Decoupled genetic and epigenetic variation in the montane endemic *Erodium cazorlanum* (Geraniaceae) and a widespread congener

Rubén Martín-Blázquez[1*], Mónica Medrano[1], Pilar Bazaga[1], Francisco Balao[2], Ovidiu Paun[3], Conchita Alonso[1*]

1. Estación Biológica de Doñana, CSIC, Avenida Américo Vespucio 26, Sevilla 41092, Spain.

2. Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Apdo. 1095, 41080 Sevilla, Spain.

3. Department of Botany and Biodiversity Research, University of Vienna, 1030 Vienna, Austria.

* Corresponding authors: rmartinblazquez@gmail.com and conalo@ebd.csic.es

Abstract

Epigenetic states offer an additional layer of variation besides genetic polymorphism that contribute to phenotypic variation and may arise either randomly or in response to environmental factors. We hypothesize that closely related species with different life-histories and habitat requirements could show distinct patterns of intraspecific epigenetic variation. We used Restriction-site Associated DNA sequencing (RADseq) and its bisulfite-converted variant (bsRADseq) to investigate in Sierra de Cazorla (SE Spain) the genetic and epigenetic population structure of two congeneric species, the endemic woody perennial *Erodium cazorlanum* and the widely distributed annual herb *E. cicutarium*. Population genomics analyses revealed no structure in either *E. cazorlanum* and *E. cicutarium,* suggesting substantial gene flow between the study populations. In contrast, we found that the mean proportion of global DNA methylation was different between populations and species, with *E. cazorlanum* DNA showing higher methylation average and across-individuals variation. For each species, we searched for the loci with the strongest epigenetic differentiation between populations (differentially methylated cytosines, DMCs), and summarized them across regions (differentially methylated regions, DMRs). Multivariate analysis and hierarchical clustering of single cytosines' methylation percentages did not group individuals by population, pointing to high epigenetic variation within populations. DMRs were enriched in *Copia* transposable elements, putatively associated with stress response in plants. Our results suggest that variation at both genetic and epigenetic levels in our study area occur mainly within natural populations of *E. cazorlanum* and *E. cicutarium,* with stronger population structure in *E. cicutarium,* and highlight the relevance of analyzing short distance spatial patterns.

39 **Keywords:** plant population epigenomics, bsRADseq, DNA methylation, epigenetic
40 variation, cytosine differential methylation

41 **1. Introduction**

42 Understanding the sources of natural phenotypic variation is a crucial question in ecology
43 and evolution. The idea that epigenetic diversity can compensate for reduced genetic
44 diversity, promoting phenotypic variation through trait plasticity in response to
45 environmental changes is relatively new (Balao et al., 2018; Bossdorf et al. 2008;
46 Richards et al., 2017; Richards, 2006; Williams et al., 2023). Trait plasticity can be
47 particularly advantageous for sessile organisms such as plants, with limited escape
48 strategies to face environmental perturbations and, thus, the magnitude and population
49 structure of genetic and epigenetic diversities in plant species with disparate genomic and
50 ecological features warrants comprehensive investigation (Anderson & Gezon, 2015;
51 Herrera et al.,2022; Richards et al., 2017; Szukala et al., 2023). Such knowledge can help
52 to optimize decisions in conservation (i.e. accurately evaluating the adaptive potential of
53 populations, Lamka et al., 2022; Williams et al., 2023), or even predict how plant species
54 would fare when introduced in foreign ecosystems (i.e., assess invasive potential, Hawes
55 et al., 2018).

56 DNA methylation remains a focus of interest among the spectrum of plant epigenetic
57 modifications. The mechanistic control and effects of DNA methylation, its relationship
58 with other chromatin regulators such as histone tail modification and small RNAs, and
59 the potential transgenerational transmission of different methylation-marker types have
60 been mainly studied in model species such as *Arabidopsis thaliana* (Cokus et al., 2008;
61 Muyle et al., 2022; Stroud et al., 2014), and economically important species such as maize
62 and rice (reviewed by Varotto et al., 2022; but see also Herrera et al., 2018). Several
63 DNA-methyltransferases set methyl groups in the carbon-5' of cytosines followed by
64 specific nucleotide motifs in the genome, referred to as methylation contexts. In plants,
65 there are three cytosine methylation contexts: CpG, CHG and CHH (where H can be any

