarum*, 109, 313–321.

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.

Verhoeven KJF, Preite V (2014) Epigenetic variation in asexually reproducing organisms. *Evolution*, 68, 644–655.

Wada Y, Miyamoto K, Kusano T, Sano H (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. *Molecular Genetics and Genomics*, 271, 658–666.

Wang W, Zhao X, Pan Y *et al.* (2011) DNA methylation changes detected by methylation-sensitive amplified polymorphism in two contrasting rice genotypes under salt stress. *Journal of Genetics and Genomics*, 38, 419–424.

Wu W-Q, Yi MR, Wang X-F *et al.* (2013) Genetic and epigenetic differentiation between natural *Betula ermanii* (Betulaceae) populations inhabiting contrasting habitats. *Tree Genetics and Genomes*, 9, 1321–1328.

Yu Y, Yang X, Wang H *et al.* (2013) Cytosine methylation alteration in natural populations of *Leymus chinensis* induced by multiple abiotic stresses. *PLoS ONE*, 8, e55772.

Zhang Y, Chen JM, Millerb JR (2005) Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, 133, 166–181.

Zhang Y-Y, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist*, 197, 314–322.

Zhong X, Wang Y, Liu X *et al.* (2009) DNA methylation polymorphism in annual wild soybean (*Glycine soja* Sieb. et Zucc.) and cultivated soybean (*G. max* L. Merr.). *Canadian Journal of Plant Science*, 89, 851–863.

Data accessibility

AFLP data, MSAP data and grid maps with individual sample and hemispherical photography coordinates are deposited at the Dryad repository under the accession doi:10.5061/dryad.m1b66.

Appendix

Appendix 3.1 AFLP and MSAP protocol

For restriction and ligation (RL) 5.6 µl genomic DNA were combined with 5.4 µl RL reaction mix containing 0.55 µl BSA (1 mg/ml; New England Biolabs, NEB), 1.1 µl 0.5 M NaCl, 5 u *Eco*RI (NEB), 1 u *Mse*I (NEB), 67 u T4 DNA ligase (NEB), 1.1 µl T4 DNA ligase buffer (NEB), 1 µl *Eco*RI adapter (5 pmol) and 1 µl *Mse*I adapter (50 pmol). The reaction was incubated for 2 h at 37 °C and diluted 1:2. For the preselective amplification (PCR1), 4 µl RL product were combined with 16 µl PCR1 reaction mix containing 1.5 ng/µl *Eco*RI- and *Mse*I preselective primers each, 200 µM dNTPs (Roth), 2 µl 10 x Dream Tag buffer (QIAGEN), 0.8 u Dream Tag polymerase (QIAGEN) and 9.84 µl H_2O. The thermocycler protocol was 72.0 °C (2 min) followed by 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min) and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler gradient. The PCR1 product was diluted 1:5. For the selective amplification (PCR2), 1 µl PCR1 product was combined with 3.4 µl PCR2 reaction mix containing 2.2 µl Multiplex PCR kit (QIAGEN) and 0.6 µl fluorescent labeled *Eco*RI primer (1 pmol/µl) and 0.6 µl *Mse*I (5 pmol/µl) selective primers each. The thermocycler protocol was 94.0 °C (2 min) followed by 10 cycles of 94.0 °C (20 s), 66.0 °C (30 s, decreasing 1 °C per cycle) and 72.0 °C (2 min) and 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min), and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler pro 384.

After an initial screening of 64 primer pairs, eight selective primer combinations (Supplementary Table S3.1) were chosen for AFLP analyses. Separation and visualization of the fragments was done on a ABI 3130 capillary sequencer (Applied Biosystems, Foster City, USA) with Genescan 500(-250) LIZ internal size standard (Applied Biosystems). GENMAPPER version 3.7 (Applied Biosystems) was used to analyze the AFLP profiles. Binning of fragments was done manually for all samples in one batch using a peak height threshold of 10 rfu. Peak height data were exported and for each fragment a specific peak height threshold was manually determined based on the peak height distribution which allowed scoring presence (1) and absence (0) of fragments. All loci that showed a monomorphic pattern or a deviation in only one individual were excluded from the data set to prevent biased parameter estimation. Error rate estimation was based on 33 replicate samples (20%) that were repeated, starting from the same DNA extracts.

The MSAP protocol was almost identical with the AFLP protocol, except exchanging the restriction enzyme *Mse*I by 5 u *Hpa*II or *Msp*I (Fermentas) and replacing the *Mse*I adaptor and primers by the respective *Hpa*II/*Msp*I adaptor and primers in equal concentrations.

Supplementary Table S3.1 Adaptor- and primer sequences used for AFLP and MSAP analyses

Primer	Sequence
Adaptors	
*Eco*RI-*adapter* top	5'-CTCGTAGACTGCGTACC-3'
*Eco*RI-*adapter* bottom	5'-AATTGGTACGCAGTCTAC-3'
*Mse*I-*adapter* top (AFLP)	5'-GAGCGATGAGTCCTGAG-3'
*Mse*I-*adapter* bottom (AFLP)	3'-TACTCAGGACTCAT-5'
*Hpa*II/*Msp*I-*adapter* top (MSAP)	5'-GATCATGAGTCCTGCT-3'
*Hpa*II/*Msp*I -adapter bottom (MSAP)	5'-CGAGCAGGACTCATGA-3'
Preselective primers	
*Eco*RI + A	5'-GACTGCGTACCAATTCA-3'
*Mse*I + C (AFLP)	5' GATGAGTCCTGAGTAAC-3'
*Hpa*II/*Msp*I (MSAP)	5'-ATCATGAGTCCTGCTCGG-3'
Selective primer AFLP	
*Eco*RI + AAC-FAM[1]	5'-GACTGCGTACCAATTCAAC-3'
*Eco*RI + ACT-FAM[2]	5'-GACTGCGTACCAATTCACT-3'
*Eco*RI + ACA-VIC[3,4]	5'-GACTGCGTACCAATTCACA-3'
*Eco*RI + AAG-NED[5,6]	5'-GACTGCGTACCAATTCAAG-3'
*Eco*RI + AGC-PET[7]	5'-GACTGCGTACCAATTCAGC-3'
*Eco*RI + AGG-PET[8]	5'-GACTGCGTACCAATTCAGG-3'
*Mse*I + CTA[1]	5'-GATGAGTCCTGAGTAACTA-3'
*Mse*I + CAA[2,3,7]	5'-GATGAGTCCTGAGTAACAA-3'
*Mse*I + CAC[5]	5'-GATGAGTCCTGAGTAACAC-3'
*Mse*I + CTC[4,6,8]	5'-GATGAGTCCTGAGTAACTC-3'
Selective primers MSAP	
*Eco*RI + AAC-FAM[1,2]	5'-GACTGCGTACCAATTCAAC-3'
*Eco*RI + ACA-VIC[3,4]	5'-GACTGCGTACCAATTCACA-3'
*Eco*RI + AAG-NED[5,6]	5'-GACTGCGTACCAATTCAAG-3'
*Eco*RI + AGG-PET[7,8]	5'-GACTGCGTACCAATTCAGG-3'
*Hpa*II/*Msp*I + TCA[1,3]	5'-ATCATGAGTCCTGCTCGGTCA-3'
*Hpa*II/*Msp*I + TCC[2,5]	5'-ATCATGAGTCCTGCTCGGTCC-3'
*Hpa*II/*Msp*I + TCG[6]	5'-ATCATGAGTCCTGCTCGGTCG-3'
*Hpa*II/*Msp*I + TCCA[4,7]	5'-ATCATGAGTCCTGCTCGGTCCA-3'
*Hpa*II/*Msp*I + TCTA[8]	5'-ATCATGAGTCCTGCTCGGTCTA-3'

Superscript numbers indicate primer combinations used for the selective amplification.

Supplementary Table S3.2 Pairwise population ϕ_{ST} for 112 AFLP loci (upper diagonal) and 444 MSAP subepiloci (lower diagonal)

	$K1_M$	$K2_M$	$K3_W$	$L1_M$	$L2_W$	$L3_W$
$K1_M$	-	0.65	0.68	0.76	0.76	0.65
$K2_M$	0.35	-	0.66	0.72	0.76	0.64
$K3_W$	0.46	0.46	-	0.77	0.84	0.70
$L1_M$	0.54	0.49	0.57	-	0.84	0.72
$L2_W$	0.55	0.55	0.57	0.67	-	0.52
$L3_W$	0.40	0.38	0.49	0.57	0.47	-

Supplementary Table S3.3 Results of simple and partial Mantel tests for genetic and epigenetic pairwise population ϕ_{ST} with geographic distance and habitat matrices

	Simple		Partial	
	Geography	Habitat	Geography controlling for habitat	Habitat controlling for geography
AFLP	r = 0.30 *p* = 0.18	r = 0.13 *p* = 0.43	r = 0.32 *p* = 0.16	r = 0.17 *p* = 0.30
MSAP all subepiloci	r = -0.01 *p* = 0.53	r = 0.16 *p* = 0.10	r = 0.002 *p* = 0.51	**r = 0.16** ***p* = 0.04**

Mantel r and *p*-values in bold indicate significant relationships

Supplementary Table S3.4 List of putative adaptive AFLP loci in the six surveyed populations of *Viola elatior* as assessed by differentiation based genome scan approaches (DFDIST/FDIST and BAYESCAN) and SAM analyses for correlation with light availability. The marker code depicts the selective primers used for fragment amplification and the respective fragment length in base pairs. Markers in bold were identified by two different genome scan approaches.

Marker Code	DFDIST/FDIST	BAYESCAN	SAM (transmitted PAR)
AAC_CTA_201	**	-	***
AAC_CTA_209	***	-	***
AAC_CTA_331	-	-	***
AAG_CAC_160	-	-	***
AAG_CAC_175	**	-	-
AAG_CAC_271	-	-	***
AAG_CAC_274	-	-	***
AAG_CTC_128	**	-	-
AAG_CTC_150	-	-	**
AAG_CTC_151	-	-	***
AAG_CTC_156	-	-	***
ACA_CAA_213	-	-	***
ACA_CTC_337	-	-	***
ACA_CTC_422	-	-	***
AGC_CAA_121	-	-	***
AGC_CAA_250			***
AGC_CAA_195	**	-	-
AGC_CAA_264	-	-	***
AGC_CAA_291	-	-	***
AGG_CTC_165	-	-	**
AGG_CTC_504	-	-	***
AGG_CTC_59	-	-	***

** $p < 0.05$, *** $p < 0.01$

Supplementary Table S3.5 List of putative adaptive MSAP subepiloci in the six surveyed populations of *Viola elatior* as assessed by SAM analyses for correlation with light environment. The marker code depicts the selective primers used for fragment amplification and the respective fragment length in base pairs. Markers with the superscript COR indicate corresponding pairs of outlier subepiloci (i.e. subepiloci resulting from one epilocus in the MSAP raw data matrix that both are identified as being outliers).

Marker code	Subepiloci	SAM (transmitted PAR)
AAC_TCA_342	n	**
AAC_TCA_410	h	***
AAC_TCC_191^{COR}	n	**
AAC_TCC_191^{COR}	m	**
AAC_TCC_224^{COR}	n	***
AAC_TCC_224^{COR}	m	***
AAC_TCC_227	m	**
AAC_TCC_253	n	**
AAC_TCC_258	n	***
AAC_TCC_259	n	***
AAC_TCC_261	n	***
AAC_TCC_390	n	***
AAC_TCC_394	n	***
AAC_TCC_430	n	**
AAG_TCC_252	n	***
AAG_TCC_381	n	***
AAG_TCG_249^{COR}	n	***
AAG_TCG_249^{COR}	m	***
AAG_TCG_254^{COR}	n	**
AAG_TCG_254^{COR}	m	**
AAG_TCG_315	m	**
AAG_TCG_320	n	***
AAG_TCG_494	n	**
ACA_TCA_168^{COR}	n	**
ACA_TCA_168^{COR}	m	**
ACA_TCA_213	n	**
ACA_TCA_217	m	***
ACA_TCA_234^{COR}	n	***
ACA_TCA_234^{COR}	m	***
ACA_TCA_240	n	***
ACA_TCA_354	h	***
ACA_TCA_366	n	***
ACA_TCA_374	n	**
ACA_TCCA_188^{COR}	n	***
ACA_TCCA_188^{COR}	m	***
ACA_TCCA_214	n	**
ACA_TCCA_354	h	***
AGG_TCCA_240	n	***
AGG_TCCA_241	n	***
AGG_TCCA_265^{COR}	n	***
AGG_TCCA_265^{COR}	m	**
AGG_TCCA_359	n	***
AGG_TCCA_55	m	**
AGG_TCCA_63^{COR}	n	**
AGG_TCCA_63^{COR}	m	**
AGG_TCTA_280^{COR}	n	***
AGG_TCTA_280^{COR}	m	***
AGG_TCTA_356^{COR}	n	***
AGG_TCTA_356^{COR}	m	***
AGG_TCTA_402	m	**

** $p < 0.05$, *** $p < 0.01$

CHAPTER 4

CAN PERSISTENT SEED BANKS BUFFER GENETIC EFFECTS OF DECLINING POPULATION SIZE AND SELECTION?

Benjamin Schulz, Walter Durka, Jiří Danihelka and R. Lutz Eckstein
Submitted manuscript

Abstract

Persistent seed banks are predicted to have an important impact on population genetic processes by increasing effective population size and probably storing past genetic diversity. Accordingly, they might buffer genetic effects of disturbance, fragmentation and/or selection. However, empirical studies surveying the relationship between aboveground (AG) and seed bank (SB) genetics under changing environments are scarce and so far based only on small data sets. Here, we compared AG and SB cohorts in 15 populations of the cleistogamous *Viola elatior* in two contrasting habitats along a successional gradient characterized by strong differences in light-availability and declining population size. Using AFLP markers, we found significantly higher AG than SB genetic diversity in early (band richness Br = 1.20 vs. 1.17; pairwise private band richness PBr_p = 0.06 vs. 0.02) but not in late successional habitats (Br = 1.14 vs. 1.13; PBr_p = 0.07 vs. 0.05). Moreover, three of eight late successional populations even showed higher SB than AG diversity, and differentiation between AG and SB was less frequent in early than in late successional habitats (AMOVA: 29% vs. 50%). Our results indicate that in *V. elatior* relative SB diversity (i.e. compared to AG diversity) and thus the SB genetic buffer capacity increases with ongoing succession and despite decreasing population size. This will counteract effects of drift and selection, and seems to assure a higher chance for the species' long term persistence after disturbance events. Corroborated by much lower small-scale genetic structure in late successional habitats, we suggest that the observed change in relative SB diversity is initially driven by a change in outcrossing rates along the gradient.

