DNA methylation shows footprints of altitude selection in the clonal plant species *Fragaria vesca*

Audrey Le Vève[1,+], Iris Sammarco[2,+], Vitek Latzel[2], Clément Lafon Placette[1,*]

[1] Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

[2] Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic

[+] shared authorship

[*] corresponding author: lafonplc@natur.cuni.cz

The paper is available on BioRXiv server :

Classification: Biological Sciences; Ecology, Environmental Sciences, Evolution, Population Biology

Abstract

Climate change poses a significant threat to plant species, potentially pushing them beyond their adaptive capacities. Epigenetic modifications, such as DNA methylation, have emerged as a key candidate mechanism enabling plants to quickly adapt to environmental changes by generating locally adapted phenotypes. These phenotypic changes can be even inherited across multiple generations, thus potentially becoming targets of natural selection. However, whether natural selection can act on these epialleles has hardly been tested directly. Addressing this knowledge gap is crucial as population survival may heavily rely on DNA methylation, especially in scenarios with restricted genetic diversity, such as within clonal plant populations.

Here, we employed population genomics approaches on seven natural populations of the clonal species wild strawberry (*Fragaria vesca*) distributed across an altitudinal range to investigate the presence of epigenetic sites under altitude-driven selection. Our genomic, epigenomic, and transcriptomic analyses found a wide intra- and inter-population epigenetic diversity despite a considerably low genetic diversity. We identified heritable epialleles exhibiting signs of positive selection related to altitude, i.e. reduced intrapopulation epigenetic diversity and increased epigenetic divergence between low and high altitude. These altitude-selected epigenetic loci overlapped with genes involved in biological processes such as DNA repair, molecular recognition, regulation of gene expression, and chromatin structure. Interestingly, some of these epialleles were independent of genetic variation, suggesting they may have arisen stochastically or in response to environmental variation.

These findings suggest that heritable epigenetic variation could help clonal species quickly adapt to environmental challenges as those related to varying altitudes and/or temperatures.

Keywords: epigenetic diversity, epigenetic inheritance, natural selection, altitude adaptation, clonal species, wild strawberry populations

Significance statement

The eventuality of epigenetic marks being inherited has led the scientific community to rethink evolution and adaptation, debating whether epigenetic information could be considered as the "fifth base". However, direct evidence supporting that epigenetic marks are under evolutionary processes such as natural selection remain scarce. To address this knowledge gap, we screened natural populations of wild strawberries and searched for footprints of altitudinal selection on epi-alleles. We found loci matching the criteria for epigenetic selection, i.e. reduced epigenetic diversity in populations, increased epigenetic divergence between populations of different altitudes and transmission to the (clonal) progeny. Some of these selected epi-alleles were not associated with any genetic variant, suggesting that epigenetic marks are another layer of substrate for selection to act.

Introduction

Given their sessile nature, plants need to constantly respond to changing environments, i.e. show phenotypic plasticity. Local adaptation and plasticity to a specific environment often lead to the same consequences on the phenotype, which suggests that "phenotypic plasticity takes the lead in adaptive evolution" (1). In this context, the role of the epigenetic marks and their inheritance in transgenerational phenotypic plasticity has been the target of interest for evolutionary biologists (2). Indeed, plants can adjust their phenotypes through epigenetic modifications and these modifications involve mitotically (and partly meiotically) heritable changes in gene expression (3-4). Thus, some epigenetic marks can be stably inherited across generations in plants, and more specifically in clonal species (5-6). However, and concurrently, epigenetic marks are reversible. They undergo changes depending on environmental stimuli, and many of these changes are reversed once the stimulus is gone (7). In addition, most of them are reset during sexual reproduction, together with stress memory (8-10). Epigenetics is expected to be particularly important for the adaptation of clonal species, as they do not experience memory resetting of sexual reproduction, and specifically in case of reduced genetic polymorphism limiting their genetic evolvability (6). If epigenetic marks are more stable in clonal species, epigenetic variation might be the substrate for selection instead of genetic variation. This assumption remains, however, to be tested in natural populations.

Among all the epigenetic marks, DNA methylation has received most of the attention in respect to evolution, as it is a covalent modification of DNA and as such appears to be more stable than other marks (3). In plants, DNA methylation occurs in three sequence contexts: CG, CHG and CHH (with H representing A, C or T; 11). In all sequence contexts, DNA methylation plays a pivotal role in repressing the mobilisation of transposable elements (TEs; 12), while CG and CHH methylation appears to be connected to gene expression in a non-linear fashion (13-17). DNA methylation variation occurs at a considerably higher rate than genetic mutations (18) and can arise in response to genetic, environmental and stochastic factors (19-21). In particular, stochastic DNA methylation variants arise from errors in the maintenance of DNA methylation during DNA replication (22). Importantly, recent studies revealed extensive inheritance of environmentally-associated natural epigenetic variation across clonal generations, with the inheritance progressively decreasing from CG to CHG and CHH contexts (23-25). While there is potential for these inherited variants to shape offspring phenotypes and facilitate rapid adaptation (2,5), substantial empirical evidence remains elusive.

Thus, DNA methylation, just as genetic mutations, has the potential to meet all the criteria required for natural selection to act: new methylation profiles, or epi-alleles, can emerge in the genome over a few generations and some are potentially heritable (18, 26). These methylation profiles can exhibit variation within and between natural populations (27), and epi-alleles can influence gene expression, and therefore the fitness of individuals (5). However, this hypothesis has hardly been tested directly and usually population epigenomics studies entails just screening for natural epigenetic diversity. If a theoretical model demonstrates the possible selection on methylated sites, and their advantages in populations (28), this phenomenon could have a singular and unclear effect on diversity compared to classical signals found on genetic variants (28-29) that must be estimated on empirical data. The neutral theory of evolution stipulates that most of the observed (epi)genetic variation is not due to selection, and thus it is fundamental to distinguish the impact of selection from other factors affecting diversity in methylated sites, including demographic history, mutation rates, drift (29-30),

and in the case of epigenetic variation, the immediate response to the environment, which is not necessarily adaptive (31). To achieve this, we can employ population genomics assumptions. First, epigenetic loci under selection should exhibit reduced epi-allelic diversity compared to the rest of the epigenome. Second, epigenetic loci under selection should display elevated epi-differentiation between populations exposed to different selection pressures and reduced between populations exposed to similar selection pressures. Lastly, these epialleles should exhibit some degree of inheritance between generations despite environmental changes.

In this study, using publicly available genomic, epigenomic, and transcriptomic data obtained from natural populations of the widespread clonal species wild strawberry (*Fragaria vesca*; 32), we searched for signatures of selection on epialleles in response to altitude. For this purpose, first, we 1) used classical population genomics metrics, the polymorphism estimator (π), the Tajima's D, fixation index (Fst) and linkage disequilibrium (LD) of epimutations to evaluate the general tendencies of the epi-allele diversity within and between natural populations. Based on these indices, we 2) identified candidate epi-alleles under altitudinal selection. To assess their heritability, an important aspect for their adaptive potential, we compared the populations collected in their natural habitats with their clonal offspring grown in a common garden. Finally, 3) we distinguished the selected epi-alleles that were under genomic control from those that putatively arose due to environmental or stochastic variation and demonstrated that some selected methylation profiles were non genetically determined.

Results

Low DNA methylation variability within and between populations, and across generations in the CG and CHG contexts

We first estimated the general variation of epigenetic diversity on CG, CHG and CHH context based on epigenetic marks in seven natural populations of the clonal species *F. vesca* (Table S1). The mean DNA methylation average in the field populations was significantly higher for the CG, then for the CHG and finally for the CHH context (Fig 1A; Table S2; *p-values U-test* <= 1.75^{e-0}). To be able to use population genomics indices and infer DNA methylation diversity, we converted these methylation levels to threshold to define the methylated sites and biallelic states in each position. For the rest of the article, all analyses use methylated/unmethylated epialleles instead of DNA methylation levels (except for the correlation between methylation of candidate regions and gene expression, and the GWA study).

For each field population, we quantified the methylation polymorphism ($\boldsymbol{\pi}_{met}$), which represents the variation in DNA methylation across individuals at each site. Across the three contexts, the mean $\boldsymbol{\pi}_{met}$ in the field populations was significantly higher for the CHH, then for the CG and finally for the CHG context (Fig 1B; Table S2; *p-values U-test* <= 0.02). In the common garden, the $\boldsymbol{\pi}_{met}$ was significantly higher for the CHH (Fig 1B; *p-values U-test* <= 1.23^{e-03}), but not different between the CG and the CHG context (*p-value U-test* = 0.06). This global polymorphism includes sites completely fixed across all the populations. The proportions of fixed sites across all the field and garden populations were 85, 88 and 22% for the CG, CHG and CHH context respectively (Table S3). These sites offer insights into DNA methylation patterns consistent across all populations and environmental conditions. However, for the rest of this article, we will focus on variable methylated sites, as these provide more valuable information about selection processes.

Previous studies tried to make assumptions about the variation of the Tajima's D in epigenome taking into account the strong violations of the infinite sites model expected. However, the predictions were completely opposite depending on the study (28-29). Thus, we estimated the Tajima's D for methylation and genomic SNPs from the field populations to test the effect of the type of methylation sites on Tajima's D. Irrespective of the methylation context considered, the Tajima's D was always lower for DNA methylation compared to genomic SNPs (Median=0.40; Fig 1C; *p-values U-test* <= 2.00^{e-16}). Then the Tajima's D was higher for the CHG (Median=0.06), the CG (Median=-0.03) and finally the CHH, which presented also a binomial distribution (Median=-0.39; Fig 1C; *p-values U-test* <= 5.54^{e-09}).