66 nucleotide but G), which are each regulated by the action of distinct enzyme complexes,
67 and appear unevenly represented in plant genomes (Law & Jacobsen, 2010). Symmetrical
68 methylation in CpGs can be preferentially found in promoters and exons within genes
69 (referred to as gene body methylation; Bewick & Schmitz, 2017) and can be more
70 frequently heritable compared to other contexts (Schmitz et al., 2011). In turn, methylated
71 CHGs and CHHs are more abundant in heterochromatic and intergenic regions, and are
72 often reset between generations or rapidly (de)methylated in response to changes in
73 environmental conditions (Bouyer et al., 2017; Du et al., 2012). Furthermore,
74 transposable elements (TEs) show DNA methylation in all three contexts, resulting in TE
75 silencing and promoting genome stability (Matzke & Mosher, 2014). Given the wide
76 range of genome sizes in plants, and the positive relationship between genome size and
77 TE proportion in the genome (Kidwell, 2002), plant species show a broad variation in the
78 level of DNA methylation, that was found to be related to phylogenetic relationships and
79 species genome size (Alonsoet al., 2019a; Alonso et al., 2015; Vidalis et al., 2016).

80 The genomic and ecological features that could be associated to the broad variation in
81 levels of DNA methylation mentioned above are still poorly understood. Particularly for
82 non-model plant species, contrasting evidences have been found regarding their variation
83 across natural populations and geographic-climatic gradients (see e. g. De Kort et al.,
84 2020; Galanti et al., 2022; Lamka et al., 2022; Richards et al., 2017; Richards et al., 2010).
85 In particular, global DNA methylation tends to be lower in woody plants than perennial
86 herbs, as well as in genomes of widespread species compared to species with restricted
87 geographic distributions (Alonso et al., 2019a). Relevance of such life history traits when
88 controlling for phylogenetic relatedness remains largely unexplored and, thus, studies
89 focused on analyzing natural variation in DNA methylation of related species with
90 contrasting ecological features should be particularly valuable to better understand

epigenomic similarities and divergences within and across plant species (Medrano et al., 2020 and references therein; Noshay & Springer, 2021). Reduced representation bisulfite DNA sequencing (RRBS) allows characterization of DNA methylation from a reduced portion of the genome and can be applied to species even when no reference genome is available, as it is usually the case for plants with narrow distribution ranges (Paun et al., 2019). Despite limitations, RRBS can provide quantitative methylation information at the different sequence contexts (CpG, CHG, CHH), together with an estimate of methylation similarity across individuals and differentiation between populations and species, and may lead to some functional characterization of methylated regions when an annotated reference genome of close relatives and/or complementary RNAseq data is available (Chen et al., 2015; Hsu et al., 2017; Kreutz et al., 2020; Nunn et al., 2021; Paun et al., 2019).

In this study, we explore the DNA methylation variation in natural plant populations of two *Erodium* species co-occurring in Cazorla mountain ranges (South East Spain): the endemic woody perennial *E. cazorlanum*, and the widespread annual herb *E. cicutarium*. Given the differences in habitat specialization, distribution range and life history of these two congeneric species, we predict they could differ in their genome-wide DNA methylation level and in the way epigenetic variation is structured within and among populations throughout a geographical area where both species co-exist. In particular, we predict that (1) individuals of the endemic *E. cazorlanum* will exhibit higher diversity in genomic methylation levels than genetic variants, and (2) that epigenetic variability should be less structured in space in the endemic species compared to its widespread congener given the expected reduced range in suitable environmental conditions for the former. To address so, we generated Restriction site-Associated DNA sequencing (RADseq) data, both untreated and treated with bisulfite (bsRADseq, Trucchi et al.,

116 2016), from leaf DNA of the two *Erodium* species, each collected from three different
117 localities. First, we checked for population genetic structure in each species using the
118 single nucleotide polymorphisms (SNPs) obtained from RADseq data. We then estimated
119 the percentage of DNA methylation per cytosine in the three methylation contexts to
120 search for quantitative and qualitative differences in genome-wide DNA methylation at
121 the species and locality levels using the bsRADseq data. Next, we identified differentially
122 methylated cytosines (DMCs), used them to estimate population epigenetic structure in
123 each species, and summarized them in differentially methylated regions (DMRs) to
124 characterize genomic loci potentially under divergent epigenetic regulation, at locality
125 and species level. Challenges derived from using a single incomplete reference genome
126 to analyze both species are discussed at the end.