Introduction

Instead of relying solely on the spatial dispersal of their seeds, many plant species have developed the ability to additionally spread their offspring on a temporal scale by accumulating long-lived, dormant seeds over time in the ground or in aerial reservoirs. Such persistent seed banks are a common feature of plants to counteract the consequences of environmental or demographic stochasticities and are present in species across a wide range of life history types, habitats and climate zones (reviewed in Leck *et al.* 1989; Baskin and Baskin 2001). Particularly for rare and isolated species or for species from highly dynamic or disturbed habitats, persistent seed banks are thought to play a crucial role for population dynamics and stability, and potentially are capable to replace standing individuals after bottlenecks or extinction events (Hölzel and Otte 2004; Honnay *et al.* 2005).

Besides that, theory predicts that seed banks have an important impact on population genetic processes as they can consist of progeny produced in the course of several generations and probably under varying selection regimes (Templeton and Levin 1979; Tonsor 1993). Accordingly, persistent seed banks increase effective population size (Vitalis *et al.* 2004) and may store even more genetic diversity than present in standing plant populations (Templeton and Levin 1979). Thus, they would buffer deleterious effects of random genetic drift and weaken patterns of genetic population structuring (Templeton and Levin 1979; McCue and Holtsford 1998). In addition, by enabling gene flow from past generations and maintaining genes in populations through periods in which they are selected against (Tonsor *et al.* 1993), seed banks could have the potential to slow down adaptation processes and damp out directional selection in response to environmental fluctuations.

In the last years many studies were carried out to empirically test these hypotheses and find evidence for the ecological and evolutionary impact of long-lived seed reservoirs. While some of them clearly could show that persistent seed banks can increase effective population size, both in annuals (Lundemo *et al.* 2009; Hanin *et al.* 2013) and perennials (Falahati-Anbaran *et al.* 2011), so far it was impossible to conclusively confirm the potential of seed banks to accumulate genetic diversity and hence to serve as a genetic memory that might influences the evolutionary fate of populations (Mandák *et al.* 2012). Some studies indeed found higher genetic diversity in seed banks (McCue and Holtsford 1998; Morris *et al.* 2002), however, most others exhibited higher diversity estimates in above-ground plant population (e.g. Tonsor *et al.* 1993; Cabin *et al.* 1998; Mandák *et al.* 2006) or could detect no significant differences at all (e.g. Shimono *et al.* 2006; Lundemo *et al.* 2009; Falahati-Anbaran *et al.* 2011; Hanin *et al.* 2013).

To shed some more light on this topic Honney *et al.* (2008) conducted a meta-analysis on 13 studies that compared the genetic diversity of seed banks and standing plant populations. Interestingly, whereas levels of heterozygosity and percentage of polymorphic loci overall appeared to be similar in the two groups, allele counts after correction for sample size were significantly higher in the seed bank. It was concluded that the differences in allele numbers thus mainly are driven by rare alleles and that selection might act as filter on seed bank alleles, preventing some of them to be established in standing plants. The

analysis additionally revealed significantly higher inbreeding and more homozygotes in the seed bank, substantiating the results of earlier studies that surveyed heterozygosity levels across plant life stages and reported a gradual increase of heterozygosity towards the adult stage (Lesica and Allendorf 1992; Alvarez-Buylla *et al.* 1996; Mandák *et al.* 2006). As discussed in detail by Vitalis *et al.* (2004) this pattern is most likely explained as an effect of selection that in the course of seed germination and recruitment progressively eliminates less fit homozygotes (Mandák *et al.* 2006).

Overall, the study of Honnay *et al.* (2008) thus gave no evidence that high levels of genetic diversity are accumulating in long-lived seed reservoirs. Instead, it was concluded that any difference in the genetic composition between seed bank and standing plants is rather the result of local selection than a buffering effect of stored seeds (Vitalis *et al.* 2004; Honnay *et al.* 2008). As a consequence, the authors reasoned that it seems not very fruitful to continue surveying the genetic diversity of the two groups, unless this is performed under different selection regimes, in order to compare the outcome of the selection process (Honnay *et al.* 2008). However, until today the aspect of selection has been largely omitted in empirical research on seed bank genetics. To our knowledge, only three studies have yet compared seed bank and standing plants in contrasting environments and thus potentially under different selection regimes. Koch *et al.* (2003) investigated *Cardamine amara* in four habitat types along a gradient of increasing flooding and found that heterozygosity of both, seed bank and standing plants, tended to decrease with decreasing disturbance frequency. Uchiyama *et al.* (2006) examined *Betula maximowicziana* in low and high-density stands, and detected differing estimates of private alleles and linkage disequilibrium in above-ground plants but not in the seed bank. Finally, Hanin *et al.* (2013) compared populations of *Eruca sativa* in desert and semi-arid habitats and found higher seed density and greater seed longevity in the latter. Notwithstanding, none of these studies could detect clear differences in genetic diversity, neither between seed bank and above-ground plants nor among habitats. At least partially, this might be attributed to comparatively small data sets and hence the intrinsic characteristics of particular populations, as in Hanin *et al.* (2013) both surveyed habitats, and in Koch *et al.* (2003) two of the four habitats were represented by only one locality.

To fill the existing gap in current knowledge and to follow up the conclusions of Honnay *et al.* (2008), here we compared the genetic variation of seed bank and above-ground plants in the perennial flood plain species *Viola elatior* along a successional gradient ranging from managed open grassland to closed alluvial forests. The ability of the study species to compete for light is low (Moora *et al.* 2003), consequently with increasing succession to closed forests, population sizes gradually decline and the species finally disappears from the aboveground vegetation (Eckstein *et al.* 2006a). As light availability is a strong environmental cue that influences an array of other biotic and abiotic parameters (e.g. competition, water availability or temperature), the surveyed gradient appears well suited to study the outcome of local selection on the interplay of seed bank and above-ground genetics. Moreover, as there is an urgent need for studies that provide data for more than only a few populations and that cover a large spatial range and thus allow for conclusions on any broad genetic differences (Mandák *et al.* 2012), we conducted a multi

population study with 15 localities, covering two main distribution areas of *Viola elatior* in Germany and the Czech Republic. Using amplified fragment length polymorphisms (AFLP) we asked the following questions: (1) Can the persistent seed bank of *Viola elatior* maintain genetic diversity under decreasing population size and a changing selection regime? (2) Do contrasting habitat types impact the small-scale spatial genetic structure of the seed bank and/or the aboveground plants?

Materials and Methods

Study species

Viola elatior (Violaceae) is a perennial iteroparous hemicryptophyte belonging to the section *Viola*, subsect. *Rostratae* (Eckstein *et al.* 2006a). The species' distribution roughly covers the submeridional and temperate zone of western Eurasia ranging from the Parisian basin to southern Siberia (Meusel *et al.* 1978). Whereas in its core area with summer-warm continental climates the species is found in steppe and forest-steppe vegetation, in Central Europe towards the western border of its distribution, *V. elatior* is confined to large river corridors (Eckstein *et al.* 2006a; Danihelka *et al.* 2009). Here it becomes increasingly rare and occurs in different floodplain habitats along a successional gradient, extending from managed open floodplain meadows to alluvial woodland fringes (Eckstein and Otte, 2005; Eckstein *et al.* 2006a). Population sizes can vary between tens and hundreds of individuals (Eckstein *et al.* 2006a).

Viola elatior has an octoploid genome (2n = 40) and exhibits a mixed mating system with potentially cross-pollinated chasmogamous (CH) and obligatory self-pollinated cleistogamous (CL) flowers, producing capsules with approximately equal numbers of seeds (~30). Nonetheless, seed production through CL flowers is dominating, resulting in a very high selfing rate. In common garden experiments only around 4% of total capsule production consisted of CH capsules (Eckstein and Otte, 2005; Schulz unpublished).

As many other violets, *V. elatior* builds up persistent soil seed banks that may lead to high seed accumulation in the ground. Hölzel and Otte (2004) found maximum seed densities of up to 2660 germinable seeds/m^2 under a densely populated floodplain meadow, with more than 80% of all seeds concentrated in the upper 5 cm of the soil layer. Especially under strongly fluctuating conditions of floodplain habitats, the seed bank seems to be an important part of the species' life strategy, which is illustrated by various reports about sudden emergence of plants in the course of disturbance events after long-term absence from the aboveground vegetation (Eckstein *et al.* 2006a, and reference therein).

Study regions

The study was conducted in two regions that represent strongholds of *V. elatior* in Europe: The Upper Rhine floodplain (R) south of Frankfurt am Main, Germany, and the Thaya/Morava floodplain (T) around Břeclav, Czech Republic. In the densely populated

and highly agriculturally influenced Upper Rhine area the study species is only found in some nature reserves with a high share of open floodplain meadows that are regularly managed by mowing or grazing and thus provide high proportions of suitable early- and mid-successional habitats. In contrast, large parts of the Thaya/Morava floodplain are less influenced by settlements and intense land-use, and the landscape is rather characterized by a higher percentage of forests and non-intensively managed patches of floodplain meadows. Here, populations of *V. elatior* are more widely scattered and mostly occur in late-successional habitats within forest stands or along forest fringes (Eckstein *et al.* 2006b).

Table 4.1 Overview of surveyed populations of *Viola elatior*

Population	Region	Latitude	Longitude	AG samples	SB samples*	Grid cells (m^2)	Habitat type	Mean transmitted PAR ± SD (%)
RM1	Ger	49°50'16"N	8°24'00"E	21	20 (11)	14	meadow	95.9 ± 0.2
RM2	Ger	49°50'01"N	8°25'32"E	19	20 (12)	~1400	meadow	87.7 ± 6.5
RM3	Ger	49°49'49"N	8°28'03"E	22	12 (8)	159	meadow	73.8 ± 8.3
RM4	Ger	49°36'08"N	8°26'50"E	22	19 (7)	170	meadow	80.5 ± 14.7
RW1	Ger	49°48'50"N	8°24'57"E	23	21 (10)	158	woodland	23.6 ± 14.2
RW2	Ger	49°35'49"N	8°26'50"E	20	20 (12)	31	woodland	12.5 ± 3.3
RW3	Ger	49°35'44"N	8°25'55"E	22	19 (10)	144	woodland	16.5 ± 9.4
TM1	Cz	48°45'50"N	16°51'57"E	21	23 (14)	32	meadow	48.3 ± 21.2
TM2	Cz	48°46'52"N	16°51'48"E	21	23 (13)	~1200	meadow	83.0 ± 4.7
TM3	Cz	48°48'51"N	16°49'53"E	21	22 (14)	160	meadow	60.1 ± 14.9
TW1	Cz	48°49'00"N	16°27'08"E	23	23 (14)	131	woodland	22.4 ± 11.7
TW2	Cz	48°49'25"N	16°46'26"E	23	23 (14)	59	woodland	25.3 ± 20.2
TW3	Cz	48°49'01"N	16°47'41"E	23	22 (13)	34	woodland	14.8 ± 7.4
TW4	Cz	48°38'21"N	16°57'19"E	22	21 (12)	56	woodland	13.1 ± 3.2
TW5	Cz	48°58'30"N	17°23'09"E	21	23 (15)	59	woodland	30.6± 13.4

* Numbers in brackets indicate germination trays that contributed with one or more emerged seedlings to the SB samples, respectively; Ger – Upper Rhine, Germany; Cz – Thaya/Morava, Czech Republic; AG – aboveground; SB – seed bank; PAR – photosynthetic active radiation.

Sampling design

In both regions we surveyed stands (hereafter called populations) from each of the two extremes of the species' environmental range, i.e. sunny floodplain meadows (M) and shady alluvial woodland fringes (W). First, all known sites in the two regions inhabiting *V. elatior* were inspected and visually classified according to their light environment. The light environment of each population (Table 4.1) was measured with hemispherical photography as mean daily percentages of transmitted total photosynthetic active radiation (for details see Schulz *et al.* 2014). Then, representative populations of both habitat types (Table 4.1) were selected and aboveground (AG) and seed bank (SB) samples (hereafter called AG and SB cohorts) were collected between May and June, i.e. after the spring germination peak and before the new seed rain (but see Mandák *et al.* 2012). Due to limited numbers of appropriate populations in the different habitat types with more than 20 AG individuals, we chose four M- and three W-sites in region R and three M- and five W-sites in region T (Table 4.1). Distances between populations ranged from 0.5 to 27 km in region R and from 1.6 to 70 km in region T.

To capture a maximum of allelic diversity and to detect potential spatial genetic structure within populations, we adopted a grid-based randomized sampling protocol. Therefore, in each population the presence of *V. elatior* was mapped on a 1 m grid. Then, for AG cohort sampling at each site 19–23 populated grid cells were randomly selected and young and undamaged leaves from one plant per cell were collected. As an exception, for RM1 2–3 individuals were sampled in 6 of the grid cells as the total number of populated grid cells did not exceed 14. Samples were immediately cooled to below 10 °C, stored at -25 °C and were then freeze-dried for 48 h.