To estimate the conservation of the methylation profiles across generations and environments, the methylation profiles of each individual from the field were compared to the profiles of the clonal offspring grown under common garden conditions (Table S4). For the CG, CHG and CHH methylation profiles respectively, 90, 81 and 78% of the methylation profiles of the individuals from the field were conserved in the common garden (Table S5). Considering that the epimutations in the common garden were inferred after three clonal generations, we estimated the epimutation rates for methylation gain (unmethylated sites in the field that became methylated in the common garden). The rates were 2.66^{e-02} for CG, 4.92^{e-02} for CHG and 15.38^{e-02} for CHH epimutations per site per generation. With the same method, we estimated that the epimutation rates involving methylation loss (methylated sites in the field that became unmethylated in the common garden) were 3.99^{e-02} for CG, 8.24^{e-02} for CHG and 5.16^{e-02} for CHH epimutations per site per generation. When considering all

contexts and sites that exhibit changes in methylation profiles between field and common garden conditions, the mean epimutation rate is approximately 6.73 x10^{-2} epimutations per cytosine per generation.

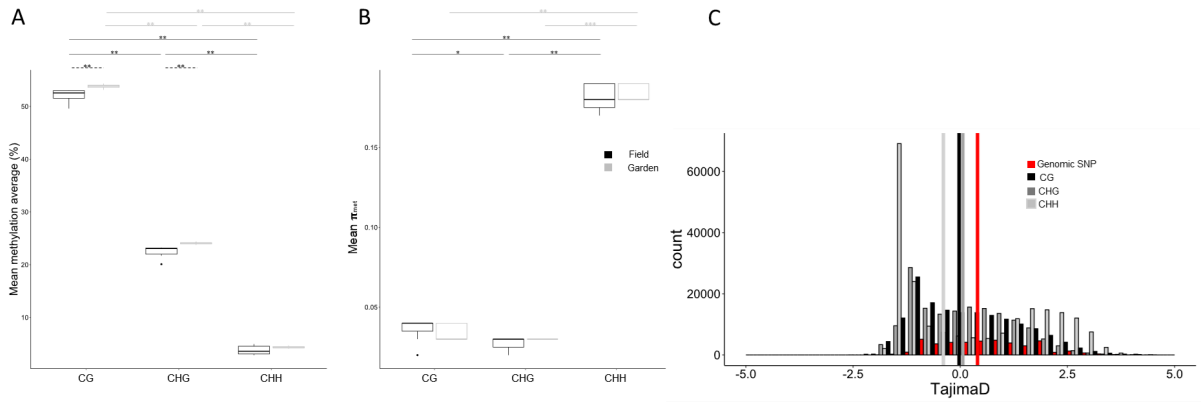

Fig 1: Polymorphism observed in the CG, CHG and CHH methylation context.

Rapid decrease of the linkage disequilibrium for the three methylation contexts

While linkage disequilibrium (LD) is expected to be high and maintained over a long distance for genetic variants in clonal species, we expected the opposite trend for epialleles, due to the high epimutations rate. Furthermore, we expected LD to be lower the higher the rate of epimutation, i.e. at CHH sites.

The LD between CG, CHG or CHH variable methylation sites separated by a maximum distance of 30 kb was compared to LD obtained by genomic SNPs (Fig 2). Over the 30kb, the mean LD between two genomic SNPs varied between 0.33 and 0.76. Specifically, the mean LD between two genomic SNPs remained high, ranging from 0.49 to 0.42 for distances between 1 and 30,000 bases. In contrast, the mean LD between two methylation sites was relatively low in all sequence contexts (Fig 2). For instance, in the field populations, the mean LD between two CG-sites never exceeded 0.56 and decreased by a factor of 10 in the 30 kb: the mean LD between two variable CG sites between 1 and 30 kb were 0.56 and 0.06 respectively. The mean LD between two variable CG sites reached for the first time the minimum value (0.06) at the distance of 3,546 bp. For the CHG sites, the mean LD between two variable sites between 1 and 30,000 bp were 0.40 and 0.06 respectively, and this value was obtained for the first time at a distance of 5,246 bp. Finally, for the CHH context, the mean LD between two variable sites between 1 and 30,000 bp were 0.21 and 0.04 respectively, and this value was obtained for the first time at a distance of 4,050 bp. We observed similar tendencies, in the common garden (Fig. S1).

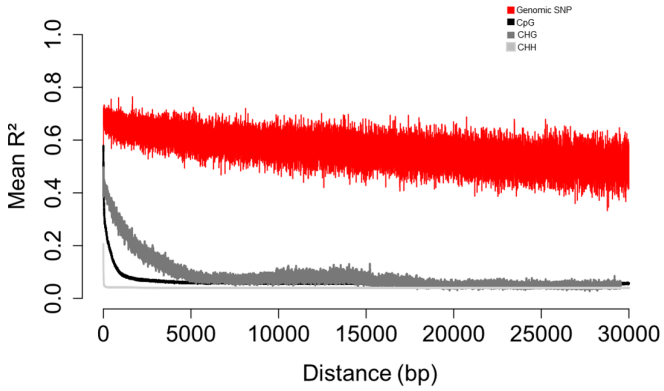

Fig 2: Decrease of the linkage disequilibrium with the distance between two variable sites across all the methylation contexts.

T*he CG and CHG contexts differentiated the populations from the field*

In order to visualise the population structure associated with variable methylation profiles, we estimated the differentiation among individuals. We used these differentiation estimates to construct a kinship matrix (Fig 3). We focused on the structuration by altitude levels of the field populations, negatively correlated with temperature (Table S1). The differentiation of CG methylation profiles between two individuals varied between 9 and 30% (average=25%). The kinship matrix for CG revealed an accurate structure by population, with the exception of individual 06 of population FV_IT_04, and interestingly a structure by altitude (Fig 3A). In the common garden, the differentiation of CG methylation profiles between two individuals from the common garden varied between 8 and 25% (average=21%). As for the field, the kinship matrix resulting from these differentiation values presented an almost perfect structuring by population (Fig S2A), except for one individual from a high-altitude population (>1,000m), FV_IT_02, which clustered with low-altitude populations.

The differentiation of CHG methylation profiles between two individuals from the field varied between 18 and 30% (average=25%). The resulting kinship matrix presented also an almost perfect structuring by population and altitude, with the exception of individual 02 of population FV_IT_06 and the individuals 03 and 05 of population FV_IT_04 (Fig 3B). In the common garden, the differentiation of CHG methylation profiles between two individuals varied between 14 and 24% (average=21%). However, the kinship matrix resulting from the CHG sites seemed to have lost part of this structuring by population and altitude (the individual 01 of population FV_IT_02 finds itself isolated, as well as one individuals of population FV_IT_04 and two from population FV_IT_03; Fig S2B).

On CHH sites for individuals from the field (Fig 3C), the differentiation of methylation profiles between two individuals in the field varied between 15 and 28% (average=23%). The differentiation between two individuals from the common garden varied between 18 and 24% (average=20%). Contrary to CG and CHG profiles, the kinships obtained in the field and common garden did not present structure by population or altitude (Fig 3C; Fig S2C).

To verify if these environmental structuring of methylation profiles were related to the genomic structure, the kinship matrices obtained from methylation profiles were compared to the kinship from SNP data of each individual (Fig 3D). The differentiation of the genomic profiles between two individuals varied between 6 and 41% (average=29%). Moreover, the kinship matrix derived from genomic data did not provide a clear structuring by population or altitude (Fig 3D). For the three methylation contexts, the kinship was significantly and positively correlated with the kinship obtained for the genomic data (Table S6), and the r^2 were always higher than 0.6. In the populations from the common garden, the same tendencies were observed (Table S6).

We further characterised population epigenetic differentiation using the mean methylation fixation index (Fst_{met}). Fst_{met} between each pair of field populations was higher in CG, then for the CHG and finally for the CHH context (*p-values U-test* < 1.22^{e-03}; Fig 4A; Table S7). In the common garden populations, only the Fst_{met} of the CHG sites significantly decreased (*p-value U-test*= 2.91^{e-08}; Fig 4A; Table S7).