127 **2. Materials and methods**

128 **2.1. Study species and sample collection**

129 *Erodium cazorlanum* (Heywood) is an endemic, long-lived woody perennial plant,
130 typically associated to a few dolomite outcrops in Sierra de Cazorla-Segura-Las Villas
131 Natural Park (Jaén, SE Spain) (Alonso & García-Sevilla, 2013). *Erodium cicutarium* (L.)
132 L'Hér. is a widespread, herbaceous annual plant species with a fast life-cycle, frequently
133 associated with ruderal habitats, and to more developed soils within the same study area
134 (Fiz-Palacios *et al.,* 2010). The genome size of *E. cicutarium* has been estimated to range
135 from 1.2 to 2.7 Gbps (Pustahija et al., 2013; Zonneveld, 2019), different ploidy levels
136 have been reported from different geographic regions, ranging from diploid to tetraploid
137 (Pustahija et al., 2013; Zonneveld, 2019), and preliminary data suggest our plants have
138 2n = 40 chromosomes. Genomic information about *E. cazorlanum* is scarce, but its
139 estimated genome size is 8.1 Gbps (Alonso *et al.* unpublished data), it is described as an

140 octoploid (Alonso & García-Sevilla, 2013; Blanca et al., 2009) and preliminary data
141 suggest our plants have 2n = 80 chromosomes.

142 Young fully-developed leaves from 18 *E. cazorlanum* and 18 *E. cicutarium* individuals
143 were collected at flowering within three localities per species in Sierra de Cazorla-Segura-
144 Las Villas Natural Park (Jaén, Spain, **Figure 1**). The number of individuals per locality
145 was selected based on available information on minimum sample sizes required for
146 accurate estimates of intra- and inter-population genetic diversity based on high-
147 throughput DNA sequencing (Nazareno et al., 2017). The study localities (hereafter
148 populations) were Fuente Bermejo (FB), Puerto Lézar (PL), and Puerto del Tejo (PT) for
149 *E. cazorlanum*; and Nava Correhuelas (CH), Fuente Bermejo (FB), and Puerto del Tejo
150 (PT) for *E. cicutarium*. Horizontal distances between the nearest and farthest sampling
151 sites were 7.5 and 19.4 km for *E. cazorlanum* (FB-PL and PL-PT, respectively), and 2.1
152 and 11.5 km for *E. cicutarium* (CH-FB and FB-PT, respectively). Note that despite
153 sharing some sampling localities, the two species were located in separated patches due
154 to differences in soil requirements (Moreno, 2008). Sampling details are thoroughly
155 described in (Medrano et al., 2020).

156 **2.2. DNA extraction, library preparation, sequencing and quality check**

157 Total genomic DNA was extracted using Bioline ISOLATE II Plant DNA Kit and
158 quantified using a Qubit fluorometer 2.0 (Thermo Fisher Scientific, Waltham, MA, USA).
159 The RADseq library preparation followed the general steps from (Balao et al., 2023) (see
160 **Supplementary Materials & Methods** for details). Half the DNA yield per sample was
161 treated with bisulfite to generate bsRADseq libraries (Trucchi et al., 2016), the other half
162 remained untreated for generating RADseq libraries.

163 Both untreated and bisulfite-treated DNA samples were sequenced at the VBCF Vienna
164 (https://www.viennabiocenter.org/vbcf/next-generation-sequencing/) as paired-end 2 x
165 150 bp reads, a single lane of an Illumina HiSeq 2500 machine was used for the untreated
166 samples of the two species (RADseq) and another lane was used for the bisulfite-treated
167 samples (bsRADseq). Reads are accessible through the sequence read archive (SRA)
168 from NCBI, under BioProject ID PRJNA1071520. The reads were demultiplexed with
169 deML (Renaud et al., 2015), and then STACKS v2.41 (Rochette & Catchen, 2017),
170 discarding reads with low quality scores according to default parameters. Demultiplexed
171 reads were trimmed and cleansed from Illumina adapter sequences with Trimmomatic
172 v0.39 (Bolger et al., 2014) using the options '*TrimmomaticPE -phred33*
173 *ILLUMINACLIP:adapterList:2:30:10 LEADING:28 TRAILING:28*
174 *SLIDINGWINDOW:4:15 MINLEN:30*'.