For SB cohort sampling we applied the seedling emergence method. Using a soil corer of 5 cm in diameter and 4 cm in depth, we took 5 soil samples each in 30 randomly selected grid cells per population. The 150 soil samples represented 0.3 m^2 of soil surface and 11.8 l of soil volume. As previous studies showed that *V. elatior* seeds are strongly dormant during summer, with almost no germination between May and September (Eckstein *et al.* 2006a), the soil cores were stored dry and dark until autumn. To reduce the volume and to optimize germination conditions, we concentrated the soil samples by washing through two sieves with mesh sizes of 2.4 and 0.7 mm (ter Heerdt *et al.* 1996). Seeds of *V. elatior* (diameter: 1.2–2.2 mm) thus accumulated in the middle fraction, whereas larger and smaller soil components were removed. Afterwards, the concentrated soil samples were spread in a ~0.5 cm thick layer on sterilized potting soil in 18 cm x 28 cm styrofoam trays. Due to logistic reasons, soil concentrates of two proximate grid cells were always pooled in one tray, respectively, resulting in 15 soil pools per population. For stratification, the trays were exposed in free air conditions starting in late November. From the following spring to autumn the trays were watered regularly, and germinated *V. elatior* seedlings were identified and carefully transferred to individual pots once every month. The potted seedlings were grown in a greenhouse until they reached the four-leaf stage and were then harvested, stored at -25 °C and finally freeze-dried for 48 h. Depending on germination success, 12–23 SB individuals per population were cho-

sen for genetic analyses. To obtain a comparable spatial sample resolution as for AG individuals, whenever possible only one to two samples per germination tray were selected (Table 4.1).

In summary, for each of the four region x habitat combinations we sampled between 3 and 5 populations, consisting of 19–23 AG cohort individuals and 12–23 SB cohort individuals, respectively (Table 4.1).

AFLP genotyping

We investigated a total of 324 AG and 311 SB samples with amplified fragment length polymorphism (AFLP). Total genomic DNA was extracted from dried leaf tissue using the DNeasy 96 Plant extraction kit (QIAGEN). AFLP methodology followed Kloss *et al.* (2011) and is described in detail in Appendix 4.1. After an initial screening of 64 primer pairs, eight selective primer combinations (Supplementary Table S4.1) were chosen for AFLP analyses. Separation and visualization of fragments was done on a ABI 3130 capillary sequencer (Applied Biosystems, Foster City, USA) with Genescan 500(-250) LIZ internal size standard (Applied Biosystems). GENMAPPER version 3.7 (Applied Biosystems) was used to analyze the AFLP profiles. We binned fragments manually for all samples in one batch using a peak height threshold of 10 rfu. Then, peak height data were exported and for each fragment a specific peak height threshold was manually determined based on the peak height distribution which allowed scoring presence (1) and absence (0) of fragments. All loci that showed a monomorphic pattern or a deviation in only one individual were excluded from the data set to prevent biased parameter estimation. Overall error rate was 0.6%, based on 58 replicate samples (9%) that were repeated starting with DNA extraction.

Data analysis

Binary AFLP data were analyzed using a band- or marker-based strategy, that is, without calculating allele frequencies. To account for the unequal sample size of AG and SB individuals (Table 4.1) we estimated genetic diversity within cohorts and within pooled samples (i.e. AG + SB) with a rarefaction-based approach as band richness (*Br*) and as percentage of polymorphic loci at the 5% level (*PLP)* using AFLPDIV 1.1 (Coart *et al.* 2005) and a standardized sample size equal to the smallest sample population (i.e. n = 12). Furthermore, to evaluate the number of bands within populations that are private either only for AG or SB individuals, we calculated pairwise private band richness (PBr_p) with rarefaction analyses according to Kalinowski (2004) separately for each population using ADZE 1.0 (Szpiech *et al.* 2008) and the same standardized sample size of n = 12.

Significance of differences between diversity estimates was tested using two-tailed *t*-tests and by calculating natural-logarithmic response ratios (LnRR) as proposed by Goldberg and Scheiner (2001):

$$LnRR = Ln(\frac{P_{SB}}{P_{AG}})$$

where P_{SB} is the mean value of SB cohorts and P_{AG} is the mean value of AG cohorts. Differences between cohorts were considered significant when the 95% confidence interval did not overlap with zero.

Genetic variation among groups of populations (ϕ_{CT}), among populations within groups (ϕ_{SC}) and within populations (ϕ_{ST}) was partitioned with hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010). To test for differentiation between AG and SB within each population, pairwise ϕ_{ST} values were calculated among cohorts and significance levels were determined after 9999 permutations. Clustering of individual samples was examined with principal component analysis (PCA) using the R package ADEGENET v1.4-2 (Jombart 2008).

To examine the small-scale spatial genetic structure (SGS) within habitats and AG and SB cohorts we used spatial autocorrelation methods implemented in SPAGeDi v1.4 (Hardy and Vekemans 2002). We chose distance limits of 4, 8, 12, 16 and 20 m to assure a sufficient number of individual pairs per distance class. For SB cohorts, only samples originating from soil pools with a maximum distance of 4 m between the two soil sampling sites (see SB sampling strategy above) were considered, using the corresponding midpoint coordinates for the SGS analyses, respectively. Thus, potential biases from the actual SB coordinates lie within the chosen distance classes. To construct spatial autocorrelograms, pairwise kinship coefficients (F_{ij}) for dominant markers (Hardy 2003) were calculated assuming an inbreeding coefficient of 0.5. Using higher inbreeding coefficients of up to 0.9 in additional trials had little effect on the results and did not change the general conclusions. Significance of mean F_{ij} per distance class was tested with 9999 permutations of multilocus genotypes. We quantified SGS for each population using restricted regression analyses (0–20 m) and estimating the *Sp* statistic (Vekemans and Hardy 2004) as $Sp = -b_{log}/(1-F_{(1)})$, where b_{log} is the regression slope of mean F_{ij} on log geographic distance and $F_{(1)}$ is the mean F_{ij} of the first distance class. To compare autocorrelation patterns, we furthermore pooled populations according to habitats and regions and tested for heterogeneous autocorrelation with heterogeneity tests for multiple populations subsets (Smouse *et al.* 2008) using GenAlEx 6.5 (Peakall and Smouse 2012) applying the same distance classes as above. Number of permutations and bootstraps were set to 9999, respectively. Following Banks and Peakall (2012) significance of the Heterogeneity Test is declared when $p < 0.01$.

Results

Seedling emergence rates

Total numbers of seedlings that emerged from the concentrated soil cores strongly varied between populations (Supplementary Table S4.2) ranging from 12 in RM3 (40 seedlings/m²) to 206 in TW2 (687.7 seedlings/m²). Overall, mean values of germinated seedlings differed significantly between regions (R = 24.9±9.4 and T = 104.8±48.5; *t*-test, $p = 0.003$) but not between habitat types (meadow = 45.1±34.7 and woodland = 87±59.4; *t*-test, $p = 0.15$).

Table 4.2 Measures of within-population diversity of *Viola elatior* for aboveground, seed bank and pooled individuals

Population	PS_{AFLP}		PBr_p		*Br*		*PLP* 5%		*Br*	*PLP* 5%
	AG	SB	AG	SB	AG	SB	AG	SB	AG + SB	AG + SB
RM1	33	60	0.07	0.03	1.16	1.12	19	16	1.14	22
RM2	16	40	0.09	0.02	1.31	1.25	35	31	1.28	38
RM3	55	50	0.03	0.01	1.34	1.32	34	32	1.33	35
RM4	23	58	0.06	0.06	1.15	1.15	18	17	1.16	25
RW1	13	33	0.07	0.06	1.17	1.16	20	20	1.17	27
RW2	45	25	0.03	0.08	1.06	1.12	9	16	1.10	20
RW3	50	53	0.07	0.04	1.08	1.06	11	8	1.09	16
TM1	57	87	0.10	0.01	1.10	1.01	13	2	1.06	14
TM2	29	52	0.02	0.02	1.06	1.06	7	8	1.06	10
TM3	10	36	0.03	0.02	1.29	1.28	31	29	1.28	31
TW1	13	30	0.04	0.03	1.09	1.08	13	11	1.09	16
TW2	57	74	0.04	0.01	1.05	1.02	8	3	1.04	9
TW3	35	36	0.05	0.15	1.31	1.40	35	50	1.37	55
TW4	0	19	0.22	0.05	1.34	1.17	43	20	1.29	46
TW5	57	48	0.01	0.02	1.04	1.05	5	6	1.04	7
Average overall	*33*	*47*	*0.06*	*0.04*	*1.17*	*1.15*	*20*	*18*	*1.17*	*25*
Average meadow	*32*	*55*	*0.06*	*0.02*	*1.20*	*1.17*	*22*	*19*	*1.19*	*25*
Average woodland	*34*	*40*	*0.07*	*0.05*	*1.14*	*1.13*	*18*	*17*	*1.15*	*25*
Average R	*34*	*46*	*0.06*	*0.04*	*1.18*	*1.17*	*21*	*20*	*1.18*	*26*
Average T	*32*	*48*	*0.06*	*0.04*	*1.16*	*1.13*	*19*	*16*	*1.15*	*24*

AG – aboveground individuals; SB – seed bank individuals; region R – Upper Rhine floodplain; region T – Thaya/Morava floodplain; PS_{AFLP} – percentage of individuals that share their AFLP phenotype with one or more other individuals within the cohort; *Br* – band richness, *PLP* 5% – percentage of polymorphic loci at the 5% level; PBr_p – pairwise private band richness (AG vs. SB); *Br*, *PLP* 5% and PBr_p were calculated after rarefaction to the minimal sample size (12).

Genetic diversity

AFLP analysis resulted in a total of 528 scorable loci of which 128 (24%) were polymorphic. Across the surveyed samples we found 323 unique AFLP phenotypes, that is, 49% of all individuals shared their AFLP phenotypes with at least one other individual, ranging within cohorts from 0% in TW4-AG to 87% in TM1-SB (Table 4.2). The mean percentage of individuals with shared AFLP phenotypes differed significantly between AG and SB samples (AG = 33% and SB = 47%; *t*-test, $p = 0.005$). However, when grouped for habitats significant differences remained only for meadow (AG = 32% and SB = 55%; *t*-test, $p = 0.003$) but not for woodland habitats (AG = 34% and SB = 40%; *t*-test, $p = 0.286$).

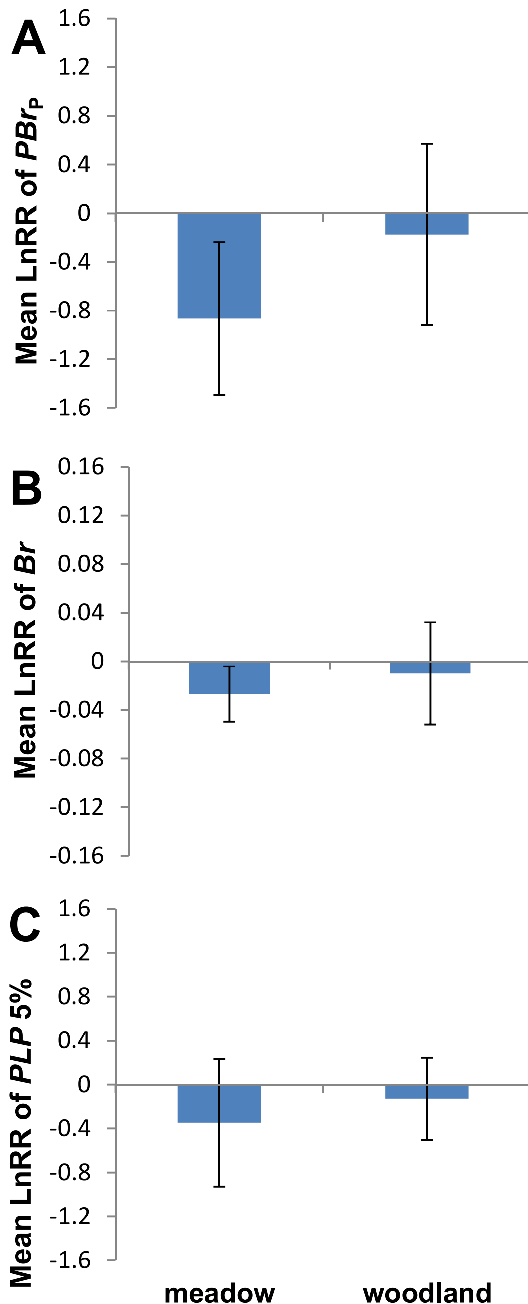


Figure 4.1 Mean (±95% CI) ln response ratio of (A) PBr_p, (B) *Br* and (C) *PLP*5% between SB and AG cohorts in meadow and in woodland habitats, respectively. Negative ln response ratios denote higher values in the AG cohort. Differences between AG and SB cohorts were considered significant when 95% CI did not overlap with zero.

As we assume that the majority of individuals with shared AFLP phenotypes do not represent vegetative clones but closely related, inbred individuals, all samples were retained in the subsequent analyses. After rarefaction, assessment of genetic diversity across populations (Table 4.2) revealed mean values of 1.17 (AG) and 1.15 (SB) for *Br*, 20% (AG) and 18% (SB) for *PLP,* and 0.06 (AG) and 0.04 (SB) for PBr_p. Significant differences between the two cohorts were only found in meadow populations for *Br* and PBr_p, with higher values in AG than in SB individuals (*Br*: 1.20 and 1.17, respectively; *t*-test, $p = 0.049$; PBr_p: 0.055 and 0.024, respectively; *t*-test, $p = 0.049$). This finding was furthermore corroborated by ln response ratios for *Br* and PBr_p that showed significant differences between cohorts only in meadow habitats (Figure 4.1). Three populations had higher genetic diversity values for SB than AG cohorts, notably all originated from woodland habitats (RW2, TW2, TW5). Between regions no consistent differences in genetic diversity could be detected, neither for the relation between AG and SB nor at the level of AG or SB alone. Overall, pooling samples (i.e. AG + SB) revealed significantly higher mean values for *PLP* (25%; *t*-test, $p < 0.01$) but not for *Br* (1.17; *t*-test, $p > 0.05$) when compared to separate cohorts.