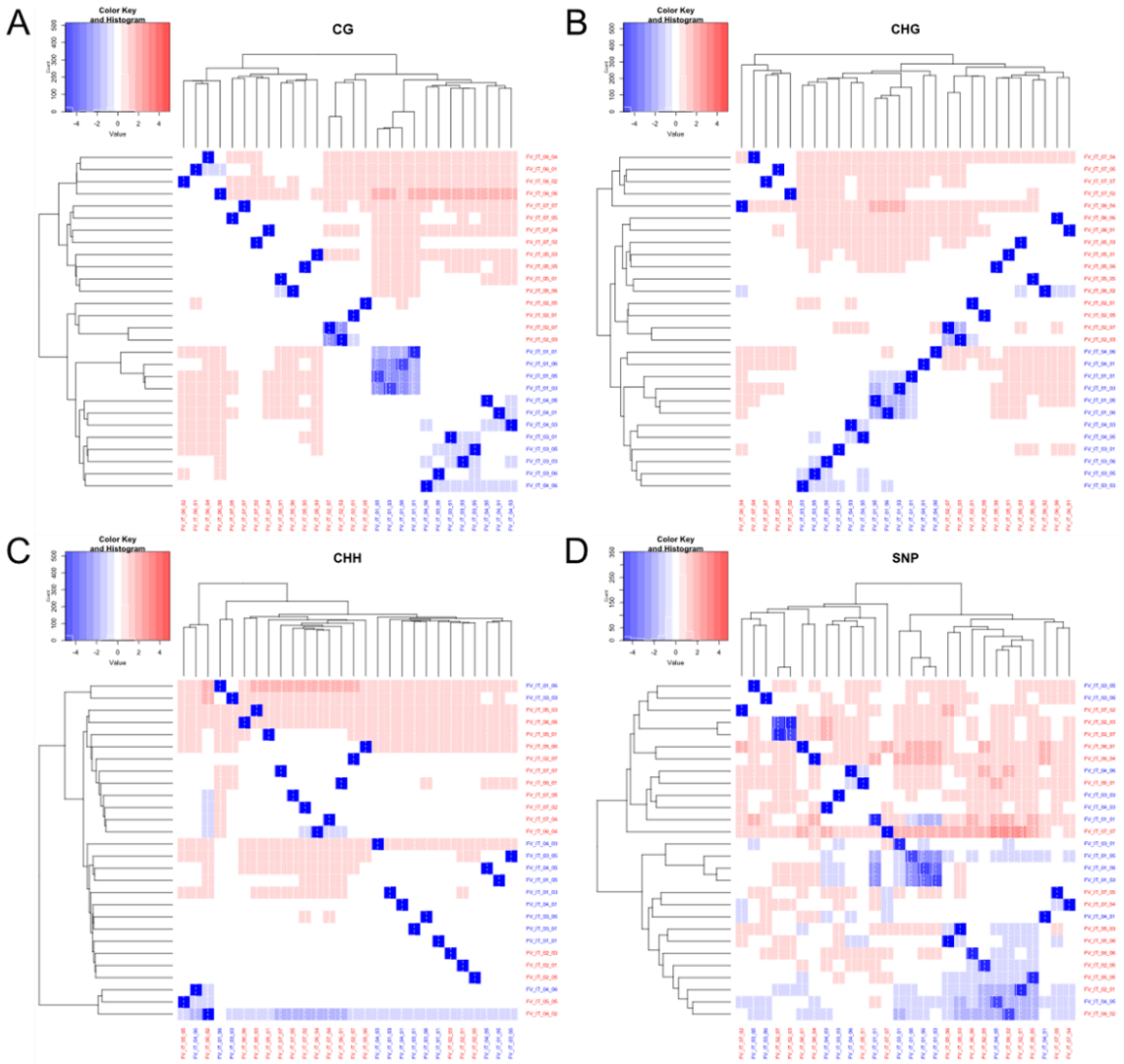

Fig 3: Structuration by altitude of the field populations for the methylation and genetic profiles.

CHH sites were more influenced by altitude

In order to estimate the global effect of altitude on differentiation of epigenetic profiles on the populations, we estimated the variation of the *Fst* between pairs of populations, depending on the altitude level in the field.

Based on genomic SNPs, the mean *Fst* from the field between two populations at high altitude levels was lower than between two populations at low altitude and then between populations of different altitude levels (*p-values U-test* < 2.2e-16; Table S6). Contrary to genomic SNPs, for the three methylation contexts, for the field conditions, the mean Fst_{met} between two populations at high altitudes were not significantly different from the Fst_{met} obtained between two populations at low altitudes for the three contexts (Fig 4B to D, black boxplot). Moreover, the mean *Fst* exhibited lower values between two populations at high altitudes compared to the Fst_{met} obtained between two populations at different altitude levels (*p-values U-test*< 0.01; Fig 4B to D; Table S7). For the CHH context, the Fst_{met} between two populations at low altitudes were also lower than the Fst_{met} obtained between two populations at different altitude levels (*p-values U-test* = 9.38^{e-03}).

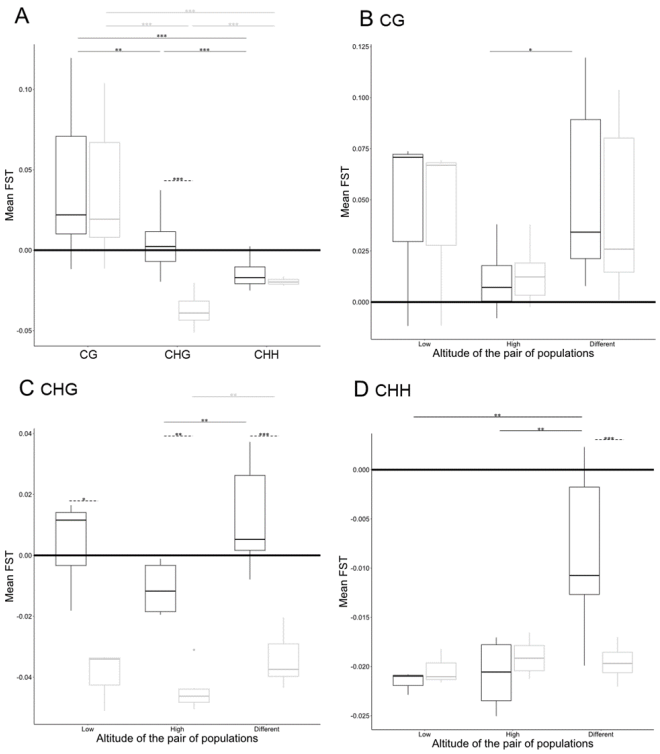

Fig 4: The mean Fst_{met} is higher between populations from different altitudes for the CHH methylation contexts.

In the common garden (Fig 4B to D, grey boxplot), the mean Fst_{met} for the CG sites did not vary significantly between populations from the field and common garden (Fig 4B), whereas the mean Fst_{met} decreased significantly for the CHG sites, irrespective of the altitude levels of the populations, in the garden compared to the field (*p-values U-test* <0.05; Fig 4C). The Fst_{met} stayed significantly lower between two populations from high altitude levels compared to two populations from different altitude levels (*p-values U-test* =8.73^{e-03}). Finally, for the CHH sites, the Fst_{met} decreased between two populations from different altitude levels (*p-values U-test* = 8.12^{e-04}; Fig 4D) and was

not anymore significantly higher than the Fst_{met} obtained between two populations from the same altitude.

Candidate regions associated with selection by altitude

Selection is classically detected by decay of polymorphism on neutral expectations. However, the high epimutation rate modifies drastically the assumptions about the neutrals expectation for selection on epiallele (28-29). Thus, we preferred to detect selection driven by altitude based on common variations of polymorphism across populations with similar altitude range rather than by comparison with neutral expectations. Thus, we defined candidate sites for selection as they exhibited high Fst_{met} (>0.2) between populations from different altitudes in the field, and consistently low Fst_{met} (<=0) between populations at the same altitude. Then, the candidate regions associated with altitude were identified as all regions exceeding 140 bases (~size of a nucleosome), containing more than three candidate sites spaced less than 1kb apart.

In the field populations, 90, 53 and 998 regions were associated with altitude for the CG, CHG and CHH contexts respectively. One of them, located between the positions 14,197,301 and 14,198,851 on chromosome 5, was also found as a candidate region under selection for altitude in CHG and CHH sites. In this region, we observed a clear increase of the methylation average and a decrease of the π_{met} on populations in high altitude for the three contexts (Fig S3A). This region is associated with the promoter and coding region of an unknown gene (FvH4_5g22911), and with the promoter of a gene encoding a fibrillin (FvH4_5g22910) and one transposable element (TE; Table S7). Eight other candidate regions were found common to either the CG and CHG or CG and CHH contexts, some of them being associated with genes and/or TEs (Table S7). In addition, one region was also found as a candidate region for both CHG and CHH contexts associated with one TE (Table S7).

Selection is expected only if the trait is inherited independently of the environment. Thus, candidate regions under selection by altitude were further distinguished to the regions associated with acclimation as the regions are also associated with altitude in samples from the common garden by the same method (Table S7). Thus, 75, 22 and 84 regions were found in the populations from the common garden for the CG, CHG and CHH contexts respectively. However only 42, 12 and 23 regions were in common with the candidate found in the field. Together, only 47, 23 and 2% of the candidate regions associated with altitude in the field for the CH, CHG and CHH context respectively were under selection. The candidate regions that were common between field and garden conditions in the CG context showed enriched gene ontology (GO) terms related to DNA repair and molecular recognition, whereas candidate regions in the CHG context showed enrichment in a term related to the regulation of gene expression. Finally, the candidate regions in the CHH context showed enrichment of terms related to the regulation of protein folding and chromatin structure (Fig 5). Between the candidate region under selection for altitude, we found the region on chromosome 5 previously mentioned that exhibited conserved methylation profiles in the common garden (Fig S3B), the eight common regions to either the CG and CHG or CG and CHH contexts and the common region for both CHG and CHH contexts. For the other candidate CG regions, 21 were located in the coding region of 23 genes, 10 in 11 promoters, and 5 in 7 TEs (Table S7). For the other candidate CHG regions under selection for altitude, two were not associated with any gene or TE, one was located in the coding region and the promoter of one gene and the last two in two promoters (Table S7). Finally, for the other candidate CHH regions under selection, four were not associated with any gene or TE,

eight were located in the coding region of eight genes, ten were located in the promoters of 13 genes and 12 were located on 17 TEs (Table S7).