175 **2.3. Population genetics analysis**

176 A draft genome of *E. cicutarium* based on PacBio single-molecule long-read sequencing
177 and Illumina short-reads sequencing was assembled and annotated (see **Supplementary**
178 **Materials & Methods** for the details). The draft genome v2 is available at NCBI
179 Genome, under accession JATABU000000000; PacBio reads are available at NCBI SRA,
180 under BioProject ID PRJNA984161. Trimmed reads from the RADseq libraries were
181 aligned to the *E. cicutarium* draft genome v2 using *bowtie2* v2.2.5 (Langmead &
182 Salzberg, 2012) with the '*--very-sensitive*' option. Mapped reads were analyzed with the
183 STACKS script *ref_map.pl*, and then single nucleotide polymorphisms (SNPs) were
184 filtered using STACKS binary *populations* with options '*-r 0.83 --min_maf 0.05 --*
185 *max_obs_het 0.7 --vcf --write_single_snp*'. For each species separately, remaining SNPs
186 were analyzed with PLINK v1.90b7 (Purcell et al., 2007) to perform Principal
187 Component Analysis (PCA) to summarize the genomic variation. Further, each SNP

188 dataset was analyzed with ADMIXTURE v1.3.0 (Alexander & Lange, 2011) to test the
189 population genetic structure of each species, testing for K values from 1 to 6, choosing
190 the best fit for the K value with the lowest cross-validation error. Pairwise genetic distance
191 matrices between individuals within a species were generated from the SNP dataset using
192 the *distance* function of PLINK, that calculates pairwise identity-by-state (IBS) distances
193 based on allele counts from alleles shared for each pairwise comparison. As an alternative
194 approach, additional analyses for SNP calling and population structure were performed
195 with different model assumptions as implemented in the genotype likelihood-based
196 methods of *ANGSD* (Korneliussen et al., 2014) and *NGSadmix* (Skotte et al., 2013) (see
197 **Supplementary Materials & Methods** for details). Despite being developed for diploid
198 organisms, ANGSD was suggested as a powerful approach to disentangle genetic patterns
199 in heteroploid datasets (Zaveska et al., 2019). Naturally occurring C/T polymorphic sites
200 found in any of the two RADseq dataset were masked in the *E. cicutarium* genome for
201 downstream analyses.

202 **2.4. Alignment of bisulfite converted reads to genome**

203 Trimmed reads from the bsRADseq were aligned to the same *E. cicutarium* draft genome
204 using *bismark* v0.23.1 (Krueger & Andrews, 2011), with *Bowtie2* as base aligner
205 (Langmead & Salzberg, 2012), using the following options: *'--bowtie2 --non_directional*
206 *--most_valid_alignments 2 --L 32 --D 10 --R 1 --score_min L,0,0.2'*. For each individual,
207 we checked the summary report recording the mapping efficiency, the number of
208 cytosines screened, and the distribution among the different contexts (CpG, CHG, and
209 CHH). The next step was to extract the methylation information of each cytosine position
210 using *bismark_methylation_extractor*, with the options *'--no_overlap --ignore 4 --report*
211 *--bedGraph --counts --CX'*. As in each read pair, the read corresponding to the sonicated
212 end incorporates an unmethylated cytosine in its first position during library construction,

213 which then is exposed to bisulfite treatment, bisulfite conversion rate (BCR) was
214 calculated for each sample using the formula $100 * T / (T + C)$, where T and C are the
215 respective percentage of thymines and cytosines in the first position of the 2nd paired read
216 (Krueger, 2013).

217 **2.5. Global DNA methylation analysis**

218 Unless otherwise stated, all statistical analyses were performed using R v4.1.2 (R Core
219 Team, 2023). The number of methylated cytosines and total number of genomic positions
220 analyzed per individual was extracted from *bismark* reports with a custom R script.
221 Global effects associated to species and population divergence on global genome
222 methylation levels were analyzed with a GLM with quasibinomial distribution and logit
223 link, using *emmeans* (Lenth, 2023) and *lme4* (Bates et al., 2015) R packages. The
224 dependent variable was the overall proportion of methylated to unmethylated cytosines
225 per individual, and the formula included species as fixed factor and populations of each
226 separate species as a nested fixed factor (formula: proportion of methylated cytosines ~
227 species + species : population). The same process was done for each methylation context
228 separately (CpG, CHG, CHH).