Genetic structure

Analyses of molecular variance resulted in global ϕ_{ST} values of 0.80 and 0.83 for AG and SB cohorts, respectively, ranging for population pairwise ϕ_{ST} between 0.24 and 0.96 for AG cohorts and between 0.28 and 0.99 for SB cohorts (Supplementary Table S4.3). Hierarchical AMOVA furthermore showed that for both cohorts around 16% of genetic variance resided between regions, whereas most variation (65.8 and 68.7%, respectively) was partitioned among populations within regions (Table 4.3).

Table 4.3 Summary of hierarchical AMOVA results for aboveground and seed bank cohorts of the surveyed populations

Source	Cohort	V	% total	*P*	ϕ Statistics
Among all populations	AG	14.62	80.06	< 0.001	$\phi_{ST} = 0.80$
Among all populations	SB	15.18	83.47	< 0.001	$\phi_{ST} = 0.83$
Among regions	AG	3.09	15.69	< 0.001	$\phi_{CT} = 0.16$
Among populations within regions	AG	12.97	65.83	< 0.001	$\phi_{SC} = 0.78$
Within populations	AG	3.64	18.47	< 0.001	$\phi_{ST} = 0.82$
Among regions	SB	3.16	16.03	< 0.001	$\phi_{CT} = 0.16$
Among populations within regions	SB	13.53	68.71	< 0.001	$\phi_{SC} = 0.82$
Within populations	SB	3.01	15.27	< 0.001	$\phi_{ST} = 0.85$

AG – aboveground individuals; SB – seed bank individuals; V – variance components.

Thus, overall we found very strong genetic differentiation among populations but no significant differences between AG and SB cohorts (Supplementary Table S4.4). Nonetheless, at the individual population level, 2 of 7 meadow population and 4 of 8 woodland populations exhibited significant genetic differentiation between cohorts with pairwise ϕ_{ST} values ranging from 0.05 for TW1 to 0.16 for TW4 (Figure 4.2; Supplementary Table S4.3).

Principal component analysis corroborated the AMOVA results, revealing a very close clustering of cohort pairs for most populations (Figure 4.2). Overall, the first three components accounted for 20.4%, 11.4% and 9.3% of genetic variation. Whereas regions were separated along the first axis, there was no consistent structuring according to habitat types.

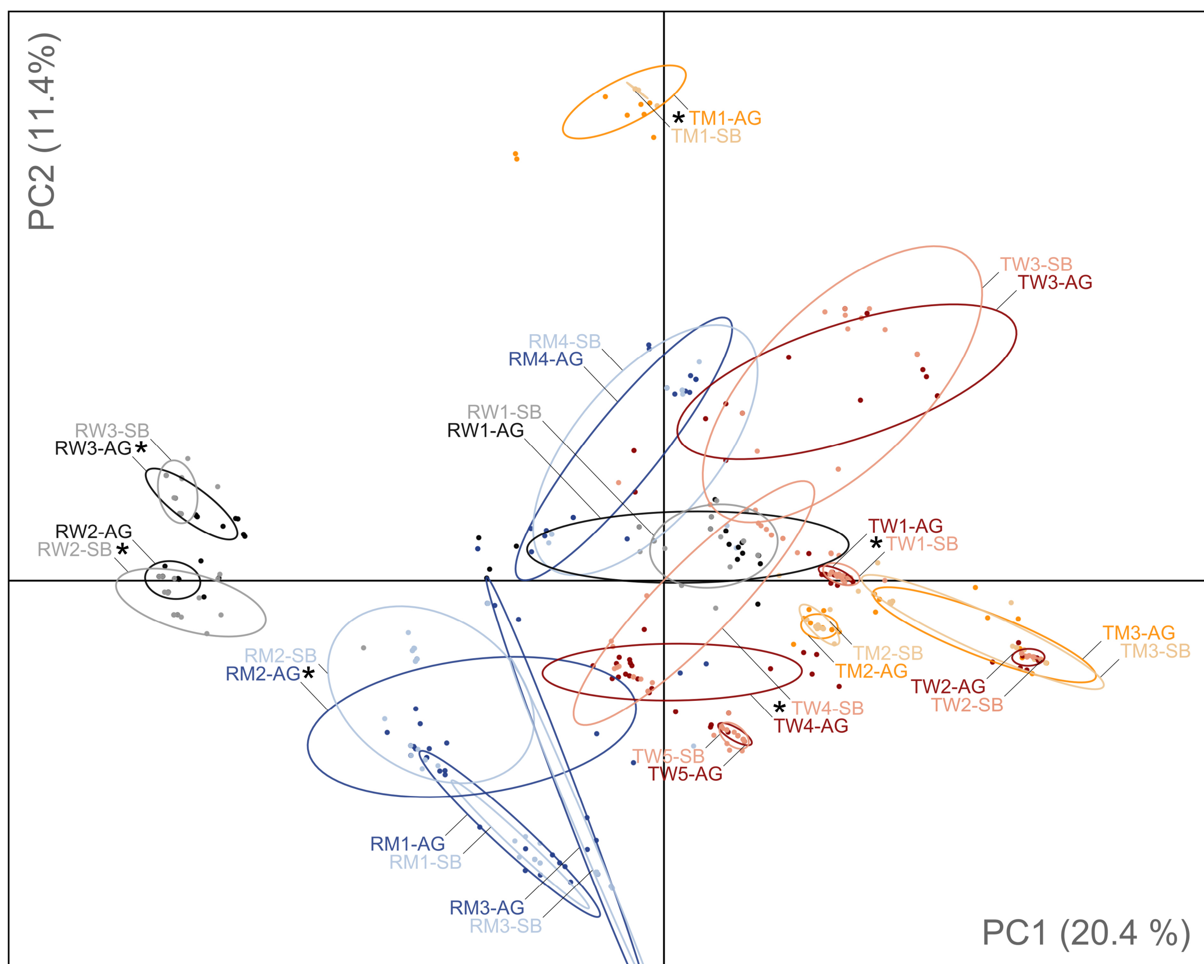


Figure 4.2 Principal component analysis (PCA) of the genetic structure in populations of *Viola elatior*. German populations are indicated by blue (meadow) and grey (woodland), and Czech populations by orange (meadow) and red (woodland). Aboveground and seed bank cohorts are depicted by strong and light colours, respectively. Inertia ellipses indicate dispersion of samples in relation to mean coordinates and include approximately three-fourths (76%) of all individuals for each group. Stars denote populations with significant genetic differentiation (ϕ_{ST}) between aboveground and seed bank cohorts (see also Supplementary Table S4.3).

Small-scale spatial genetic structure

For both, AG and SB samples 5 out of 15 populations showed significant SGS, respectively, with *Sp* values ranging from 0.186 to 2.198 for AG and from 0.073 to 0.253 for SB (Table 4.4). Even though *Sp* values were generally lower for SB samples, the presence of SGS was comparable among AG and SB, and in all but one case, population showing SGS at the AG level also showed SGS at the SB level. Overall, the proportion of populations with significant SGS was higher in meadow than in woodland habitats (AG and SB: 42.9% vs. 25%) and meadow populations showed a stronger degree of SGS with markedly higher average *Sp* values (AG: 0.60 vs. 0.11; SB: 0.32 vs. 0.03). This was also corroborated by the heterogeneity test with pooled data sets, revealing for both cohorts significantly stronger SGS in meadow than in woodland populations, overall and at the regional scale (Figure 4.3; Supplementary Figure S4.1; data only shown for AG). Heterogeneity tests furthermore confirmed that SGS was not significantly different between AG and SB (Figure 4.3). However, for meadow populations kinship coefficients of AG and SB samples showed significant differences in the second distance class.

Table 4.4 Small-scale genetic structure for aboveground and seed bank individuals in populations of *Viola elatior*

Population	n		$F_{(1)}$		b_{log}		*Sp*	
	AG	SB	AG	SB	AG	SB	AG	SB
RM1	21	20	0.000	0.004	-0.004	-0.002	0.004	0.002
RM2	19	7	0.102	0.948	-0.079	-0.085	0.088	1.646
RM3	22	12	0.678**	0.357*	-0.707**	-0.144**	2.198	0.223
RM4	22	18	0.391**	0.425**	-0.428**	-0.062**	0.703	0.108
RW2	23	9	0.329**	0.058	-0.291**	-0.001	0.433	0.001
RW3	20	14	0.016	-0.050	-0.021	0.018	0.021	-0.017
RW4	22	10	0.147	0.358*	-0.159*	-0.085**	0.186	0.133
TM1	21	18	-0.021	-0.019	0.007	0.011	-0.007	-0.011
TM2	21	10	0.041	0.334	-0.134	-0.011	0.140	0.017
TM3	21	20	0.515**	0.548**	-0.487**	-0.114**	1.003	0.253
TW1	23	21	-0.008	-0.059	-0.010	-0.017	0.010	0.016
TW2	23	20	-0.001	-0.109*	-0.042	0.012	0.042	-0.011
TW3	23	19	0.005	0.055	-0.065	-0.013	0.066	0.013
TW4	22	16	0.014	0.265**	-0.008	-0.054*	0.008	0.073
TW5	21	22	0.160*	0.030	-0.067	-0.008	0.079	0.009
Average	*21.6*	*15.7*	*0.158*	*0.210*	*-0.166*	*-0.037*	*0.332*	*0.164*

AG – aboveground individuals; SB – seed bank individuals; n– sample number; $F_{(1)}$ – kinship coefficient of the first distance class; b_{log} – regression slope of spatial genetic autocorrelation; *Sp* – statistic. Significant values are indicated by stars (*$p < 0.05$; **$p < 0.01$).

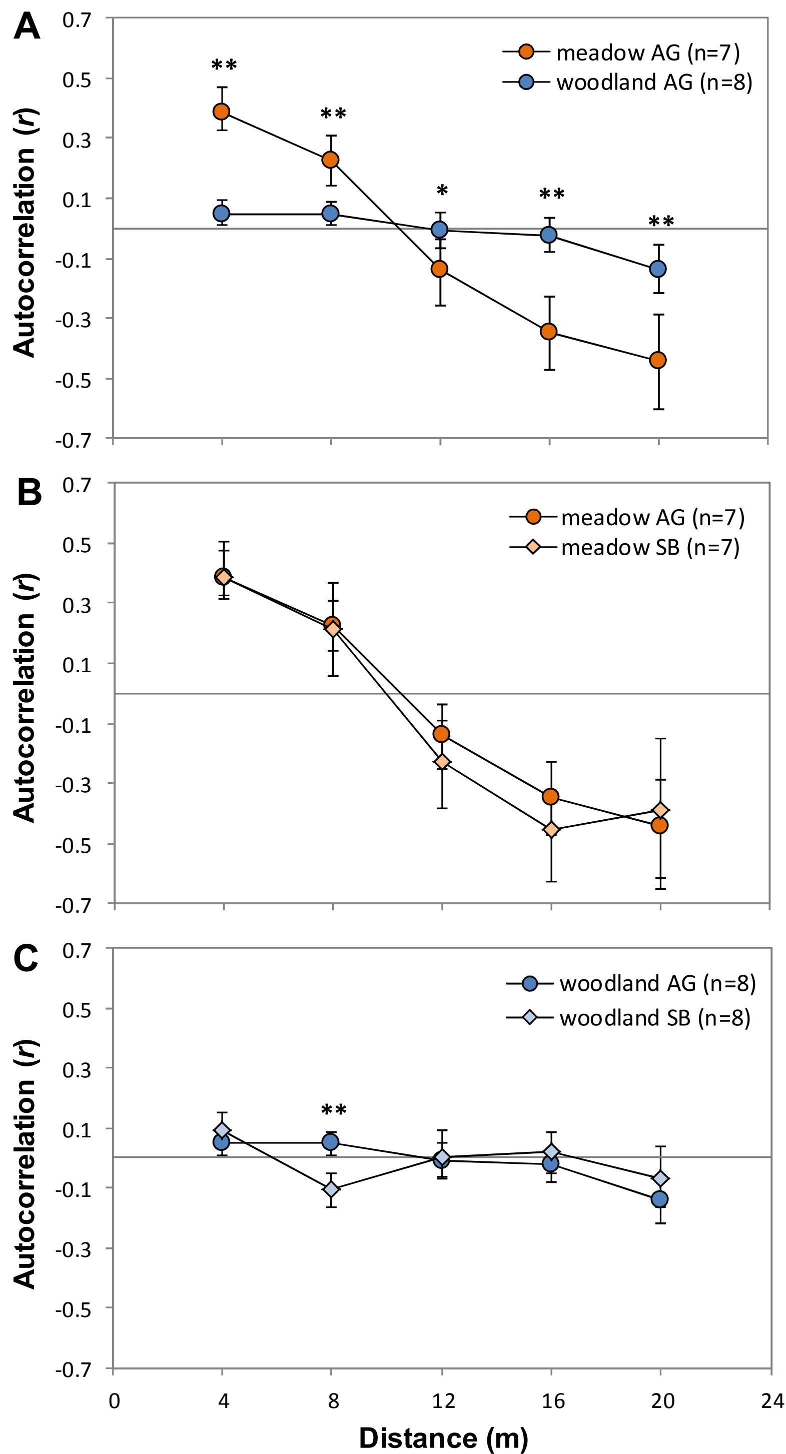


Figure 4.3 Correlograms of spatial genetic autocorrelation in populations of *Viola elatior*. Comparison of correlogram homogeneity is shown for (A) aboveground samples grouped for habitats, and for above-ground and seed bank samples either only from (B) meadow or (C) woodland habitats, respectively; ω-test indicates overall significance (A: $\omega = 73.74$, $p = 0.0001$; B: $\omega = 9.33$, $p = 0.497$; C: $\omega = 19.85$, $p = 0.030$). $^{*}p < 0.05$ and $^{**}p < 0.01$ indicate significant differences for single distance classes. AG – aboveground; SB – seed bank; n – number of populations.