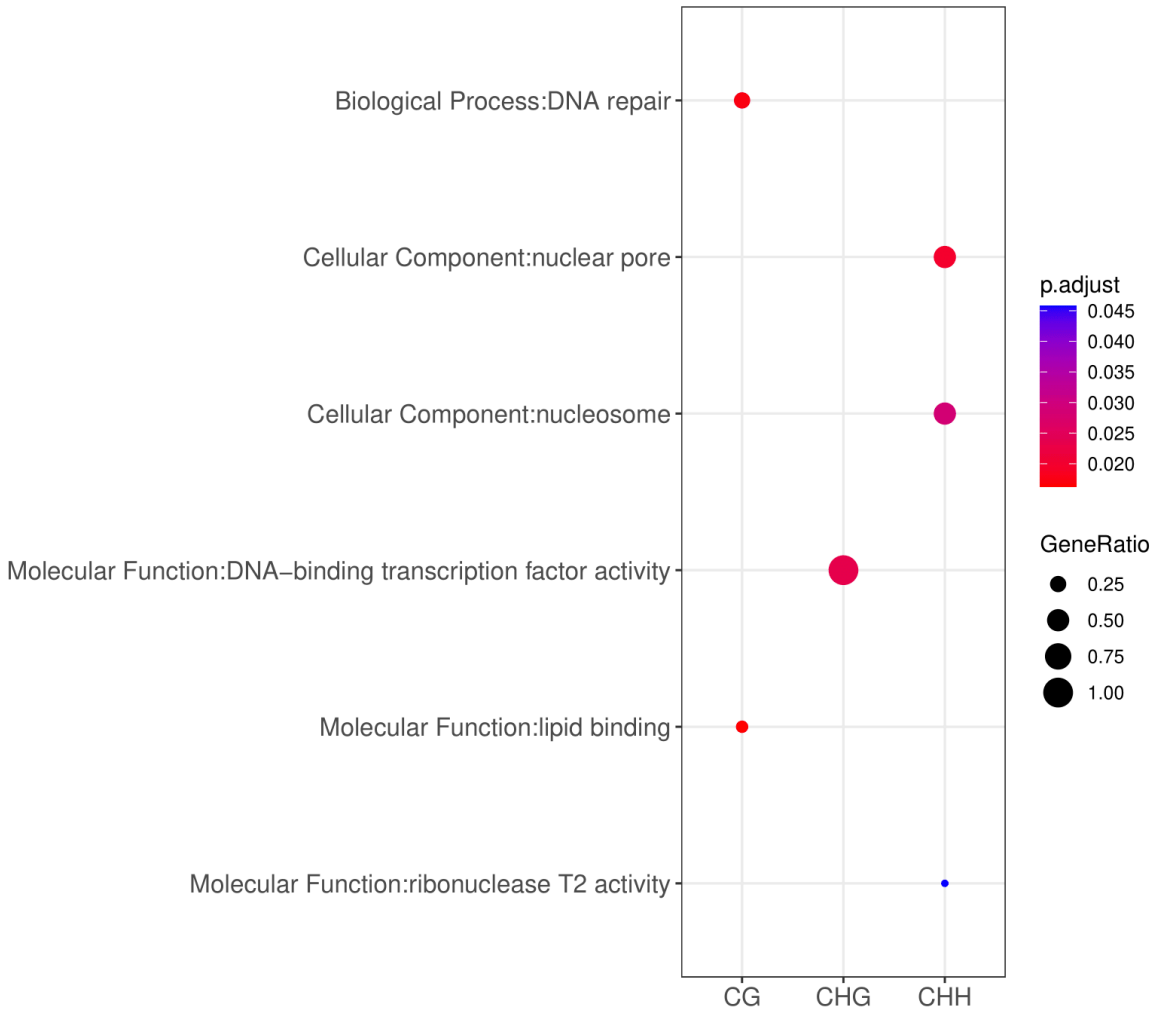

Fig 5: Gene Ontology (GO) enrichment analysis performed on the candidate genes associated with altitude selection that were conserved across both field and garden conditions.

We took advantage of these results to verify the predictions of Charlesworth and Jain (28) concerning Tajima's D variation for a selected methylated region. The Tajima's D obtained in windows of 1 kb that overlapped the candidate regions associated with altitude in the field and in the common garden were compared with the Tajima's D obtained in windows that did not overlap candidate regions for each context independently. Interestingly, regardless of the context considered, the values of the Tajima's D in candidate regions were always significantly higher than those in other genomic regions (*p-values U-test* <=1.25^{e-09}; Fig S4).

Correlation analysis between methylation of candidate regions associated with altitude selection and expression of DEGs

To assess the potential impact of methylation in candidate regions on gene expression in *cis*, we overlapped the list of genes associated with these regions with those that were differentially expressed (DEGs) between low- and high-altitude populations. For this analysis, we focused on the candidate regions conserved in field and garden conditions, as these could potentially have an evolutionary role. However, we found no overlap between the candidate regions common to field and garden and DEGs.

We then assessed whether we could find any *trans*-acting correlation between methylation of candidate regions and gene expression. We found 26, 16 and 3 positive and 10, 4 and 5 negative correlations in the CG, CHG and CHH contexts respectively (Fig S5; Table S8). The candidate regions correlated with gene expression presented a GO enrichment in DNA repair for CG, and in RNA and protein processing for CHH (Table S9). On the other hand, the genes whose expression was significantly correlated with methylation were enriched in terms related to cell wall degradation and protein kinase activity.

Genetic basis of candidate regions: Genome-Wide association (GWA) analysis

The conservation of methylation patterns of selection candidate regions between field and common garden could be explained by the involvement of genetic factors. To test this hypothesis, we conducted a GWA analysis using methylation of the candidate regions as phenotypes and SNPs as predictors (Table S10). Significant associations were detected with SNPs for 15 candidate regions in CG (36% of the candidate regions), 3 in CHG (25%) and 7 in CHH (29%). The SNPs detected were majoritively associated with non overlapping candidate regions (100, 98 and 98% for the CG, CHG and CHH context respectively). The majority of the SNPs overlapped with TEs (133, 44 and 115 for the CG, CHG and CHH context respectively).

Consequently, we could not find any association between SNPs and the methylation of candidate regions for 27 regions for CG context (64% of candidate regions), 9 for CHG (75%) and 16 for CHH (70%). The GO enrichment analysis of genes included in these regions revealed an enrichment of a term related to DNA binding activity for the CHG and of terms related to protein folding and structure, chromatin structure, and DNA repair for the CHH sequence (Fig. S6). Interestingly, the candidate region highly conserved across populations in the field and in the garden for the three different contexts (Fig S3A) was not associated with any SNPs for any of the three contexts.

Discussion

DNA methylation, as genetic mutations, can fulfil the prerequisites of natural selection: methylation profiles, known as epi-alleles, can be heritable across generations, affecting gene expression, and eventually fitness (5). In clonal species, DNA methylation could be particularly essential for adaptation to environmental fluctuation (6). However, although studies of whole epigenome diversity at single-nucleotide resolution exist, these types of analyses have mainly been applied to model organisms under controlled conditions (21, 27). Studies in natural populations are still mostly limited to low-resolution molecular analyses (33). Here, we propose a comprehensive genomic, epigenomic and transcriptomic analysis at single-nucleotide resolution of natural populations of *F. vesca*, sampled across an altitudinal gradient. We tested whether we could identify epiloci under altitude-driven selection in a predominantly clonally reproducing plant species and clarify the methylation dynamics in natural populations.

The CHH context is the most dynamic methylation context

Modelling studies predicted that an increase in mutation rates can have short-term adaptive advantages in clonal organisms when they are in changing environments (34). In *F. vesca*, we estimated a mean epimutation rate ~6.73×10^{-2} per cytosine per generation, whereas the genetic mutation rate for the strawberries genus is ~2.8×10^{-9} per site per year (35). Consistent with this observation, an extremely higher epimutation rate of CG sites compared to the genetic mutation rate was also reported in *Arabidopsis thaliana* (36). Moreover, as in *A. thaliana*, the epimutation rate was higher for demethylation than methylation (36).

In our study, the observed high frequency of epimutations suggests that they do not arise from genetic mutations, especially considering that we used a clonal species with reduced genetic diversity. Our estimates of linkage disequilibrium (LD) also support this idea, as LD remained high for genetic mutations but fastly decreased for epimutations. This was particularly true for CHH methylation, which showed higher dynamics than CG and CHG.

In agreement with this observation, the epimutation rate in CHH was twice as high compared to both CG and CHG. Moreover, contrary to CG and CHG contexts, the CHH context exhibited a high DNA methylation polymorphism (π_{met}) and a low conservation of methylation profiles across populations and between field and garden conditions in natural populations of *F. vesca*. These results contradict previous studies in *A. thaliana*, which demonstrated that the CG context exhibits the most variability between individuals and generations grown under controlled laboratory conditions, while the CHG and CHH contexts display the lowest levels of spontaneous epimutation rates (18, 36). This discrepancy could be attributed to the predominantly clonal reproduction of *F. vesca*. Vegetative propagation could lead to higher epimutation rates in CHH methylation compared to sexual reproduction due to the absence of meiosis that resets many DNA methylation patterns especially in this context (7-8). Furthermore, CHH methylation seems to be less faithfully transmitted through vegetative propagation than CG and CHG methylation, likely due to the presence of a mirrored cytosine on the opposite strand in these latter contexts (23). This conclusion explain the extremely low linkage disequilibrium found in the CHH context: if meiosis acts as a reset for CHH methylation, in clonal species that lack this process, CHH methylation gets easily (and randomly) lost through mitosis (6). Another complementary explanation for the different results obtained in previous studies in *A. thaliana* could be the encompassed populations across diverse environments in our study, whereas previous studies worked under highly environmentally unified environments (18, 36). In fact,

environmental variation is recognized for modifying DNA methylation patterns especially in the CHH context. Therefore, it is likely that environmental fluctuations play a major role in shaping DNA methylation diversity patterns in nature, more than the reproductive strategies. This is consistent with the fact that while we found a high methylation divergence (Fst_{met}) for CHH methylation between low- and high-altitude populations in their natural conditions, this difference vanished once they were grown in a common garden. This is in line with several studies finding that the plastic response to environmental cues is stronger for CHH methylation than CG and CHG (37-38).

However, as much as fast epimutation rates can give a short-term adaptive advantage to a population, they can also hinder its adaptive potential if epimutations are too unstable/reversible (39). This might be the case for a large part of CHH epimutations in our study, since CHH methylation appeared to be highly variable, and CHH methylation divergence between populations of different altitudes vanished when grown in the same environment.