229 **2.6. Differential methylation analysis**

230 Detection of differentially methylated cytosines (DMCs) between populations was
231 conducted with *DSS* v2.46.0 R package (Feng et al., 2014; Park & Wu, 2016). *DSS* is
232 based on a beta-binomial regression model with arcsine link function, and it takes
233 advantage of shrinkage estimation of the dispersion parameter based on a Bayesian
234 hierarchical model to reduce the dependence of variance on mean. P-values were adjusted
235 for multiple testing using the Benjamini and Hochberg false discovery rate (FDR)
236 procedure, retaining those with FDR < 0.05 (Benjamini & Hochberg, 1995). Analyses

were conducted for each species separately. Although *DSS* allowed fitting a model with a single factor with three populations as levels, handling RRBS data reduces the probability of finding single cytosine positions well covered for all three populations (i.e., a position showing acceptable coverage for two populations, but not covered in the remaining population), which would discard potential DMCs due to insufficient coverage. Thus, differences in cytosine methylation between pairs of populations were calculated fitting the data to one-factor models consisting in pairs of populations, generating six datasets (FB x PL, FB x PT and PL x PT for *E. cazorlanum*; CH x FB, CH x PT and FB x PT for *E. cicutarium*). A position was considered a DMC when its FDR between populations was < 0.05, and the difference in methylation was > 20 % between any of the paired-population comparisons. This approach implies that the definition of hyper- and hypo-methylation is relative: a single DMC can be considered as hyper-methylated in one or more populations, and at the same time hypo-methylated in other population(s). Number of overlapping DMCs (either hyper- or hypo-methylated DMCs) from two populations were analyzed with a hypergeometric test to investigate whether overlaps were greater than expected by chance.

Hierarchical clustering and visualization based on methylation percentage per DMC was calculated and plotted using *pheatmap* R package (Kolde 2019). The variation observed in percentage of methylation across DMCs was also analyzed with non-metric multidimensional scaling (NMDS) in *vegan* R package (Oksanen et al., 2022), with 100 random starting configurations and choosing the Manhattan distance method. Sets of DMCs which showed 100 % methylation in a single individual and methylation percentages close to zero in the rest of the individuals were named “drastic” DMCs, the rest of DMCs were named “gradual” DMCs. Hierarchical clustering and NMDS analyses were performed again with only “gradual” DMCs.

262 Epigenetic Euclidean distance matrices between individuals within a species were
263 calculated from the methylation percentage matrix of gradual DMCs using *vegan* R
264 package (Oksanen et al., 2022). The complete DMC set and only "gradual" DMC set
265 epigenetic matrices were compared with previously calculated genetic distance matrices
266 with a Mantel test, to test for intra-specific Spearman correlation between genetic and
267 epigenetic distance, with 9,999 permutations, using *vegan* R package. Additional
268 Mantel's tests to check the relationship between genetic and epigenetic distances were
269 also performed on each separated methylation context (CpG, CHG and CHH).

270 Differentially methylated regions (DMRs) between populations were detected using
271 *callDMRs()* function from the *DSS* package with default parameters. Methylation site
272 density (MSD) per DMR was calculated by dividing the number of DMCs found in the
273 DMR by the total length in base pairs of the respective DMR. After observing the MSD
274 frequency distribution, percentile 90 of each population comparison's MSD distribution
275 was calculated and used as threshold to filter potential false positive DMRs (i. e. long
276 DMRs with a relatively small number of DMCs). A DMR was retained when three or
277 more DMCs were present within a genomic region with a minimum length of 50 bps and
278 when its MSD value was higher than the percentile 90 of the MSD distribution for the
279 analyzed comparison. After considering the biological differences of CpG, CHG and
280 CHH methylation contexts, we opted to call DMRs as groups of DMCs from the same
281 methylation context (CpG, CHG and CHH).

282 **2.7. Inter-species comparison**

283 The mean percentage of uniquely mapped bsRADseq reads to *E. cicutarium* draft
284 reference genome was 19.00 % [13.6 % - 25.6 %] for *E. cazorlanum* samples, and 58.40
285 % [47.2 % - 72.6 %] for *E. cicutarium* samples (**Supplementary Table S1**), which
286 showed statistically significant differences (ANOVA F statistic = 515.91, df = 1, P-value

287 < 0.001). Given the different mapping rates, biases in differential methylation analysis
288 were expected when cytosine and methylated cytosine coverages were compared between
289 species. Thus, a normalization factor was applied to *E. cicutarium* read counts before
290 comparing methylation levels between species. The normalization factor was calculated
291 as the mean mapping efficiency from *E. cazorlanum* samples divided by the mean
292 mapping efficiency for *E. cicutarium*, and *E. cicutarium* read counts were normalized by
293 multiplying both the coverage and methylated cytosines counts per locus by the obtained
294 normalization factor. Raw read counts in *E. cazorlanum* and normalized read counts in *E.*
295 *cicutarium* were used for DMC and DMR calling analyses in the interspecies comparison.
296 Priors for considering a DMC and a DMR were the same as the ones explained in previous
297 sections above.