Discussion

Genetic diversity

Both, overall and within AG cohort genetic diversity of *V. elatior* was relatively low compared to other plant species (Nybom 2004; Reisch and Bernhard-Römermann 2014). This is consistent with earlier studies on *V. elatior* (Eckstein *et al.* 2006b; Schulz *et al.* 2014) and with findings in other plants with predominant CL seed production (Sun 1999; Durka *et al.* 2012) that exhibit high levels of inbreeding, little or no genetic variability within populations and strong population differentiation (Culley and Wolfe 2001). Additional factors like spatial isolation and past population bottlenecks may have even reinforced this pattern in the rare *V. elatior*.

Generally, it is widely assumed that decreasing population sizes lead to a loss of genetic variation through effects of increased random genetic drift, higher inbreeding rates and the accumulation of deleterious mutations (e.g. van Treuren *et al.* 1991; Young *et al.* 1996). Furthermore, under changing environmental conditions during succession the loss of genetic variation might be aggravated by an increased probability of local extinction of certain genotypes due to selection (Raffl *et al.* 2006). Consequently, for our study system we initially hypothesized to find a decrease of AG genetic diversity towards woodland habitats, while SB genetic diversity should stay rather constant along the gradient, probably storing genotypes that are lost aboveground and thus enabling to regain former AG genetic diversity under more favorable conditions (e.g. after disturbance).

However, even though we found slightly higher mean values for *Br* and *PLP*5% in meadow populations, overall AG genetic diversity showed no significant differences between the two habitats that differ strongly in population size. This lack of association between population size and genetic variation might be related to multiple causes. First, the time span of reduced individual numbers in the course of succession could have been too short to distinctively affect genetic variation. As genetic drift will be stronger as more generations have passed, long lived, perennial species like *V. elatior* might suffer less from the negative genetic consequences of reduced population sizes (Leimu *et al.* 2006). Second, in self-compatible species with a long history of inbreeding the genetic load might have been purged, making them less susceptible to genetic erosion (Busch 2005). Indeed, Leimu *et al.* (2006) showed that the positive relationship between population size and genetic variation is stronger in self-incompatible than in self-compatible species. Third, the rarity of *V. elatior* might be an additional cause, as rare species typically are considered to exhibit comparatively low genetic variation, independently of their respective population size (Leimu *et al.* 2006). Finally, also a higher relative outcrossing rate in woodland populations and thus a stronger buffer capacity of the seed bank potentially could have balanced the effects of decreasing population size (but see below).

Similar to AG cohorts also SB cohorts showed no significant differences in genetic diversity along the gradient. However, when examining the relationship between AG and SB, we found clear differences between the two habitat types. Whereas meadow populations overall exhibited higher AG than SB diversity – i.e. higher *Br* and PBr_p values and

lower number of shared AFLP phenotypes in AG cohorts – no significant differences could be detected in woodland populations. Thus, strikingly relative SB genetic diversity (i.e. compared to AG cohorts) seems to increase with ongoing succession and despite decreasing population size, resulting for three of eight woodland populations even in higher SB than AG genetic diversity.

Generally, higher AG than SB genetic diversity, as found in meadow populations, might be explained by selection against homozygotes or inbred individuals during germination and recruitment as suggested by earlier seed bank genetic studies (Vitalis *et al.* 2004; Mandák *et al.* 2006; Honnay *et al.* 2008). However, as we did not find reduced seed bank genetic diversity in late successional stages, in which similar or even stronger selection pressure can be expected, we alternatively suggest that in woodland an increase in the relative contribution of outcrossed seeds to the seed bank may be causal for the observed pattern, counteracting effects of selection and reduced population size.

Several non-mutually exclusive and linked processes may have fostered outcrossing in woodland habitats. First, a change in the balance between CH and CL capsules along the gradient might have led to higher CH/CL capsule ratios in woodland habitats. It is well known that in cleistogamous species the allocation to CH and CL capsules may depend on environmental factors (Culley and Klooster 2007) and indeed some authors found increasing CH/CL ratios with decreasing plant density (Cheplick 2007), decreasing light availability (Mattila and Salonen 1995; Cheplick 2007) and increasing soil water availability (Brown 1952). Hence, the same environmental gradients that are present between meadow and woodland habitats. Although some studies (e.g. Le Corff 1993) indicate that the relationship between environmental parameters and CH/CL ratio is species specific, the only study in the genus *Viola* we are aware of, found that the CH/CL ratio increased with shading (Mattila and Salonen 1995). Second, besides changes in relative CH and CL capsule production, also variation in CH seed abortion rates or seed quality might be causal. As CH capsules mature in the hottest period of the year (i.e. June–July) differences in water availability between meadow and woodland habitats are particularly distinct during CH seed development. Thus, CH seed production and quality could increase along the successional gradient and hence the relative contribution of outcrossed seeds to the seed bank. This is corroborated by a survey in 4 populations from the Upper Rhine region (Schulz, unpublished data) revealing for CH seeds considerably higher abortion rates (30% vs. 6%) and lower seed mass (10 mg vs. 18 mg) in meadow than in woodland sites, respectively. In contrast, differences for CL seeds in the same populations were less pronounced (seed abortion: 4.8% vs. 2.2%; seed mass: 15 mg vs. 17 mg). Similarly, also in *Viola blanda* shading increased average seed mass (Griffith Jr 1998). Third, seed longevity may change along the successional gradient. Given that lower seed mass is associated with lower seed bank survival (van Groenendael 1985; Thompson *et al.* 1993) the persistence of CH seeds could be reduced in meadow populations. Moreover, it might be possible that seed longevity generally is higher in woodland habitats, as soil parameters like moisture and temperature are more constant and not as extreme as in open grassland. Accordingly, woodland seed banks potentially could be assembled from more seed generations, leading to higher relative genetic diversity. Finally, also anthropogenic effects could have

had an impact on CH seed contribution, as most of the floodplain meadows are regularly managed by a one-time mowing in early June. Hence large amounts of ripening CH capsules might be destroyed every year, whereas CL capsules freely develop from July to October. For grasses it was even suggested that mowing and grazing pressure generally can increase the share of CL spikelets (but see Cheplick 2007 and reference therein).

Comparable to results of Eckstein *et al.* (2006b) we found no significant differences in genetic diversity between German and Czech populations, neither for AG nor SB cohorts. This strongly implies that the detected habitat related differences seem to present a general effect that is independent of geographic location.

Genetic structure

Genetic differentiation in *V. elatior* was very high, with 80.1% and 83.5% of genetic variation residing among populations for AG and SB cohorts, respectively. Similar differentiation (up to 82%) has been also reported in earlier studies on this species (Gygax 2001; Eckstein *et al.* 2006b; Schulz *et al.* 2014), reflecting the predominant selfing and relatively small population sizes and hence a strong influence of genetic drift on population structure. This is further substantiated by a lack of correlation between genetic differentiation and geographic distances in populations from the Upper Rhine Valley (Schulz *et al.* 2014), which indicates that spatial isolation and thus gene flow do not play a major role for population divergence. Overall, population differentiation in *V. elatior* was virtually identical among SB and AG cohorts at all hierarchical levels (Table 4.3). In contrast, some other studies reported lower differentiation among SB cohorts than among AG cohorts (Tonsor *et al.* 1993; McCue and Holtsford 1998; Zaghloul *et al.* 2013), indicating that by chance or selection, aboveground populations can become more differentiated than the potentially multigenerational and thus more homogeneous seed pools they derived from (McCue and Holtsford 1998). Accordingly, in *V. elatior* habitat related change in selection might have been not as high as expected. However, at least the fact that some individual populations exhibited significant differentiation between cohorts suggests that differential seed bank recruitment over time, and hence changes in post-germination selection, do also exist in *V. elatior*. Interestingly, populations with significant cohort differentiation were obviously more frequent in woodland than in meadow habitats (4/8 vs. 2/7, respectively) indicating that the genetic buffering capacity of the seed bank increases towards late successional stages.

Small-scale spatial genetic structure

Spatial genetic structure within populations arises due to spatially restricted gene dispersal and is mainly related to the amount of gene flow by seeds and pollen (Zeng *et al.* 2011). Therefore, the observed significant decrease of SGS from meadow to woodland habitats strongly supports the assumption that in *V. elatior* the relative impact of outcrossing and thus pollen dispersal increases towards late successional stages. Very high selfing rates in meadow populations are furthermore corroborated by high *Sp* values (mean $Sp = 0.59$ and 0.32, for AG and SB, respectively) that even exceeded mean values reported for other

predominantly selfing species (*Sp* = 0.14, Vekemans and Hardy. 2004). In contrast, *Sp* values in woodland populations (mean *Sp* = 0.11 and 0.3, for AG and SB, respectively) were markedly lower and tended to reflect more those for mixed mating species (*Sp* = 0.04, Vekemans and Hardy 2004). Besides outcrossing rates, also the density of co-occurring individuals is a major driver of SGS and under isolation by distance, *Sp* statistics are expected to be inversely proportional to plant density (Heywood 1991; Vekemans and Hardy 2004). In *V. elatior* however, the effect of lower population densities in woodland habitats might have been compensated by larger gene dispersal distances. Indeed, direct measurements in other plant species showed that pollen dispersal distances increased with decreasing plant densities (Schmitt 1983; Fenster 1991). Moreover, if in meadow populations almost all seeds result from self-fertilization, only seed dispersal (~ 1.3 m, Eckstein *et al.* 2006a) does impact gene dispersal distances. In contrast, in woodland populations with considerable outcrossing, both seed and pollen dispersal would contribute and consequently significantly increase overall gene dispersal (Vekemans and Hardy 2004). But still, also other succession related changes may have decreased SGS in woodland habitats. More scattered patterns of light availability and microclimatic conditions could have reduced overall reproductive success and thus the clustering of siblings (Born *et al.* 2008), and a reduced spatial aggregation of individuals could have increased the overlap of seed shadows (Hamrick and Nason 1996; Zeng *et al.* 2011). Moreover, as *V. elatior* is a diplochorous species (i.e. ballistic and ant dispersal) also differences in secondary dispersal by ants may have contributed. However, in four diplochorous violet species from North-America dispersal by ants was the least important factor determining neighborhood size, whereas pollen dispersal played the major role (Beattie and Culver 1979).

Both, AG and SB cohorts showed significant SGS and even though *Sp* values were generally lower in SB cohorts, overall there were no significant differences between the two groups. Notwithstanding, lower spatial autocorrelation of SB cohorts in some of the populations might indicate that at small spatial scales, seed germination and establishment could be under stronger selection than seed persistence in the soil (Tonsor *et al.* 1993). Similar to our results, most of the few other available studies on that topic also detected significant SGS in both cohorts (Cabin *et al.* 1998; Shimono *et al.* 2006; Ottewell *et al.* 2011) suggesting that the spatial structure of AG and SB individuals is often mutually dependent (Shimono *et al.* 2006; but see Mandak *et al.* 2006).

Taken together, the overall SGS patterning in *V. elatior* seems to reflect the different outcrossing rates along the environmental gradient. However, SGS appears to break down relatively fast with ongoing succession and is not stored over longer periods of time in SB cohorts. Otherwise much stronger spatial autocorrelation would have been present in SB cohorts of woodland populations. Instead, the high correlation of AG and SB cohorts in both habitats suggests that the longevity of soil seeds and adult plants in *V. elatior* largely overlaps and that seeds do not persist in the seed bank for more than few adult generations (Tonsor *et al.* 1993).

Conclusions

To our knowledge, the present study is the first one that has compared AG and SB genetic diversity along a successional gradient. Surveying a relative large number of populations of the cleistogamous *V. elatior* in two different regions, we could show that the contribution of outcrossing to reproduction seems to increase from early to late successional stages, leading to higher relative seed bank diversity and lower within population SGS. This suggests that under favorable early successional conditions with high plant densities, populations maintain their approved genotypes mainly by selfing. Contrary, under more unfavorable conditions an increase in outcrossing may keep genetic diversity in *V. elatior* at a constant level and hence seems to compensate the detrimental effects of small population size. Ultimately, in late successional habitats with a higher risk of extinction, the resulting increased relative SB genetic diversity also assures a higher chance for population recovery and thus the long term persistence of the species. Moreover, the finding that pooled AG and SB samples overall had higher mean *PLP* values than individual cohorts, generally supports the stabilizing effect of the seed bank.

To close the cycle and to come back to the implications of Honnay *et al.* (2008), we might conclude from our results that the relationship between seed bank and aboveground plants potentially can be driven by both, post-germination selection as seen in meadow habitats and genetic buffering through stored seeds that seems to counteract the effects of drift and selection as seen in woodland populations. However, to substantiate these results and to test if this is not a specific situation in cleistogamous plants, further studies in species with other mating systems are needed that survey seed bank genetics along environmental gradients.

Acknowledgements

We want to thank Josef Scholz vom Hofe and Lena Föhr for help during field work, and Martina Herrmann and Ina Geier for assistance during laboratory work. Financial support was obtained from the Deutsche Forschungsgemeinschaft (DFG, grant EC209/7-1). The contribution by J. Danihelka was supported as a long-term research development project no. RVO 67985939.

References

Alvarez-Buylla ER, Chaos A, Pinero D, Garay AA (1996) Demographic genetics of a pioneer tropical tree species: patch dynamics, seed dispersal, and seed banks. *Evolution*, 50, 1155–1166.

Banks SC, Peakall R (2012) Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Molecular Ecology*, 21, 2092–2105.

Baskin CC, Baskin JM (2001) *Seeds: ecology, biogeography, and evolution of dormancy and germination*. Academic Press, San Diego, USA.

Beattie AJ (1976) Plant dispersion, pollination and gene flow in *Viola*. *Oecologia*, 25, 291–300.

Beattie AJ, Culver DC (1979) Neighborhood size in *Viola*. *Evolution*, 33, 1226–1229.

Born C, Hardy OJ, Chevallier M *et al.* (2008) Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology*, 17, 2041–2050.

Busch J (2005) Inbreeding depression in self-incompatible and self-compatible populations of *Leavenworthia alabamica*. *Heredity*, 94, 159–165.