The CG and the CHG context drive epigenetic differentiation in natural populations

In our study, the individuals of each population appeared to be more structured by population for the CG and CHG contexts compared to the CHH context. This epigenetic structure was partially correlated with the genetic structure, suggesting that genomic factors partially explained it. In *A. thaliana*, only CG sites appeared to be structured by wild populations, and this structure was explained by genomic factors (27). Nevertheless, instead of a cause-consequence relationship, epigenetic and genetic structure could be correlated as a result of demography affecting both structures. Thus, CG and CHG methylation patterns may be an indicator of demographic history, implying their possible use in reconstructing phyloepigenetic trees (40).

We also found that the epigenetic structure in the three contexts appeared to be related to the altitude levels. Altitude is a broad environmental characteristic that can influence more specific climatic factors, such as temperature. For the CG methylated sites, the differentiation between populations was conserved across generations in the common garden conditions. On the contrary, the differentiation between populations decreased intensively in the common garden for the CHG context, regardless of the initial altitude levels of the populations in the field conditions. Nevertheless, the Fst_{met} value between populations at different altitudes was higher than the Fst_{met} value between populations at high altitudes under natural conditions. This pattern persisted even after the plants were transplanted to the common garden. This observation suggests that a substantial proportion of altitude-specific variants were inherited across clonal generations, which could potentially impact their adaptive potential. Our observations concur with multiple studies indicating that DNA methylation patterns can differ among populations inhabiting distinct natural environments (20, 41-42). Moreover, these variations are occasionally linked with ecological variables and are separate from genetic differences (43). Therefore, our findings underscore the potential role of epigenetic variation in shaping the adaptive responses of clonal species across diverse environmental gradients.

The selection of epigenomes and the interactions with the genomes

Similar to genetic mutations, epimutations can significantly impact traits under selection, especially in clonal plants, which display a high degree of transgenerational conservation of methylation profiles (5-6). Moreover, the survival of clonal or highly inbred plant populations could be highly dependent on methylation in the case of limited genetic diversity (44-45). However, to our

knowledge, no study attempted to detect traces of selection on epigenomes. This is partially explained by the bias for the "classical" neutral expectations found in genetics by the high epimutation rates (28-29, 31). In agreement with these assumptions, in our study, the high epimutation rates promoted the decrease of the Tajima's D for methylated sites compared to genomic data.

Thus, we proposed another method to find regions under selection, not based on comparison with neutral expectations but on comparison with global variation of the epigenomic diversity. We searched for *Fst* outliers among populations at different altitudes in order to detect regions under selection. High altitudes were montane to subalpine altitudes (1,436-1,905 m), which display environmental conditions not as extreme as alpine climate, and may thus represent a case of mild selection. In order to locate regions whose methylation was heritable, we compared the regions associated with altitude in the field with those associated in the garden. Regardless of the context considered, we found that the regions conserved across clonal generations were a minority among the regions associated with altitude in the field. The proportions of the candidate regions conserved in the common garden, which exhibited a significant *Fst* variation due to altitude in the field, were variable among methylation contexts. The CHH context presented the highest number of regions associated with significant *Fst* variation in the field. However, this context also presented the lowest proportion of heritable methylation profiles in the common garden (divided by a factor 24 and 12 compared with the CG and CHG context respectively). This suggests that methylation patterns in the CHH context are more susceptible to environmental influences compared to those in the CG and CHG contexts. However, they are less inherited across generations, making them potentially less significant for natural selection. Interestingly, the candidate regions for selection by altitude exhibited a significant increase in the Tajima's D for all the sequence contexts. This observation suggests that, even if the epimutation rate biassed this index for the neutral departure, selection on methylated sites could be detected by the variation of the Tajima's D across the epigenome of individuals in different populations or/and growth conditions.

Interestingly, among the 77 candidate regions under selection associated with altitude, 27% overlapped with a candidate region in another context. This observation suggests that an important part of the adaptation through methylation requires the intervention of different methylation profiles. Moreover, we observed that the genomic location of the candidate methylation profiles was context-dependent: most CG candidates were located within genes, whereas the majority of CHG and CHH candidates were located in promoters and TEs, respectively. These observations align with the location of differentially methylated regions across wild populations in *A. thaliana* (21). Interestingly, as in *A. thaliana*, we found enrichment in functions associated with transport and cytoskeleton, but, contrary to *A. thaliana*, this was observed in all sequence contexts (21). However, we did not find overlap between differentially expressed genes associated with altitude and candidate regions, which may be due to the relation between gene expression and methylation being still unclear and variable (21,46). Kawakatsu et al. (21) suggested that this relation is overestimated and dependent on the growth conditions. Moreover, methylation changes could influence gene expression in *trans*. In our study, we found some *trans* correlations between methylation of candidate regions and expression. These correlations were both positive and negative, indicating a complex relationship between methylation and gene expression.

An important theoretical point that remains debated is the possibility of selection on the epigenome independent of genomic variation. This aspect is especially critical for clonal species with low genetic

diversity (6). In our study, among the 77 heritable candidate regions associated with altitude, the majority were not associated with any genetic variants, regardless of the context considered. This observation suggests that selection could act on methylation profiles without the intervention of genetic mutations. It is worth noting that the limited number of associations observed between methylation of candidate regions and genetic variation may be attributed to the low genetic diversity within the studied populations, as indicated by the relatively low number of SNPs detected. Moreover, in this study we could not account for *de novo* TE insertions among the genetic variants. Given that these could be significant contributors to epigenetic variation, it is possible that we overlooked certain associations between genetic and DNA methylation variants. In this sense, we also found that the methylation of the majority of candidate regions was associated with genetic variants acting in *trans,* and these regions were overlapping with TEs. Methylation is particularly important for TE inactivation (5), especially in asexually reproducing populations. In these populations, TEs could quickly accumulate and lead to population extinction if not tightly repressed by DNA methylation (47-48).

In conclusion, our comprehensive study on *F. vesca* across an altitudinal gradient highlights the dynamic and complex nature of DNA methylation in natural populations, particularly emphasising the differential roles of methylation contexts (CG, CHG, and CHH) in epigenetic differentiation and potential adaptation. We observed that CHH methylation, being more variable and influenced by environmental conditions, may have a prominent role for epimutations in short-term responses, especially in clonally reproducing species. Our findings support the idea that epigenetic variation, particularly in non-CG contexts, can contribute to population differentiation and potentially to adaptation in changing environments, independent of genetic variation. This underscores the importance of considering epigenetic mechanisms in ecological and evolutionary studies, especially for species with limited genetic diversity or those reproducing primarily clonally. Furthermore, the interplay between genetic and epigenetic factors in shaping phenotypic plasticity and adaptive potential highlights the complexity of natural selection processes in wild populations. Our study adds to the growing body of evidence that epigenetic mechanisms play a crucial role in the adaptive capacity of plants, offering new insights into the evolutionary implications of epigenetic variation in natural environments. Speculatively, considering that many plant traits are controlled by multiple genes (polygenic control), and if epimutations are heritable, this could offer a powerful mechanism for generating phenotypic variation without necessitating changes in DNA sequence variation. This notion introduces the possibility of epigenetic mechanisms acting as a dynamic and reversible layer of regulation that can quickly respond to environmental changes, thereby providing populations with an additional tool for adaptation. Such epigenetic flexibility, when combined with the polygenic control of traits, could enhance the evolutionary potential of species by facilitating the exploration of phenotypic landscapes through mechanisms other than genetic mutation alone. This could be particularly beneficial in changing environments, where the ability to quickly adjust and adapt can be crucial for survival and reproductive success especially for populations with low standing genetic variation. Thus, the inheritance of epimutations opens up exciting avenues for understanding the generation of biodiversity and the adaptability of organisms in the face of environmental pressures.

Materials and methods

Study system

Fragaria vesca L. (*Rosaceae*) is a widespread herbaceous perennial species with a native range that includes Europe, northern Asia, North America, and northern Africa (49). It reproduces both clonally through stolons and sexually through seeds, although sexual reproduction is very rare in natural conditions (50). *F. vesca* is primarily a selfing species, but outcrossing is also possible (51-52).

Plant collection and growth

The dataset studied was composed of 7 populations of *F. vesca* from Italy (25), initially consisting of 21 natural populations from three European countries. The 7 populations selected represented the largest altitudinal range (Table S1). These populations were classified into two groups of altitudes, those situated at low altitudes (<1000m) and those situated at high altitudes (>1000m). With one exception, these populations also followed a temperature gradient, with low-altitude populations correlating with higher temperatures (Mean temperature >= 8°C), and vice versa (Table S1). The exception of one high-altitude population (FV_IT_02) that was associated with lower temperatures.

Briefly, Sammarco et al. (25) collected a fully developed leaf from 4 distinct individuals directly from field conditions from each of 7 populations of *F. vesca* (n = 28). They used these leaves for whole genome bisulfite sequencing (WGBS) analysis. They then dug up the same plants and planted them individually in the common garden of the Institute of Botany of the Czech Academy of Sciences in Průhonice, Czechia (Table S1). They let the plants propagate clonally for one year, and collected a fully developed leaf from offspring ramets of at least the third generation. These samples were also used for WGBS analysis (n = 49). We retained only the individuals grown in the garden as these were clones of the field ones and thus genetically identical. From a subset of 21 plants in the common garden (3 plants per population), they also collected a fully developed leaf from each individual plant for RNA-sequencing.

Genomic data

Single nucleotide polymorphisms (SNPs) from WGBS data were inferred following the methods described in Sammarco et al. (25). Specifically, the EpiDiverse SNP pipeline with default parameters (https://github.com/epidiverse/snp; 53) was used for each individual of all the populations grown under garden conditions (n=28). We then assessed the distribution of coverage for positions with a genotyping quality of 30 or higher to determine the mean coverage for 90% of the distribution (57 reads). Positions exceeding this mean coverage were excluded to minimise false SNP inference from repetitive elements. Finally, only sites covered by at least 3 reads in 80% of individuals from all populations were retained. The summary of the number of sites concerned is represented in Table S11.