298 **2.8. Functional characterization of differentially methylated regions**

299 DMCs included in DMRs that overlapped either with exons, introns, 5' UTRs, 3' UTRs
300 or intergenic regions as described in the *E. cicutarium* draft genome were identified with
301 custom scripts. DMCs that overlapped with repeated elements from the *E. cicutarium*
302 genome, including transposable elements (TEs), were retrieved with the *CNEr* R package
303 (Tan et al., 2019). Fisher's exact test was used to test whether the proportion of TEs
304 overlapping with the *E. cicutarium* set of DMCs/DMRs was over- or under-represented
305 compared to the proportion of all the TEs present in the *E. cicutarium* draft genome.

306 GO terms were extracted from the gene annotation file of the *E. cicutarium* draft genome
307 with a custom script. Lists of DMRs overlapping with annotated genes were then analyzed
308 with *topGO* (Alexa & Rahnenfuhrer, 2023) to find enriched GO terms in each DMR list,
309 using a weighted Fisher's exact test and correcting the p-value with FDR. A GO term was
310 considered enriched/over-represented in a set when the FDR < 0.05.

311 **3. Results**

312 **3.1. Sequencing quality and bisulfite treatment efficiency.**

313 For the 36 samples, Illumina sequencing produced a total of 332.2 million 2 x 150 bp
314 length paired-end reads for RADseq (ranging from 3.4 to 12.2 million per sample in *E.*
315 *cazorlanum*, and from 4.4 to 33.8 million per sample in *E. cicutarium*), and a total of
316 323.4 million 2 x 150 bp length paired-end reads for bsRADseq (ranging from 3 to 13.6
317 million per sample in *E. cazorlanum*, and from 3.8 to 30 million per sample in *E.*
318 *cicutarium*). After removing adapters and checking the distribution of sequencing quality
319 per nucleotide, 315.3 million reads were retained in the RADseq, and 302.4 million reads
320 remained in the bsRADseq. Mean bisulfite conversion rate per sample was 99.09 %
321 [99.07 % - 99.13 %] (**Supplementary Table S1**).

322 **3.2. Genetic polymorphism and population structure**

323 A total of 24,977 SNPs in *E. cazorlanum* and 31,696 SNPs in *E. cicutarium* passed the
324 filtering process. In the endemic *E. cazorlanum*, the PCA shows clustering by sampling
325 locality mainly on PC1, but the position of the groups did not follow a geographical
326 pattern, with some individuals from different localities positioned very close together in
327 the multidimensional PCA. Individuals from the FB population appeared more clearly
328 separated from the other two populations, although all principal components explained
329 less than 10 % of the variance on their own (**Figure 1**). ADMIXTURE considered *E.*
330 *cazorlanum* as a single genetic entity (K = 1) based on minimizing the cross-validation
331 error, whereas higher numbers of potential populations did not differentiate between the
332 different localities. For the widespread *E. cicutarium*, the PCA shows three separated
333 groups: one with all the individuals from CH and FB populations, another with four
334 individuals from PT population, and the third one with two individuals from PT

335 population (**Figure 1**); the difference between the first two groups is supported by PC1
336 (20.3 % of explained variance) and the difference between the first and third group is
337 described by PC2 (13.5 % of explained variance). ADMIXTURE again considered *E.*
338 *cicutarium* individuals as a single genetic entity (K = 1) based on the cross validation
339 error (**Supplementary Figure S1**). **Figure 1** shows assignment of *E. cazorlanum* and *E.*
340 *cicutarium* individuals to subpopulations, assuming models with two to six
341 subpopulations. **Supplementary Table S2** shows the cross-validation error values.
342 Alternative analyses performed with the genotype likelihood-based method of *ANGSD*
343 showed consistent results with the ones shown above (see **Supplementary Materials &**
344 **Methods** for details).