Cabin RJ, Mitchell RJ, Marshall DL (1998) Do surface plant and soil seed bank populations differ genetically? A multipopulation study of the desert mustard *Lesquerella fendleri* (Brassicaceae). *American Journal of Botany*, 85, 1098–1109.

Cheplick GP (2007) Plasticity of chasmogamous and cleistogamous reproductive allocation in grasses. *Aliso: A Journal of Systematic and Evolutionary Botany*, 23, 286–294.

Coart E, Glabeke SV, Petit RJ, Bockstaele EV, Roldán-Ruiz I (2005) Range wide versus local patterns of genetic diversity in hornbeam (*Carpinus betulus* L.). *Conservation Genetics*, 6, 259–273.

Le Corff J (1993) Effects of light and nutrient availability on chasmogamy and cleistogamy in an understory tropical herb, *Calathea micans* (Marantaceae). *American Journal of Botany*, 80, 1392–1399.

Culley TM, Klooster MR (2007) The cleistogamous breeding system: a review of its frequency, evolution, and ecology in angiosperms. *The Botanical Review*, 73, 1–30.

Culley TM, Wolfe AD (2001) Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. *Heredity*, 86, 545–556.

Danihelka J, Niklfeld H, Šípošová H (2009) *Viola elatior*, *V. pumila* and *V. stagnina* in Austria, Czechia and Slovakia: a story of decline. *Preslia*, 81, 151–171.

Durka W, Nossol C, Welk E *et al.* (2012) Extreme genetic depauperation and differentiation of both populations and species in Eurasian feather grasses (Stipa). *Plant Systematics and Evolution*, 299, 259–269.

Eckstein RL, Hölzel N, Danihelka J (2006a) Biological Flora of Central Europe: *Viola elatior*, *V. pumila* and *V. stagnina*. *Perspectives in Plant Ecology, Evolution and Systematics*, 8, 45–66.

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

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.

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.

Falahati-Anbaran M, Lundemo S, Agren J, Stenoien HK (2011) Genetic consequences of seed banks in the perennial herb *Arabidopsis lyrata* subsp. *petraea* (Brassicaceae). *American Journal of Botany*, 98, 1475–1485.

Fenster CB (1991) Gene flow in *Chamaecrista fasciculata* (Leguminosae) I. Gene dispersal. *Evolution*, 45, 398–409.

Goldberg D, Scheiner S (1993) ANOVA and ANCOVA: field competition experiments. In: *Design and analysis of ecological experiments* (eds Scheiner S, Gurevitch J), pp. 77–98. Oxford University Press Inc., Oxford.

Griffith Jr C (1996) Distribution of *Viola blanda* in relation to within-habitat variation in canopy openness, soil phosphorus, and magnesium. *Bulletin of the Torrey Botanical Club*, 123, 281–285.

Van Groenendael J (1985) Differences in life histories between two ecotypes of *Plantago lanceolata*. In: *Studies on plant demography* (ed White J), pp. 51–67. Academic Press, London.

Gygax A (2001) Populationsentwicklung, Ansiedlungsversuche und genetische Variabilität von *Viola elatior* Fr. in der Schweiz. Université de Neuchâtel, Switzerland.

Hanin N, Quaye M, Westberg E, Barazani O (2013) Soil seed bank and among-years genetic diversity in arid populations of *Eruca sativa* Miller (Brassicaceae). *Journal of Arid Environments*, 91, 151–154.

Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology*, 12, 1577–1588.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618–620.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, 22, 335–355.

Hölzel N, Otte A (2004) Assessing soil seed bank persistence in flood-meadows: The search for reliable traits. *Journal of Vegetation Science*, 15, 93–100.

Honnay O, Bossuyt B, Jacquemyn H, Shimono A, Uchiyama K (2008) Can a seed bank maintain the genetic variation in the above ground plant population? *Oikos*, 117, 1–5.

Honnay O, Jacquemyn H, Bossuyt B, Hermy M (2005) Forest fragmentation effects on patch occupancy and population viability of herbaceous plant species: Tansley review. *New Phytologist*, 166, 723–736.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics*, 5, 539–543.

Kloss L, Fischer M, Durka W (2011) Land-use effects on genetic structure of a common grassland herb: A matter of scale. *Basic and Applied Ecology*, 12, 440–448.

Koch M, Huthmann M, Bernhardt KG (2003) *Cardamine amara* L. (Brassicaceae) in dynamic habitats: genetic composition and diversity of seed bank and established populations. *Basic and Applied Ecology*, 4, 339–348.

Leck MA, Parker VT, Simpson RL (Eds.) (1989) *Ecology of soil seed banks*. Academic Press, San Diego, USA.

Leimu R, Mutikainen P, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology*, 94, 942–952.

Lesica P, Allendorf FW (1992) Are small populations of plants worth preserving? *Conservation Biology*, 6, 135–139.

Lundemo S, Falahati-Anbaran M, Stenoien HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Molecular Ecology*, 18, 2798–2811.

Mandák B, Bímová K, Mahelka V, Plačková I (2006) How much genetic variation is stored in the seed bank? A study of *Atriplex tatarica* (Chenopodiaceae). *Molecular Ecology*, 15, 2653–2663.

Mandák B, Zákravský P, Mahelka V, Plačková I (2012) Can soil seed banks serve as genetic memory? A study of three species with contrasting life history strategies. *PLoS ONE*, 7, e49471.

Mattila T, Salonen V (1995) Reproduction of *Viola mirabilis* in relation to light and nutrient availability. *Canadian Journal of Botany*, 73, 1917–1924.

McCue KA, Holtsford TP (1998) Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). *American Journal of Botany*, 85, 30–36.

Meusel H, Jäger E, Rauschert S, Weinert E (1978) *Vergleichende Chorologie der zentraleuropäischen Flora. Vol. 2.* Gustav Fischer Verlag, Jena, Germany.

Moora M (2003) Responses of a rare (*Viola elatior*) and a common (*Viola mirabilis*) congeneric species to different management conditions in grassland—is different light competition ability responsible for different abundances? *Acta Oecologica*, 24, 169–174.

Morris AB, Baucom RS, Cruzan MB (2002) Stratified analysis of the soil seed bank in the cedar glade endemic *Astragalus bibullatus*: evidence for historical changes in genetic structure. *American Journal of Botany*, 89, 29–36.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular ecology*, 13, 1143–1155.

Ottewell KM, Bickerton D, Lowe AJ (2011) Can a seed bank provide demographic and genetic rescue in a declining population of the endangered shrub *Acacia pinguifolia*? *Conservation Genetics*, 12, 669–678.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537–2539.

Raffl C, Schönswetter P, Erschbamer B (2006) "Sax-sess"—genetics of primary succession in a pioneer species on two parallel glacier forelands. *Molecular ecology*, 15, 2433–2440.

Reisch C, Bernhardt-Römermann M (2014) The impact of study design and life history traits on genetic variation of plants determined with AFLPs. *Plant Ecology*, 215, 1493–1511.

Schmitt J, Stinchcombe JR, Heschel MS, Huber H (2003) The adaptive evolution of plasticity: phytochrome-mediated shade avoidance responses. *Integrative and Comparative Biology*, 43, 459–469.

Schulz B, Eckstein RL, Durka W (2014) Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Molecular Ecology*, 23, 3523–3537.

Shimono A, Ueno S, Tsumura Y, Washitani I (2006) Spatial genetic structure links between soil seed banks and above-ground populations of *Primula modesta* in subalpine grassland. *Journal of Ecology*, 94, 77–86.

Smouse PE, Peakall R, Gonzales E (2008) A heterogeneity test for fine-scale genetic structure. *Molecular Ecology*, 17, 3389–3400.

Sun M (1999) Cleistogamy in *Scutellaria indica* (Labiatae): effective mating system and population genetic structure. *Molecular Ecology*, 8, 1285–1295.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, 24, 2498–2504.

Templeton AR, Levin DA (1979) Evolutionary consequences of seed pools. *American Naturalist*, 114, 232–249.

Ter Heerdt G, Verweij G, Bekker R, Bakker J (1996) An improved method for seed-bank analysis: seedling emergence after removing the soil by sieving. *Functional Ecology*, 10, 144–151.

Thompson K, Band S, Hodgson J (1993) Seed size and shape predict persistence in soil. *Functional Ecology*, 7, 236–241.

Tonsor SJ, Kalisz S, Fisher J, Holtsford TP (1993) A life-history based study of population genetic structure: seed bank to adults in *Plantago lanceolata*. *Evolution*, 47, 833–843.

Van Treuren R, Bijlsma R, Van Delden W, Ouborg N (1991) The significance of genetic erosion in the process of extinction. I. Genetic differentiation in *Salvia pratensis* and *Scabiosa columbaria* in relation to population size. *Heredity*, 66, 181–189.

Uchiyama K, Goto S, Tsuda Y, Takahashi Y, Ide Y (2006) Genetic diversity and genetic structure of adult and buried seed populations of *Betula maximowicziana* in mixed and post-fire stands. *Forest Ecology and Management*, 237, 119–126.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, 13, 921–935.

Vitalis R, Glémin S, Olivieri I (2004) When genes go to sleep: the population genetic consequences of seed dormancy and monocarpic perenniality. *The American Naturalist*, 163, 295–311.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution*, 11, 413–418.

Zaghloul M, Reisch C, Poschlod P (2013) Soil seed bank contributes significantly to genetic variation of *Hypericum sinaicum* in a changing environment. *Plant Systematics and Evolution*, 299, 1819–1828.

Zeng X, Michalski SG, Fischer M, Durka W (2011) Species diversity and population density affect genetic structure and gene dispersal in a subtropical understory shrub. *Journal of Plant Ecology*, 5, 270–278.

Appendix

Appendix 4.1 AFLP protocol

For restriction and ligation (RL) 5.6 µl genomic DNA were combined with 5.4 µl RL reaction mix containing 0.55 µl BSA (1 mg/ml; New England Biolabs, NEB), 1.1 µl 0.5 M NaCl, 5 u *Eco*RI (NEB), 1 u *Mse*I (NEB), 67 u T4 DNA ligase (NEB), 1.1 µl T4 DNA ligase buffer (NEB), 1 µl *Eco*RI adapter (5 pmol) and 1 µl *Mse*I adapter (50 pmol). The reaction was incubated for 2 h at 37 °C and diluted 1:2. For the preselective amplification (PCR1), 4 µl RL product were combined with 16 µl PCR1 reaction mix containing 1.5 ng/µl *Eco*RI and *Mse*I preselective primers each, 200 µM dNTPs (Roth), 2 µl 10 x Dream Tag buffer (QIAGEN), 0.8 u Dream Tag polymerase (QIAGEN) and 9.84 µl H_2O. The thermocycler protocol was 72.0 °C (2 min) followed by 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min) and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler gradient. The PCR1 product was diluted 1:5. For the selective amplification (PCR2), 1 µl PCR1 product was combined with 3.4 µl PCR2 reaction mix containing 2.2 µl Multiplex PCR kit (QIAGEN) and 0.6 µl fluorescent labeled *Eco*RI primer (1 pmol/µl) and 0.6 µl *Mse*I (5 pmol/µl) selective primers each. The thermocycler protocol was 94.0 °C (2 min) followed by 10 cycles of 94.0 °C (20 s), 66.0 °C (30 s, decreasing 1 °C per cycle) and 72.0 °C (2 min) and 20 cycles of 94.0 °C (20 s), 56.0 °C (30 s) and 72.0 °C (2 min), and a final extension at 60.0 °C (30 min), performed on an Eppendorf Mastercycler pro 384.

Supplementary Table S4.1 Adaptor- and primer sequences used for AFLP analyses

Primer	Sequence
Adaptors	
*Eco*RI-*adapter* top	5'-CTCGTAGACTGCGTACC-3'
*Eco*RI-*adapter* bottom	5'-AATTGGTACGCAGTCTAC-3'
*Mse*I-*adapter* top	5'-GAGCGATGAGTCCTGAG-3'
*Mse*I-*adapter* bottom	3'-TACTCAGGACTCAT-5'
Preselective primers	
*Eco*RI + A	5'-GACTGCGTACCAATTCA-3'
*Mse*I + C	5'-GATGAGTCCTGAGTAAC-3'
Selective primer	
*Eco*RI + AAC-FAM[1]	5'-GACTGCGTACCAATTCAAC-3'
*Eco*RI + ACT-FAM[2]	5'-GACTGCGTACCAATTCACT-3'
*Eco*RI + ACA-VIC[3,4]	5'-GACTGCGTACCAATTCACA-3'
*Eco*RI + AAG-NED[5,6]	5'-GACTGCGTACCAATTCAAG-3'
*Eco*RI + AGC-PET[7]	5'-GACTGCGTACCAATTCAGC-3'
*Eco*RI + AGG-PET[8]	5'-GACTGCGTACCAATTCAGG-3'
*Mse*I + CTA[1]	5'-GATGAGTCCTGAGTAACTA-3'
*Mse*I + CAA[2,3,7]	5'-GATGAGTCCTGAGTAACAA-3'
*Mse*I + CAC[5]	5'-GATGAGTCCTGAGTAACAC-3'
*Mse*I + CTC[4,6,8]	5'-GATGAGTCCTGAGTAACTC-3'

Superscript numbers indicate primer combinations used for the selective amplification.