Methylation calling

In our study, methylation calling was performed separately for the three sequence contexts (CG, CHG and CHH), following the methods described in Sammarco et al. (25). Specifically, the EpiDiverse WGBS pipeline (https://github.com/EpiDiverse/wgbs) was used for quality control, adaptor trimming, bisulfite reads mapping and methylation calling (53). We used the most recent version of the *F. vesca* genome (v4.0.a2) in the mapping step (54). All the cytosines having coverage ≥ 5 on the individual

bedGraph files of methylated positions for each sample and sequence context were retained. For mapping statistics and bisulfite non-conversion rate, please refer to Sammarco et al. (25).

Conversion of methylation profiles into SNP-like data

For each methylation context, the average methylation percentage by methylated position was calculated by our custom pipeline (https://github.com/leveveaudrey/methylation_population_analysis) across all the samples to determine the methylation thresholds for categorising cytosines as methylated or unmethylated. The methylation averages were of 52.9, 23.2 and 4.1% for the CG, CHG and CHH contexts, respectively (Table S2). The methylation thresholds were defined as the rounded values to the nearest whole number of these methylation averages, resulting in thresholds of 50%, 20%, and 4% for the CG, CHG, and CHH contexts, respectively. Based on these thresholds, the label "M" (indicating methylation) was assigned to sites exhibiting methylation percentages greater or equal to their corresponding threshold values. Conversely, the label "U" (indicating unmethylation) was assigned to cytosines exhibiting methylation percentages lower than their corresponding threshold value. Each individual was considered homozygous for either "U" or "M" alleles. The summary of the number of cytosines concerned is presented in Table S3. The methylation profiles of each individual sample were converted in vcf format for polymorphism, Fst and linkage disequilibrium analyses (see below).

Heritability of methylated profiles across clonal generations

To determine whether methylation profiles among populations persisted across three clonal generations in the common garden and estimate the heritability, we quantified the amount of methylated and unmethylated cytosines identical between the individual of the field and their respective clones grown in the common garden (n= 28) based on vcf files of each population in each condition.

Genetic and epigenetic structure

To construct kinship matrices for both methylation and genomic SNPs, we estimated the proportion of positions exhibiting distinct methylation or genomic profiles for each pairwise comparison of individuals across all the populations, and separately for field and garden conditions. When the methylation profile of one of the individuals for a given position was unknown, the site was ignored. The resulting degrees of kinship among individuals were represented through a heatmap plotted with the gplots package V3.1.3 (https://github.com/talgalili/gplots).

Polymorphism on SNP and methylated sites by population

For the SNP and the methylated sites on the three contexts, we estimated the π and the fixation index between each population pair (*Fst*) for all sites, the linkage disequilibrium (LD) between all sites in overlapping windows of 30 kb in all the populations and the Tajima's D on non-overlapping windows of 1 kb by population using vcftools v0.1.15 (56). The significance of differences between π, *Fst* and Tajima's D values for all pair of comparisons mentioned in this study were estimated by the Mann–Whitney–Wilcoxon test.

To assess the strength of the non-random association of alleles at different loci (linkage disequilibrium, LD), the coefficient of determination (R^2) was calculated separately for the field and garden populations . The R^2 values were calculated between variable methylated positions located in windows of 30kb and present in at least 80% of the individuals.

In population genomics, an increase in this index signifies a high level of differentiation between populations, which indicates directional selection. The Fst_{met} values were compared according to the altitude levels of the population pairs involved. Specifically, the Fst_{met} values obtained from population pairs from different altitudes, or from either low or high altitude were compared. These analyses were repeated for the populations grown in the common garden. Finally, the Fst_{met} values obtained from population pairs from the field and common garden were compared. We used this estimator to find our candidate regions associated with altitude (see below).

Another index typically used to detect selection in population genomics is the Tajima's D. A decrease in Tajima's D points to the prevalence of genetic variations that are either extremely common or extremely rare, suggesting directional selection. However, the neutral expectation for this index could be biassed for epialleles (28-29). We estimated this index to 1) verify if the Tajima's D is effectively biassed on epigenome and 2) estimate these variations for our candidate regions under selection.

Candidate regions associated with altitude

We defined the potential epigenomic regions associated with altitude as the regions of at least 140 bp (~size of a nucleosome), including a minimum of four sites separated by less than 1kb and with:

- Fst_{met} >0.2 for all pairwise comparisons between populations from different altitudes. This value of Fst was higher than the 4% of the *Fst* found between all pairs of populations in the same altitude levels in the field.

- Fst_{met} ≤0 for all pairwise comparisons between populations from either low or high altitude.

Between the potential epigenetic regions associated with altitude, the regions found in field and common garden conditions were considered as the candidate regions under selection, because they were inherited across generations even in the absence of altitudinal range in the common garden. On the other hand, the regions found only in the field were considered as candidate regions for local adaptation.

The overlaps of these regions with transposable elements (TEs) and genes were assessed using bedtools v2.26.0 (56), with the v4.0.a2 gene annotations downloaded from the Genome Database for Rosaceae (GDR; https://www.rosaceae.org/species/fragaria_vesca/genome_v4.0.a2; 57), and a TE annotation carried out using the EDTA annotation pipeline v1.9.6 (58) on the substituted genome using default parameters (59). We conducted a Gene Ontology (GO) enrichment analysis on the candidate genes associated with altitude selection. This analysis was performed using the R package clusterProfiler (version 3.18.1; 60). The results were considered statistically significant if they had a False Discovery Rate (FDR) adjusted P-value < 0.05.

Differential expression analysis

We analysed the RNA-sequencing data from Sammarco et al. (25) to identify differentially expressed genes (DEGs) between populations from low and high altitudes. We performed the analysis using the DESeq2 package for R (v1.30.1; 61). Genes were considered differentially expressed if they had an adjusted P-value < 0.05 (Benjamini-Hochberg) and an absolute value of fold change (FC) ≥ 1.5.

Correlation analysis between methylation of candidate regions associated with altitude selection and expression of DEGs

To evaluate the correlation between the methylation of candidate regions linked to altitude selection and the expression of DEGs, we used the multi-block discriminant analysis DIABLO (Data Integration Analysis for Biomarker discovery using Latent cOmponents; 62). For this analysis, we focused on the candidate regions common in field and garden conditions. We ran the analysis separately for each sequence context, using the R package mixOmics (version 6.24.0; 63). We removed predictors with zero or near-zero variance, and we fine-tuned the number of components to retain in the model using the DIABLO's perf function, with Mfold validation (n = 5) and cross-validation (nrepeat = 1000). The best performance was determined based on both the balanced error rate and centroid distances. We used methylation and gene expression of each individual sample (21 samples, 7 populations), and we used as predictive variables the altitude of the populations (high, low). We included all variables as these were part of a pre-selected group that showed differences in our predictive variables. This approach was feasible and efficient as the pre-selected group comprised a relatively low number of variables (30 DEGs, 42 candidate regions in CG, 12 in CHG and 31 in CHH). We built the final model using the block.splsda function designed to identify signatures of highly correlated variables across the multiple matrix sets (62). We visualised the correlations with the circosplot function.

Genome-wide association (GWA) analysis

To assess the putative genetic basis of altitude selection-associated candidate regions, we performed a genome-wide association (GWA) analysis. We conducted this analysis only for the candidate regions common in field and garden conditions. To increase the statistical robustness of the analysis, we included all the samples from Sammarco et al. (25), encompassing populations from distinct climatic and geographic origins grown under garden conditions (N = 147). Our rationale was that if a specific methylation pattern is under genetic control, this effect should be observable consistently across all populations. We conducted GWA analysis as outlined in Galanti et al. (64), using the R package rrBLUP (4.6.1; 65). For genetic variants, we imputed the missing genotype calls using BEAGLE 5.2 (66) and filtered for minor allele frequency (MAF) > 0.04. After filtering, we retained 83,095 SNPs. To correct for population structure, we used an identity by state (IBS) matrix derived from variants filtered for MAF ≥ 0.01. These were also pruned for LD with an LD threshold (r^2) of 0.8 for SNP pairs in a sliding window of 50 SNPs, sliding by 5. For phenotypes, we employed individual average methylation across each candidate region. This was calculated using the regionCounts R function from the methylKit package (version 1.16.1; 67), with a minimum cytosine coverage of 5. Finally, we selected the threshold P-value using Bonferroni correction to adjust for multiple comparisons.

Acknowledgements

We thank Frank Johannes and Stéphane Maury for carefully reading and improving the manuscript. ALV and CLP have been supported by the Czech Science Foundation (GACR 22-20240S). IS and VL have been supported by the Czech Science Foundation (GACR 23-04749S) and by institutional research project RVO 67985939. IS has also been supported by a Postdoctoral Fellowship (PPLZ) awarded by the Czech Academy of Science.

Computational resources were supplied by the project 'e-Infrastruktura CZ' (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Author contributions

AL, IS and CLP developed and designed the experiments for the study. IS and VL performed the experiments. AL and IS analysed and interpreted the data. AL, IS and CLP wrote the manuscript. All authors edited the manuscript.

Declaration of interests

The authors declare no competing interests.

Data sharing plans

The pipeline for the analysis of methylated sites developed in this paper is available on Github (https://github.com/leveveaudrey/methylation_population_analysis).