345 **3.3. Global DNA methylation analysis.**

346 *Bismark* analyzed a total of 11,921,613 cytosines in any methylation context (1.49 % of
347 *E. cicutarium* draft genome total length). Percentages of global DNA methylation ranged
348 from 6.66 % to 12.71 % in the endemic *E. cazorlanum* and from 7.76 % to 11.07 % in the
349 widespread *E. cicutarium* (**Supplementary Table S1**). Within species variance of global
350 methylation percentages was higher for *E. cazorlanum* (**Table 1**). **Figure 2** shows the
351 methylated cytosine percentages per sequence context by species and by population. On
352 average, there were significant differences between the two species in the ratio of
353 methylated to unmethylated cytosines per individual (**Table 2**) with the genomes of *E.*
354 *cazorlanum* being more methylated (mean ± SD = 9.51 ± 1.83) compared to *E. cicutarium*
355 (mean ± SD = 9.08 ± 0.94). Similar results were also obtained for each of the three
356 methylation contexts (**Table 2**, **Supplementary Figure S2**). Differences in percentages
357 of global DNA methylation were also statistically significant between the study
358 populations within a species, as were all pairwise within-species population contrasts

359 (**Table 2**, **Supplementary Figure S2**). Further, within species variance was always
360 higher for *E. cazorlanum* global and context specific methylation percentages (**Table 1**).

361 **3.4. Differential cytosine methylation analysis within species.**

362 A total of 12,795 and 16,131 naturally occurring C/T polymorphic sites obtained from
363 RADseq analyses were masked in bsRADseq data before analyses in *E. cazorlanum* (0.11
364 % of analyzed Cs) and *E. cicutarium* (0.14 % of analyzed Cs), respectively. For *E.*
365 *cazorlanum*, a total of 13,643 DMCs (5,632 CpG, 2,605 CHG, 5,406 CHH) were detected
366 between the three populations (ca. 0.11 % of total analyzed Cs), and 47,615 DMCs (ca.
367 0.40 % of total analyzed Cs) were detected between the three populations of *E. cicutarium*
368 (20,670 CpG, 8,278 CHG, 18,667 CHH; **Supplementary Figure S3**, **Supplementary**
369 **Tables S3, S4, S5**). In each of the two species, the number of overlapping hyper- or hypo-
370 methylated DMCs between each population pair was never significantly greater than
371 expected by chance (hypergeometric test, P-value = 1, **Supplementary Figures S4, S5**).
372 The overlap of DMCs between species was also not greater than predicted by chance
373 either (hypergeometric test, P-value = 1, **Supplementary Figure S6**).

374 Hierarchical clustering of DMCs showed that most clusters were defined by cytosines
375 with "drastic" changes in methylation percentage within the study dataset (i. e., 100 %
376 methylation for one individual and ca. 0 % methylation for the rest, **Supplementary**
377 **Figure S7**). Removing these "drastic" DMCs and leaving only those with "gradual"
378 change in methylation percentage still did not group samples following a clear population
379 structure in any of the two species (**Figure 3, Supplementary Figure S7**). When
380 analyzing the "gradual" DMCs by separated methylation context, the CpG context DMCs
381 showed clusters with at least three samples from the same population for *E. cazorlanum*,
382 and more patently for *E. cicutarium*; while CHG and CHH contexts did not group more

383 than three samples from the same population within any cluster (**Supplementary Figures**
384 **S8, S9**).

385 Multivariate analysis based on all DMCs showed no clear population segregation in *E.*
386 *cazorlanum* and a slight differentiation of the PT population for *E. cicutarium* (**Figure**
387 **3**). The same analysis, but performed only with gradual DMCs, marginally increased the
388 differentiation of PT population from FB and CH in *E. cicutarium*, while not affecting *E.*
389 *cazorlanum* grouping pattern (**Figure 3**). Performing multivariate analyses with DMCs
390 separated by context did not improve segregation of populations (**Supplementary**
391 **Figures S10, S11**).

392 Mantel tests revealed no statistically significant correlation between pairwise genetic and
393 epigenetic distances across individuals in any of the two study species (*E. cazorlanum*:
394 Mantel's R = 0.104, p-value = 0.271; *E. cicutarium*: Mantel's R = 0.003, p-value = 0.471).
395 When considering epigenetic distances exclusively for each separate methylation context,
396 Mantel tests also showed no statistical significance (p-value > 0.05).