Supplementary Table S4.2 Seedling emergence rates of *Viola elatior* in soil samples from the surveyed populations

Population	Total seedlings	Seedlings per m^2
RM1	20	66.7
RM2	26	86.7
RM3	12	40.0
RM4	18	60.0
RW1	26	86.7
RW2	44	146.7
RW3	28	93.3
TM1	61	203.3
TM2	116	386.7
TM3	63	210.0
TW1	156	520.0
TW2	206	686.7
TW3	72	240.0
TW4	85	283.3
TW5	79	263.3

Supplementary Table S4.3 Pairwise ϕ_{ST}-Matrix for aboveground and seed bank cohorts of the surveyed populations

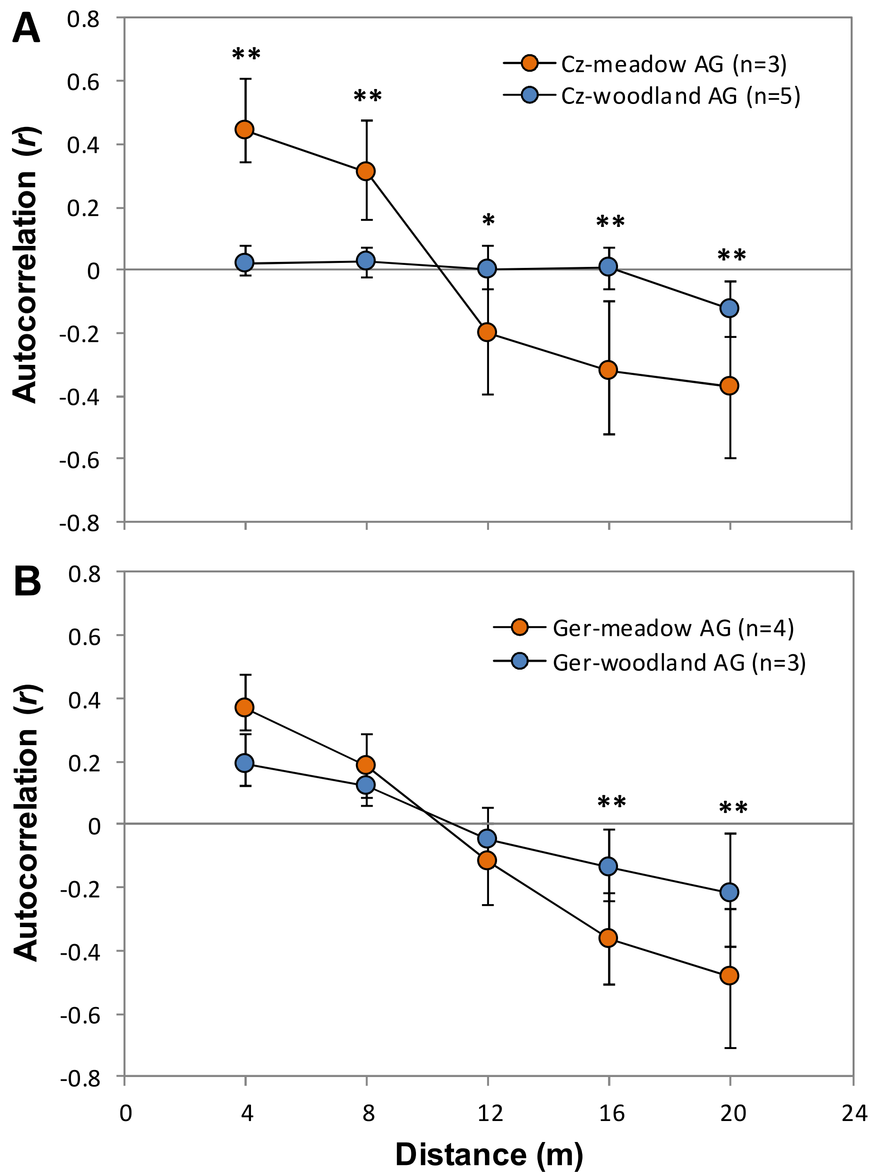
	RM 1 AG	RM 1 SB	RM 2 AG	RM 2 SB	RM 3 AG	RM 3 SB	RM 4 AG	RM 4 SB	RW 1 AG	RW 1 SB	RW 2 AG	RW 2 SB	RW 3 AG	RW 3 SB	TM 1 AG	TM 1 SB	TM 2 AG	TM 2 SB	TM 3 AG	TM 3 SB	TW 1 AG	TW 1 SB	TW 2 AG	TW 2 SB	TW 3 AG	TW 3 SB	TW 4 AG	TW 4 SB	TW 5 AG	TW 5 SB
RM1 AG	-	**0.17**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM1 SB	**0.03**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM2 AG	0.62	0.67	-	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM2 SB	0.67	0.71	**0.11**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM3 AG	0.53	0.56	0.51	0.56	-	**0.43**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM3 SB	0.63	0.69	0.58	0.64	**-0.03**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM4 AG	0.82	0.85	0.68	0.70	0.70	0.77	-	**0.32**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RM4 SB	0.83	0.87	0.69	0.72	0.70	0.78	**0.00**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW1 AG	0.79	0.83	0.65	0.69	0.61	0.72	0.75	0.76	-	**0.23**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW1 SB	0.81	0.85	0.68	0.71	0.62	0.73	0.77	0.78	**0.01**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW2 AG	0.88	0.92	0.74	0.72	0.74	0.85	0.87	0.89	0.88	0.89	-	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW2 SB	0.85	0.90	0.70	0.69	0.72	0.82	0.85	0.87	0.85	0.87	**0.16**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW3 AG	0.86	0.91	0.74	0.74	0.73	0.83	0.85	0.87	0.85	0.87	0.71	0.64	-	**0.01**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW3 SB	0.88	0.92	0.75	0.75	0.74	0.85	0.86	0.88	0.87	0.88	0.74	0.67	**0.09**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM1 AG	0.90	0.93	0.81	0.82	0.78	0.87	0.86	0.87	0.86	0.87	0.94	0.92	0.92	0.93	-	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM1 SB	0.93	0.96	0.85	0.86	0.82	0.91	0.89	0.91	0.90	0.91	0.97	0.95	0.96	0.97	**0.06**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM2 AG	0.89	0.93	0.82	0.83	0.77	0.86	0.86	0.87	0.87	0.88	0.95	0.92	0.93	0.95	0.94	0.97	-	**0.75**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM2 SB	0.90	0.94	0.83	0.84	0.79	0.87	0.87	0.88	0.88	0.89	0.95	0.93	0.94	0.95	0.95	0.98	**-0.02**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM3 AG	0.73	0.76	0.63	0.66	0.60	0.65	0.70	0.70	0.64	0.65	0.83	0.81	0.83	0.83	0.79	0.83	0.72	0.74	-	**0.81**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TM3 SB	0.73	0.75	0.62	0.65	0.60	0.65	0.69	0.69	0.63	0.63	0.83	0.80	0.82	0.82	0.78	0.82	0.71	0.73	**-0.04**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TW1 AG	0.88	0.92	0.80	0.83	0.77	0.85	0.86	0.87	0.86	0.87	0.95	0.93	0.93	0.94	0.93	0.96	0.93	0.93	0.75	0.74	-	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TW1 SB	0.89	0.92	0.80	0.83	0.77	0.86	0.86	0.88	0.86	0.88	0.95	0.93	0.93	0.94	0.93	0.96	0.93	0.94	0.75	0.73	**0.05**	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TW2 AG	0.90	0.94	0.81	0.83	0.75	0.85	0.87	0.89	0.84	0.86	0.96	0.95	0.95	0.96	0.95	0.98	0.93	0.94	0.24	0.26	0.93	0.93	-	**0.67**	0.00	0.00	0.00	0.00	0.00	0.00
TW2 SB	0.91	0.95	0.82	0.84	0.76	0.87	0.88	0.90	0.86	0.87	0.97	0.95	0.96	0.97	0.96	0.99	0.95	0.95	0.25	0.28	0.94	0.94	**-0.01**	-	0.00	0.00	0.00	0.00	0.00	0.00
TW3 AG	0.75	0.78	0.68	0.70	0.62	0.68	0.69	0.69	0.66	0.66	0.83	0.81	0.80	0.81	0.73	0.77	0.76	0.77	0.50	0.49	0.77	0.77	0.70	0.71	-	**0.27**	0.00	0.00	0.00	0.00
TW3 SB	0.73	0.76	0.66	0.67	0.60	0.66	0.67	0.67	0.63	0.63	0.80	0.79	0.78	0.78	0.70	0.74	0.72	0.73	0.46	0.46	0.73	0.72	0.67	0.68	**0.01**	-	0.00	0.00	0.00	0.00
TW4 AG	0.73	0.76	0.58	0.60	0.61	0.67	0.69	0.70	0.72	0.73	0.80	0.77	0.80	0.80	0.82	0.85	0.78	0.79	0.59	0.58	0.77	0.77	0.77	0.78	0.64	0.62	-	**0.00**	0.00	0.00
TW4 SB	0.81	0.85	0.67	0.69	0.69	0.76	0.75	0.76	0.78	0.80	0.87	0.84	0.86	0.87	0.86	0.89	0.84	0.85	0.66	0.65	0.85	0.62	0.84	0.86	0.71	0.68	**0.16**	-	0.00	0.00
TW5 AG	0.89	0.93	0.78	0.80	0.74	0.84	0.89	0.90	0.86	0.88	0.95	0.93	0.94	0.95	0.95	0.98	0.94	0.94	0.72	0.71	0.94	0.94	0.95	0.96	0.78	0.75	0.77	0.85	-	**0.79**
TW5 SB	0.88	0.92	0.78	0.80	0.74	0.84	0.88	0.90	0.86	0.87	0.95	0.92	0.93	0.95	0.94	0.97	0.93	0.94	0.72	0.71	0.93	0.93	0.94	0.95	0.78	0.75	0.77	0.84	**-0.02**	-

Lower and upper diagonal represent pairwise ϕ_{ST} and corresponding *p* values after 9999 permutations, respectively; light and dark gray cells depict non-significant and significant ϕ_{ST} *p* values for population pairs of above-ground (AG) and seed bank (SB) individuals, respectively.

Supplementary Table S4.4 Summary of hierarchical AMOVA results for German and Czech populations grouped by cohort type

Source	region	V	% total	*P*	ϕ statistics
Among cohorts	Ger	-1.80	-12.43	0.994	ϕ_{CT} = -0.12
Among populations within cohorts	Ger	12.60	87.12	< 0.001	ϕ_{SC} = 0.77
Within populations	Ger	3.66	25.30	< 0.001	ϕ_{ST} = 0.75
Among cohorts	Cz	-1.70	-11.22	0.982	ϕ_{CT} = -0.11
Among populations within cohorts	Cz	13.74	90.9	< 0.001	ϕ_{SC} = 0.82
Within populations	Cz	3.07	20.32	< 0.001	ϕ_{ST} = 0.80

Ger – Germany; Cz – Czech Republic; V – variance components.



Supplementary Figure S4.1 Correlograms of genetic autocorrelation in populations of *Viola elatior*. Comparison of correlogram homogeneity is shown for (A) Czech and (B) German aboveground samples grouped for habitats, respectively; ω-test indicates overall significance (A: $\omega = 67.18$, $p = 0.0001$; B: $\omega = 33.91$, $p = 0.0003$). *$p < 0.05$ and **$p < 0.01$ indicate significant differences for single distance classes. Cz – Czech Republic; Ger – Germany; AG – aboveground, n – number of populations.

Summary

The ability to adjust to varying biotic and abiotic conditions is a key feature of plants to survive in an ever changing environment. In the course of evolution plants therefore have developed numerous mechanisms to cope with environmental heterogeneity and its unpredictability. This thesis investigates two of these mechanisms and their population ecological consequences in the cleistogamous species *Viola elatior*. The conducted studies intended to (1) extend our knowledge on the impact of epigenetic variation in environmental adjustment and (2) to investigate the population genetic effects of persistent soil seed banks under changing selection regimes and decreasing population size.

Viola elatior is a perennial hemicryptophyte that within Central Europe is restricted to floodplain habitats of large river corridors, ranging from floodplain meadows to alluvial woodland fringes. Due to disturbance and succession the transition from one habitat type into the other may proceed rather fast. As the ability of *V. elatior* to compete for light is low, with increasing succession to closed forests, population sizes gradually decline and the species finally disappears from the aboveground vegetation. However, due to its persistent soil seed bank populations may recover even after years of absence. *Viola elatior* has an octoploid genome and exhibits a mixed mating system with potentially cross-pollinated chasmogamous (CH) and obligatory self-pollinated cleistogamous (CL) flowers. Seed production through CL flowers is dominating, leading to very high selfing rates and low genetic diversity within populations.

In the first and second study epigenetic variation was investigated in different populations from the Upper Rhine Valley (Germany). Generally epigenetic variation is thought to play an important role for the rapid adjustment of plants to dynamic environmental conditions. Modulating gene expression without changing the underlying genetic code, it might compensate for relative slow adaptations at the genetic level and could counteract the lack of genetic diversity. One of the most extensively studied epigenetic marks is the reversible methylation of DNA that is connected with numerous biological processes and might be inherited through meiosis over several generations.

To investigate the impact of DNA methylation in an ecological context, epigenetic studies mostly use methylation-sensitive amplification polymorphism (MSAP) analyses. However, no consensus exists on how to interpret and score the multistate information obtained from MSAP banding patterns. Therefore, the first study was intended to assess the effects of different MSAP scoring approaches on a small test data set of three *V. elatior* populations, and to justify a common scoring approach that allows for detailed and unbiased estimates. Overall, eight different scoring approaches – previously used variants and new alternatives – were applied to analyze 168 polymorphic MSAP markers.

Depending on the particular approach, between 78 and 286 polymorphic epiloci were scored that resulted in strongly varying estimates of epigenetic diversity and differentiation. Notwithstanding, linear regression and principal coordinates analysis revealed rather similar patterns and thus suggest that for multilocus analyses there seems to be not one best data analysis approach. However, for single-locus analyses like the search for epiloci under selection or the correlation of epiloci with environmental data, a new scoring variant was advocated that separately takes into account methylated as well as unmethylated MSAP fragments and thus seems appropriate to draw more detailed conclusions on population ecological processes.