References

1. R. Radersma, D.W.A. Noble, T. Uller. Plasticity leaves a phenotypic signature during local adaptation. Evolution Letters 4, 360–370. (2020). https://doi.org/10.1002/evl3.185

2. A. Ashe, V. Colot, B.P. Oldroyd. How does epigenetics influence the course of evolution? Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20200111. (2021). https://doi.org/10.1098/rstb.2020.0111

3. A.D. Riggs, T.N. Porter. Overview of Epigenetic Mechanisms. Cold Spring Harbor Monograph Archive 32, 29–45. (1996). https://doi.org/10.1101/0.29-45

4. M. Lauria, V. Rossi. Epigenetic control of gene regulation in plants. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, Epigenetic control of cellular and developmental processes in plants 1809, 369–378. (2011). https://doi.org/10.1016/j.bbagrm.2011.03.002

5. M. Miryeganeh, H. Saze. Epigenetic inheritance and plant evolution. Population Ecology 62, 17–27. (2020).https://doi.org/10.1002/1438-390X.12018

6. K.J.F. Verhoeven, V. Preite. Epigenetic variation in asexually reproducing organisms. Evolution 68, 644–655. (2014). https://doi.org/10.1111/evo.12320

7. S. Feng, S.E. Jacobsen, W. Reik. Epigenetic reprogramming in plant and animal development. Science (New York, N.Y.) 330, 622. (2010). https://doi.org/10.1126/science.1190614

8. J.P. Calarco, F. Borges, M.T.A. Donoghue, F. Van Ex, P.E. Jullien, T. Lopes, R. Gardner, F. Berger, J.A. Feijó, J.D. Becker, R.A. Martienssen. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA.Cell 151, 194–205. (2012). https://doi.org/10.1016/j.cell.2012.09.001

9. C.A. Ibarra, X. Feng, V.K. Schoft, T.-F. Hsieh, R. Uzawa, J.A. Rodrigues, A. Zemach, N. Chumak, A. Machlicova, T. Nishimura, D. Rojas, R.L. Fischer, H. Tamaru, D. Zilberman. Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. Science 337, 1360–1364. (2012). https://doi.org/10.1126/science.1224839

10. A. Wibowo, C. Becker, G. Marconi, J. Durr, J. Price, J. Hagmann, R. Papareddy, H. Putra, J. Kageyama, J. Becker, D. Weigel, J. Gutierrez-Marcos. Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. eLife 5, e13546. (2016). https://doi.org/10.7554/eLife.13546

11. E.J. Finnegan, R.K. Genger, W.J. Peacock, E.S. Dennis. Dna Methylation in Plants. Annual Review of Plant Physiology and Plant Molecular Biology 49, 223–247. (1998). https://doi.org/10.1146/annurev.arplant.49.1.223

12. A. Zemach, D. Zilberman. Evolution of Eukaryotic DNA Methylation and the Pursuit of Safer Sex. Current Biology 20, R780–R785. (2010). https://doi.org/10.1016/j.cub.2010.07.007

13. X. Li, J. Zhu, F. Hu, S. Ge, M. Ye, H. Xiang, G. Zhang, X. Zheng, H. Zhang, S. Zhang, Q. Li, R. Luo, C. Yu, J. Yu, J. Sun, X. Zou, X. Cao, X. Xie, J. Wang, W. Wang. Single-base resolution maps of

cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics 13, 300. (2012). https://doi.org/10.1186/1471-2164-13-300

14. J.I. Gent, N.A. Ellis, L. Guo, A.E. Harkess, Y. Yao, X.Zhang, R.K. Dawe. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 23, 628–637. (2013). https://doi.org/10.1101/gr.146985.112

15. Z. Lang, Y. Wang, K. Tang, D. Tang, T. Datsenka, J.Cheng, Y. Zhang, A.K. Handa, J.-K. Zhu. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. Proceedings of the National Academy of Sciences 114, E4511–E4519. (2017). https://doi.org/10.1073/pnas.1705233114

16. M.S. Rajkumar, R. Shankar, R. Garg, M. Jain. Bisulphite sequencing reveals dynamic DNA methylation under desiccation and salinity stresses in rice cultivars. Genomics 112, 3537–3548. (2020). https://doi.org/10.1016/j.ygeno.2020.04.005

17. G.T. Martin, D.K. Seymour, B.S. Gaut. CHH Methylation Islands: A Nonconserved Feature of Grass Genomes That Is Positively Associated with Transposable Elements but Negatively Associated with Gene-Body Methylation. Genome Biology and Evolution 13, evab144. (2021). https://doi.org/10.1093/gbe/evab144

18. C. Becker, J. Hagmann, J. Müller, D. Koenig, O.Stegle, K. Borgwardt, D. Weigel. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480, 245–249. (2011). https://doi.org/10.1038/nature10555

19. E.J. Richards. Inherited epigenetic variation — revisiting soft inheritance. Nat Rev Genet 7, 395–401. (2006). https://doi.org/10.1038/nrg1834

20. M. Medrano, C.M. Herrera, P. Bazaga. Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb. Molecular Ecology 23, 4926–4938. (2014). https://doi.org/10.1111/mec.12911

21. T. Kawakatsu, S.C. Huang, F. Jupe, E. Sasaki, R.J. Schmitz, M.A. Urich, et al. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell 166, 492–505.(2016). https://doi.org/10.1016/j.cell.2016.06.044

22. F. Johannes, R.J. Schmitz. Spontaneous epimutations in plants. New Phytologist 221, 1253–1259. (2019). https://doi.org/10.1111/nph.15434

23. A. Wibowo, C. Becker, J. Durr, J. Price, S. Spaepen, S. Hilton, H. Putra, R. Papareddy, Q. Saintain, S. Harvey, G.D. Bending, P. Schulze-Lefert, D. Weigel, J. Gutierrez-Marcos. Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation. Proceedings of the National Academy of Sciences 115, E9145–E9152. (2018). https://doi.org/10.1073/pnas.1805371115

24. B.D. Rodríguez, D. Galanti, A. Nunn, C. Peña-Ponton, P. Pérez-Bello, I. Sammarco, K. Jandrasits, C. Becker, E.D. Paoli, K.J.F. Verhoeven, L. Opgenoorth, K. Heer. Epigenetic variation in the Lombardy poplar along climatic gradients is independent of genetic structure and persists across clonal reproduction. (2022). https://doi.org/10.1101/2022.11.17.516862

25. I. Sammarco, B. Díez Rodríguez, D. Galanti, A. Nunn, C. Becker, O. Bossdorf, Z. Münzbergová, V. Latzel. DNA methylation in the wild: epigenetic transgenerational inheritance can mediate adaptation in clones of wild strawberry (Fragaria vesca). New Phytologist. (2023). https://doi.org/10.1111/nph.19464

26. R.J. Schmitz, Y. He, O. Valdés-López, S.M. Khan, T. Joshi, M.A. Urich, J.R. Nery, B. Diers, D. Xu, G. Stacey, J.R. Ecker. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res. 23, 1663–1674. (2013). https://doi.org/10.1101/gr.152538.112

27. R.J. Schmitz, M.D. Schultz, M.A. Urich, J.R. Nery, M. Pelizzola, O. Libiger, A. Alix, R.B. McCosh, H. Chen, N.J. Schork, J.R. Ecker. Patterns of population epigenomic diversity. Nature 495, 193–198. (2013). https://doi.org/10.1038/nature11968

28. B. Charlesworth, K. Jain. Purifying Selection, Drift, and Reversible Mutation with Arbitrarily High Mutation Rates. Genetics 198, 1587–1602. (2014). https://doi.org/10.1534/genetics.114.167973

29. J. Wang, C. Fan. A Neutrality Test for Detecting Selection on DNA Methylation Using Single Methylation Polymorphism Frequency Spectrum. Genome Biology and Evolution 7, 154–171. (2015). https://doi.org/10.1093/gbe/

30. J.J. Vitti, S.R. Grossman, P.C. Sabeti. Detecting Natural Selection in Genomic Data. Annual Review of Genetics 47, 97–120. (2013). https://doi.org/10.1146/annurev-genet-111212-133526

31. A. van der Graaf, R. Wardenaar, D.A. Neumann, A. Taudt, R.G. Shaw, R.C. Jansen, R.J. Schmitz, M. Colomé-Tatché, F. Johannes. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proceedings of the National Academy of Sciences 112, 6676–6681. (2015). https://doi.org/10.1073/pnas.1424254112

32. I. Sammarco, Z. Münzbergová, V. Latzel. Response of Fragaria vesca to projected change in temperature, water availability and concentration of CO2 in the atmosphere. Sci Rep 13, 10678. (2023). https://doi.org/10.1038/s41598-023-37901-8

33. A.W. Schrey, M. Alvarez, C.M. Foust, H.J. Kilvitis, J.D. Lee, A.L. Liebl, L.B. Martin, C.L. Richards, M. Robertson. Ecological Epigenetics: Beyond MS-AFLP. Integrative and Comparative Biology 53, 340–350. (2013). https://doi.org/10.1093/icb/ict012

34. P.D. Sniegowski, P.J. Gerrish, T. Johnson, A. Shaver. The evolution of mutation rates: separating causes from consequences. BioEssays 22, 1057–1066. (2000). https://doi.org/10.1002/1521-1878(200012)22:12

35. Q. Qiao, P.P. Edger, L. Xue, L. Qiong, J. Lu, Y. Zhang, Q. Cao et al. Evolutionary history and pan-genome dynamics of strawberry (Fragaria spp.). Proc Natl Acad Sci U S A 118, e2105431118. (2021). https://doi.org/10.1073/pnas.2105431118

36. R.J. Schmitz, M.D. Schultz, M.G. Lewsey, R.C. O'Malley, M.A. Urich, O. Libiger, N.J. Schork ,J.R. Ecker. Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334, 369–373. (2011). https://doi.org/10.1126/science.1212959

37. M.R. Atighi, B. Verstraeten, T. De Meyer, T. Kyndt. Genome-wide DNA hypomethylation shapes nematode pattern-triggered immunity in plants. New Phytologist 227, 545–558. (2020). https://doi.org/10.1111/nph.16532

38. R.H. Dowen, M. Pelizzola, R.J. Schmitz, R. Lister, J.M. Dowen, J.R. Nery, J.E. Dixon, J.R. Ecker. Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci U S A 109, E2183-2191. (2012). https://doi.org/10.1073/pnas.1209329109