In the second study the new approach was applied to a larger data set of six populations, three from meadow habitats and three from alluvial woodland habitats. Comparing amplified fragment length polymorphism (AFLP) and MSAP data revealed low levels of genetic (H'_{gen} = 0.19) and epigenetic (H'_{epi} = 0.23) diversity and high genetic (ϕ_{ST} = 0.72) and epigenetic (ϕ_{ST} = 0.51) population differentiation. Diversity and differentiation estimates were significantly correlated, suggesting that epigenetic variation partly depends on the same driving forces as genetic variation. However, even though correlation analyses detected almost equal levels of genetic (17.0%) and epigenetic (14.2%) markers that were correlated with site-specific light availability, principal coordinates analyses and Mantel tests showed that overall epigenetic variation was more closely related to habitat conditions than genetic variation. In agreement with these findings, genome scan analysis identified only very few AFLP markers (0-4.5%) that seemed to be under positive selection. This suggests that environmentally induced methylation changes indeed may play a major role for the transient or even heritable adjustment to dynamic environmental condition. Additionally, the new MSAP-scoring approach revealed that mainly unmethylated and CG-methylated states of epiloci contributed to population differentiation and putative habitat-related adaptation, whereas CHG-hemimethylated states only played a minor role.

In the third study, aboveground (AG) and seed bank (SB) samples of *V. elatior* were compared in 15 populations from the Upper Rhine Valley (Germany) and the Thaya/Morava floodplain (Czech Republic). To test if persistent seed banks might buffer the detrimental genetic effects of declining population sizes and/or changing selection regimes, the sampling was conducted along the same successional gradient from meadow to woodland habitats. Strikingly, AFLP analysis revealed significantly higher AG than SB genetic diversity in meadow (band richness Br = 1.20 vs. 1.17; pairwise private band richness PBr_p = 0.06 vs. 0.02) but not in woodland habitats (Br = 1.14 vs. 1.13; PBr_p = 0.07 vs. 0.05). Moreover, three of eight late successional populations even showed higher SB than AG diversity, indicating that persistent seed banks can accumulate genetic diversity. The results overall suggest that in *V. elatior* relative SB diversity (i.e. compared to AG diversity) and thus the SB genetic buffer capacity increases with ongoing succession and despite decreasing population size, counteracting effects of drift and selection, and hence assuring a higher chance for the species' long term persistence after disturbance events. Most likely and as corroborated by much lower small-scale genetic structure in late successional habitats, the observed change in relative SB diversity seems to be initially driven by a change in the allocation to CL and CH capsules along the gradient.

Zusammenfassung

Die Fähigkeit, sich an variierende biotische und abiotische Bedingungen anzupassen, ist eine entscheidende Eigenschaft von Pflanzen, um in einer sich ständig verändernden Umwelt zu überleben. Im Laufe der Evolution haben Pflanzen daher eine Reihe von Mechanismen entwickelt, die helfen mit der Heterogenität und Unvorhersagbarkeit ihrer Lebensbedingungen zurechtzukommen. Die vorliegende Thesis untersucht zwei dieser Mechanismen und ihre populationsökologischen Folgen in der kleistogamen Pflanzenart *Viola elatior*. Ziel der durchgeführten Studien war es, eine besseres Verständnis über (1) den Einfluss epigenetischer Variabilität in Bezug auf Umweltanpassungen zu erlangen und (2) die populationsgenetischen Effekte von persistenten Samenbanken unter sich verändernden Selektionsregimen und sinkender Populationsgröße zu erforschen.

Viola elatior ist eine mehrjährige, hemikryptophytisch wachsende Pflanzenart, deren Verbreitung innerhalb Zentral Europas auf Stromtalhabitate entlang von großen Flusssystemen begrenzt ist. Hier kommt sie entlang eines Lichtgradienten von offenen Stromtalwiesen hin zu Lichtungen und Rändern von Auwäldern vor. Durch Störungsereignisse und Sukzession kann der Wandel von einem Habitat Typ in den anderen innerhalb sehr kurzer Zeiträume stattfinden. Da die Konkurrenzfähigkeit von *V. elatior* mit sinkendem Lichtangebot abnimmt, kommt es mit zunehmender Sukzession hin zu Auwald Habitaten zu einer graduellen Abnahme der Populationsgröße, bis die Art schließlich ganz aus der Vegetation verschwindet. Aufgrund einer persistenten Bodensamenbank ist es jedoch möglich, dass Populationen auch nach Jahren der oberirdischen Abwesenheit wieder neu entstehen können. *Viola elatior* hat ein octoploides Genom und weist ein gemischtes Fortpflanzungssystem, mit potentiell fremdbestäubten chasmogamen (CH) und obligatorisch selbst-bestäubten kleistogamen (CL, cleistogamous) Blüten auf. Die Samenbildung durch CL Blüten ist jedoch vorherrschend, was zu einer sehr hohen Selbstbefruchtungsrate und vergleichsweise geringer genetischer Diversität innerhalb von Populationen führt.

Die ersten beiden Studien dieser Thesis untersuchen die epigenetische Variation innerhalb verschiedener Populationen aus dem Oberen Rheintal in Deutschland. Generell wird angenommen, dass epigenetische Variation eine wichtige Rolle bei der Anpassung von Pflanzen an sich verändernde Umweltbedingungen spielt. Da sie die Expression von Genen modelliert ohne den zugrunde liegenden genetischen Code zu verändern, ist es möglich, dass epigenetische Veränderungen relativ langsame Adaptionen auf der genetischen Ebene kompensieren und zudem die Konsequenzen von geringer genetischer Diversität ausgleichen können. Einer der bisher am besten untersuchten epigenetischen Mechanismen ist die reversible Methylierung von DNA. Sie wird mit einer Vielzahl von

biologischen Prozessen in Verbindung gebracht und kann meiotisch über mehrere Generation weitervererbt werden.

Um den Einfluss von DNA-Methylierung in einem ökologischen Kontext zu untersuchen, benutzen die meisten epigenetischen Studien MSAP-Analysen (Abk. für methylation-sensitive amplification polymorphism). Jedoch besteht bisher Uneinigkeit darüber, wie genau die multistaten Informationen aus den MSAP-Bandenmustern interpretiert und bewertet werden sollen. Aus diesem Grund diente die erste Studie dazu, den Effekt verschiedener „Scoring"-Varianten anhand eines Testdatensatzes von drei *V. elatior* Populationen zu untersuchen und zudem neue Möglichkeiten der Datenauswertung zu entwickeln, die detaillierte und unverfälschte Berechnungen erlauben. Insgesamt wurden acht verschiedene „Scoring"-Varianten getestet, um 168 polymorphe MSAP-Marker zu analysieren. Abhängig von der jeweils verwendeten Methode wurden zwischen 78 und 286 polymorphe Epiloci gewertet, die in stark variierende Werte für epigenetische Diversität und Differenzierung resultierten. Nichtsdestotrotz ergaben lineare Regressions- und Hauptkoordinatenanalysen nur sehr geringe Unterschiede zwischen den Ansätzen, was somit zeigt, dass es für Multilocus-Analysen generell keinen besten und alleingültigen „Scoring"-Ansatz gibt. Im Gegensatz dazu schlussfolgert die Studie, dass für Einzellocus-Untersuchungen, wie z.B. der Suche nach Loci unter Selektion oder der Korrelation von Epiloci mit Umweltparametern, eine neu entwickelte „Scoring"-Variante verwendet werden sollte. Diese berücksichtigt sowohl die Informationen von methylierten als auch unmethylierten MSAP Fragmenten und erscheint somit geeignet, noch detailliertere Rückschlüsse auf populationsökologische Prozesse zu ziehen.

In der zweiten Studie wurde die neue „Scoring"-Variante auf einen größeren Datensatz von insgesamt 6 Populationen angewandt, drei aus offenen Wiesen-Habitaten und drei aus Auwald-Habitaten. Der Vergleich von AFLP- (Abk. für amplified fragment length polymorphism) und MSAP-Daten ergab relativ geringe Werte für die genetische ($H'_{gen} = 0.19$) und epigenetische ($H'_{epi} = 0.23$) Diversität und eine insgesamt sehr hohe genetische ($\phi_{ST} = 0.72$) und epigenetische ($\phi_{ST} = 0.51$) Differenzierung zwischen den Populationen. Sowohl die Diversitäts- als auch Differenzierungmaße waren signifikant korreliert, was vermuten lässt, dass epigenetische Variation zumindest teilweise den gleichen treibenden Kräften wie genetische Variation unterliegt. Obwohl mit korrelationsbasierten Analysen vergleichbar hohe Anteile an genetischen (17.0%) und epigenetischen (14.2%) Markern gefunden wurden, die signifikant mit der populationsspezifischen Lichtverfügbarkeit korreliert waren, konnten Hauptkoordinatenanalysen und Mantel Tests zeigen, dass insgesamt epigenetische Variation deutlich dichter mit den Habitat-Bedingungen in Bezug steht als genetische Variation. Damit im Einklang stehen auch die Ergebnisse von zwei Genome Scan Analysen, die nur sehr wenige AFLP-Marker (0-4.5%) detektieren konnten, die mutmaßlich unter positiver Selektion stehen. Insgesamt lassen die Ergebnisse darauf schließen, dass umweltabhängig induzierte Veränderungen von Methylierungsmustern tatsächlich eine zentrale Rolle für die transiente oder sogar transgenerationale Anpassung von Populationen an dynamische Umweltbedingungen spielen. Zusätzlich zu diesen Ergebnissen zeigte die Anwendungen der neuen „Scoring"-Variante auch neue, funktionelle Einblicke in das Zusammenspiel verschiedener Methylierungsvarianten. So

erklärten hauptsächlich Epoloci in unmethylierten und CG-methylierten Zuständen die Differenzierung zwischen Populationen und Habitat bezogene Adaptation, wohingegen CHG-Hemimethylierungen insgesamt nur eine untergeordnete Rolle spielten.

In der dritten Studie wurden schließlich oberirdische (AG, aboveground) und Samenbank-Proben (SB) von 15 Population aus dem Oberen Rheintal (Deutschland) und dem Thaya/Morava Stromtal (Tschechien) verglichen. Um zu testen, ob persistente Samenbanken die nachteiligen genetischen Effekte von verringerter Populationsgröße und/oder der Veränderung von Selektionsregimen puffern können, wurde die Beprobung entlang des gleichen Sukzessionsgradienten von Wiesen- hin zu Auwald-Habitaten durchgeführt. Beeindruckenderweise, ergaben die AFLP-Analysen signifikant höhere genetische Diversitätswerte für AG als SB Individuen in Wiesen (Band Richness $Br = 1.20$ vs. 1.17; paarweise Private Band Richness $PBr_p = 0.06$ vs. 0.02), jedoch nicht in Auwald-Habitaten ($Br = 1.14$ vs. 1.13; $PBr_p = 0.07$ vs. 0.05). Darüber hinaus zeigten drei von acht Auwald-Populationen sogar höher SB als AG Diversität, was darauf hindeutet, dass persistente Samenbanken möglicherweise genetische Diversität akkumulieren können. Insgesamt zeigen die Ergebnisse, dass in *V. elatior* die relative SB Diversität (d.h. verglichen mit der AG Diversität) und entsprechend die Fähigkeit der Samenbank genetische Veränderungen abzupuffern, mit fortschreitender Sukzession und trotz sich verringernder Populationsgröße ansteigt. Somit kann sie den Effekten von Drift und Selektion entgegenwirken und langfristig dazu beitragen, das Bestehen der Art nach Störungsereignissen zu sichern. Die wahrscheinlichste Erklärung für den beobachteten Anstieg der relativen SB-Diversität ist vermutlich eine Änderung der Allokation zu CL und CH Samenkapseln entlang des Gradienten. Dies wird auch durch eine deutlich geringere kleinräumliche genetische Struktur in Auwald-Habitaten bekräftigt.

Acknowledgments

The completion of thesis would not have been possible without the support and helpful contributions of numerous people. I would like to express my deepest gratitude to all of them. In particular, I thank …

- Dr. Lutz Eckstein who initiated this project and gave me the opportunity to work on this interesting research topic. With helpful advises, fruitful discussions and patience he alway supported me in all aspects of research and writing. Moreover, many enriching conversations with him gave me food for thought on numorous topics besides the professional life.
- Prof. Dr. Dr. Annette Otte for enabling a very pleasant and constructive working environment in the division of Landscape Ecology and giving me a wide scope to develop and build on my scientific skills.
- Dr. Walter Durka for mentoring me in the fields of molecular ecology and population genetics, and for giving the impetus to include also epigenetic studies into this project.
- Prof. Dr. Tom Wilke for his willingness to be the second referee.
- Josef Scholz vom Hofe for his invaluable help in almost all aspects of field and greenhouse work, and for his many good ideas and extensive experience.
- Dr. Jiří Danihelka for his support to find and sample *V. elatior* populations in the Czech Republic.
- Lena Kretz, Zahed Shakeri, Martina Herrmann and Ina Geier for their assistance during field and laboratory work.
- Dr. Tobias Donath, Dr. Dietmar Simmering, Miriam Bienau, Sarah Harvolk, Martin Wiesmair, Tim Theißen, Anja Magiera, Dr. Kristin Ludewig, Theresa Rhül, Fang Xu and all my other colleagues from the Division of Landscape Ecology for the good time and the inspiring discussions in the coffee and lunch breaks.
- the Darmstadt Regional Council for issuing the necessary permits.
- the Deutsche Forschungsgemeinschaft (DFG) for providing financial support.
- and last but not least, Lena, and all of my family and friends who always supported and encouraged me throughout my PhD time.

List of publications

International journals

Schulz B., Eckstein R.L., Durka W. (2013) Scoring and analysis of methylation sensitive amplification polymorphisms (MSAP) for epigenetic population studies. *Molecular Ecology Resources*, 13, 642–653.

Schulz B., Eckstein R.L., Durka W. (2014) Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Molecular Ecology*, 23, 3523–3537.

Reiker J., Schulz B., Wissemann V., Gemeinholzer B. Does origin always matter? Evaluating the influence of non-local seed proveniences for ecological restoration purposes in a widespread and outcrossing plant species. (accepted in *Ecology & Evolution*)

Submitted manuscripts

Schulz B., Durka W., Danihelka J., Eckstein R.L. Can persistent seed banks buffer genetic effects of declining population size and selection?

Versicherung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe.

Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

______________________________ ______________________________

(Benjamin Schulz) Ort, Datum