39. I. Kronholm, S. Collins. Epigenetic mutations can both help and hinder adaptive evolution. Molecular Ecology 25, 1856–1868. (2016). https://doi.org/10.1111/mec.13296

40. M.D. Sow, O. Rogier, I. Lesur, C. Daviaud, E. Mardoc, E. Sanou, L. Duvaux, P. Civan, A. Delaunay, M.-C.L.- Descauses, V. Benoit, I. Le-Jan, C. Buret, C. Besse, H.Durufle, R. Fichot, G. Le-Provost, E. Guichoux, C. Boury, A. Garnier, A. Senhaji-Rachik, V. Jorge, C. Ambroise, J. Tost, C. Plomion, V. Segura, S. Maury, J. Salse. Epigenetic Variation in Tree Evolution: a case study in black poplar (Populus nigra). (2023). https://doi.org/10.1101/2023.07.16.549253

41. C.F. Lira-Medeiros, C. Parisod, R.A. Fernandes, C.S. Mata, M.A. Cardoso, P.C.G. Ferreira. Epigenetic Variation in Mangrove Plants Occurring in Contrasting Natural Environment. PLOS ONE 5, e10326. (2010). https://doi.org/10.1371/journal.pone.0010326

42. C.L. Richards, A.W. Schrey, M. Pigliucci. Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. Ecology Letters 15, 1016–1025. (2012). https://doi.org/10.1111/j.1461-0248.2012.01824.x

43. B.D. Rodríguez, D. Galanti, A. Nunn, C. Peña-Ponton, P .Pérez-Bello, I. Sammarco, K. Jandrasits, C. Becker, E.D. Paoli, K.J.F. Verhoeven, L. Opgenoorth, K. Heer. Epigenetic variation in the Lombardy poplar along climatic gradients is independent of genetic structure and persists across clonal reproduction. (2022). https://doi.org/10.1101/2022.11.17.516862

44. V. Latzel, E. Allan, A. Bortolini Silveira, V. Colot,M. Fischer, O. Bossdorf. Epigenetic diversity increases the productivity and stability of plant populations. Nat Commun 4, 2875. (2013). https://doi.org/10.1038/ncomms3875

45. V. Latzel, A.P. Rendina González, J. Rosenthal. Epigenetic Memory as a Basis for Intelligent Behavior in Clonal Plants. Frontiers in Plant Science 7. (2016).

46. M.J. Dubin, P. Zhang, D. Meng, M.-S. Remigereau, E.J. Osborne, F. Paolo Casale, P. Drewe, A. Kahles, G. Jean, B. Vilhjálmsson, J. Jagoda, S. Irez, V. Voronin, Q. Song, Q. Long, G. Rätsch, O. Stegle, R.M. Clark, M. Nordborg. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. eLife 4, e05255. (2015). https://doi.org/10.7554/eLife.05255

47. B. Charlesworth, P. Sniegowski, W. Stephan. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215–220. (1994). https://doi.org/10.1038/371215a0

48. E.S. Dolgin, B. Charlesworth. The Fate of Transposable Elements in Asexual Populations. Genetics 174, 817–827. (2006). https://doi.org/10.1534/genetics.106.060434

49. G.M. Darrow. The Strawberry. The Strawberry. (1966).

50. B. Schulz, R.L. Eckstein, W. Durka. Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. Molecular Ecology 23, 3523–3537. (2014). https://doi.org/10.1111/mec.12835

51. J. Li, M.H. Koski, T.-L. Ashman. Functional characterization of gynodioecy in Fragaria vesca ssp. bracteata (Rosaceae). Annals of Botany 109, 545–552. (2012). https://doi.org/10.1093/aob/mcr279

52. H.S. Hilmarsson, T. Hytönen, S. Isobe, M. Göransson, T. Toivainen, J.H. Hallsson. Population genetic analysis of a global collection of Fragaria vesca using microsatellite markers. PLOS ONE 12, e0183384. (2017). https://doi.org/10.1371/journal.pone.0183384

53. A. Nunn, C. Otto, M. Fasold, P.F. Stadler, D. Langenberger. Manipulating base quality scores enables variant calling from bisulfite sequencing alignments using conventional bayesian approaches. BMC Genomics 23, 477. (2022). https://doi.org/10.1186/s12864-022-08691-6

54. Y. Li, M. Pi, Q. Gao, Z. Liu, C. Kang. Updated annotation of the wild strawberry Fragaria vesca V4 genome. Horticulture Research 6, 61. (2019). https://doi.org/10.1038/s41438-019-0142-6

55. P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. (2011). https://doi.org/10.1093/bioinformatics/btr330

56. A.R. Quinlan, I.M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. (2010). https://doi.org/10.1093/bioinformatics/btq033

57. S. Jung, T. Lee, C.-H. Cheng, K. Buble, P. Zheng, J. Yu, J. Humann, S.P. Ficklin, K. Gasic, K. Scott, M. Frank, S. Ru, H. Hough, K. Evans, C. Peace, M. Olmstead, L.W. DeVetter, J. McFerson, M. Coe, J.L. Wegrzyn, M.E. Staton, A.G. Abbott, D. Main. 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Research 47, D1137–D1145. (2019). https://doi.org/10.1093/nar/gky1000

58. S. Ou, W. Su, Y. Liao, K. Chougule, J.R.A. Agda, A.J. Hellinga, C.S.B. Lugo, T.A. Elliott, D. Ware, T. Peterson, N. Jiang, C.N. Hirsch, M.B. Hufford. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biology 20, 275. (2019). https://doi.org/10.1186/s13059-019-1905-y

59. M.-E. López, D. Roquis, C. Becker, B. Denoyes, E. Bucher. DNA methylation dynamics during stress response in woodland strawberry (Fragaria vesca). Horticulture Research 9, uhac174. (2022). https://doi.org/10.1093/hr/uhac174

60. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, 284–287. (2012). https://doi.org/10.1089/omi.2011.0118

61. M.I.Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. (2014). https://doi.org/10.1186/s13059-014-0550-8

62. A. Singh, C.P. Shannon, B. Gautier, F. Rohart, M. Vacher, S.J. Tebbutt, K.-A. Lê Cao. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics 35, 3055–3062. (2019). https://doi.org/10.1093/bioinformatics/bty1054

63. F. Rohart, B. Gautier, A. Singh, K.-A.L. Cao. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology 13, e1005752. (2017). https://doi.org/10.1371/journal.pcbi.1005752

64. D. Galanti, D. Ramos-Cruz, A. Nunn, I. Rodríguez-Arévalo, J.F. Scheepens, C. Becker, O. Bossdorf. Genetic and environmental drivers of large-scale epigenetic variation in Thlaspi arvense. PLOS Genetics 18, e1010452. (2022). https://doi.org/10.1371/journal.pgen.1010452

65. JB. Endelman. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4: 250–255. (2011).

66. B.L. Browning, Y. Zhou, S.R. Browning. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet 103, 338–348. (2018). https://doi.org/10.1016/j.ajhg.2018.07.015

67. A. Akalin, M. Kormaksson, S. Li, F.E. Garrett-Bakelman, M.E. Figueroa, A. Melnick, C.E. Mason. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology 13, R87. (2012). https://doi.org/10.1186/gb-2012-13-10-r87

Figure legend

Fig 1: Polymorphism observed in the CG, CHG and CHH methylation contexts. *Boxplot of the A) mean methylation average (%) and of B) mean methylation polymorphism (π_{met}). We distinguished the values obtained by individuals from the field (black) and common garden (grey). ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. The non-significant p-values (>0.05) were not specified. C) the histogram of distribution of the Tajima's D for the CG, CHG and CHH methylation contexts. The Tajima's D values were obtained by individuals from the field in non-overlapping windows of 1kb for the CG (black), CHG (dark grey), CHH (light grey) methylation context and for genetic SNPs (red). The vertical lines represent the median values obtained for the three methylation contexts and genetic SNPs. N = 28 plants for both field and garden conditions.*

Fig 2: Decrease of the linkage disequilibrium with the distance between two variable sites across all the methylation contexts. *Mean R^2 obtained depending on the distance (base) of the two variables sites for CG (black), CHG (dark grey), CHH (light grey) methylation context and for the genetic SNP (red). The values were obtained by individuals from the field. We considered only sites separated by less than 30,001 bases.*

Fig 3: Structuration by altitude of the field populations for the methylation and genetic profiles. *Kinship matrices between each individual from the field, based on A) CG, B) CHG, C) CHH methylation profiles and on D) the genotype (SNPs). The matrices are based on the estimation of the differentiation and are represented by a heatmap. The low normalised differentiations are represented in blue. The high normalised differentiations are represented in red. Hierarchical clustering are depicted at the top and on the left. The colours assigned to the names represent the different altitude clusters in the field condition, with red: > 1000m and blue: <1000m. N = 28 plants.*

Fig 4: The mean Fst_{met} is higher between populations from different altitudes for the CHH methylation contexts in populations from the field. *A) Boxplot of the mean Fst_{met} for each population pair in the CG, CHG and CHH methylation contexts. Boxplot of the mean Fst_{met} for each population pair found in low, high or different altitude levels for B) the CG, C) CHG and D) CHH methylation contexts. The values were obtained by individuals from the field (black) and from the common garden (grey). Each point represents one value obtained in one comparison between two populations. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. The non-significant p-values (>0.05) were not specified. N = 28 plants for both field and garden conditions.*

Fig 5: Gene Ontology (GO) enrichment analysis performed on the candidate genes associated with altitude selection that were conserved across both field and garden conditions. *The GeneRatio refers to the fraction of genes associated with each specific GO term. This ratio serves as an indicator of the frequency of each GO term within the considered gene set. The figure only includes GO terms that have an adjusted P-value < 0.05.*