397 **3.5. Genomic and functional characterization of DMRs.**

398 Only DMRs with high density of DMCs (defined as DMRs above the percentile 90 of the
399 MSD distribution; **Supplementary Figure S12, Supplementary Table S6**) were
400 selected for further characterization. **Supplementary Table S7** shows the full list of
401 DMRs that passed the filter, including a total of 43 DMRs defined between populations
402 of the endemic *E. cazorlanum* that vary in length between 53 and 165 bp, and 145 DMRs
403 defined between populations of the widespread *E. cicutarium* that vary in length between
404 51 and 343 bp. GO term enrichment analysis did not reveal enriched GO terms in any set
405 of DMRs. In *E. cazorlanum*, we found 19 DMRs whose available annotation was
406 associated with genes involved in plant development, amino acid metabolism, or stress

407 response. In *E. cicutarium*, 49 out of 145 DMRs showed available annotation, and were
408 associated with plant development, stress response, signaling, and cell wall homeostasis,
409 among other biological processes.

410 The two species showed an overall similar pattern regarding DMR overlapping with
411 genomic features (**Figure 4**). DMRs in CpG context (17 in *E. cazorlanum*, 18 in *E.*
412 *cicutarium*) overlapped most frequently with exons in both species (93.8 % in *E.*
413 *cazorlanum*, 61.1 % in *E. cicutarium*), while DMRs in CHH context (24 in *E. cazorlanum*,
414 125 in *E. cicutarium*) overlapped evenly with exons, introns and intergenic regions. The
415 number of DMRs in CHG context (three in *E. cazorlanum*, two in *E. cicutarium*, all
416 overlapping with exons) is too low to reach any clear conclusion. Altogether, from the 43
417 DMRs found in *E. cazorlanum*, 29 (67.4 %) overlapped with exons, six (14.0 %) did so
418 with introns, four (9.3 %) with intergenic regions, two overlapped simultaneously with
419 an exon and an intron (4.7 %), one with a 5'UTR (2.3 %), and one (2.3 %) with a 3' UTR.
420 We found more evenly distributed overlaps in the 145 DMRs found between populations
421 of *E. cicutarium*: 44 (30.6 %) overlapped with intergenic regions, 42 (29.2 %) with exons,
422 41 (28.5 %) with introns, eleven with an exon and an intron (7.6 %), four (2.8 %) with a
423 3'UTR, one with an intron and a 3'UTR (0.7 %) and one with a 5'UTR and an exon (0.7
424 %). A higher proportion of DMRs was found in intergenic regions in *E. cicutarium* (30.56
425 %) compared to *E. cazorlanum* (9.30 %).

426 Regarding DMRs overlapping with transposable elements (TEs), most of them occurred
427 in DMRs associated to CHH methylation context: 14 out of 15 DMRs overlapping with
428 TEs involved methylation changes in CHH in *E. cazorlanum* (93.33 %), and all the 26
429 DMRs overlapping with TEs in *E. cicutarium* involved only methylation changes in CHH
430 (**Figure 4**). In *E. cazorlanum*, TEs with most overlaps with DMRs were LTRs (with
431 eleven *Copia* and two *Gypsy* elements overlapping with a DMR), followed by three DNA

432 TEs (one *MULE-MuDR*, one *CMC-EnSpm*, and one uncharacterized DNA element
433 overlapping with a DMR). For *E. cicutarium*, the most overlapped TE family was also
434 *Copia* with 13 DMRs, in addition, it showed six *LINE/L1*, three *PIF-Harbingers*, one
435 *MULE-MuDR*, one *hAT-Tag1*, one *CMC-EnSpm*, and one *Gypsy* element overlapping
436 with DMRs. **Supplementary Table S8** shows all overlaps between a DMR with a
437 genome feature, a TE, or both.

438 The set of DMRs from *E. cicutarium* showed a higher than expected proportion of
439 overlaps with *Copia* TEs compared with the proportion of *Copia* TEs in the genome of
440 *E. cicutarium* (χ^2 = 3.522, df = 1, p-value = 0.045, **Supplementary Figure S13,**
441 **Supplementary Table S9**). An analogous analysis in *E. cazorlanum* was not conducted
442 because the TE set in *E. cazorlanum* genome may be substantially different from *E.*
443 *cicutarium* genome, making the comparison between *E. cazorlanum* TEs in DMRs and
444 *E. cicutarium* TE genome set inadequate.

445 **3.6. Comparative analysis between species.**

446 Given the difference in read mapping efficiency between species (**Supplementary**
447 **Figure 14**), we used the ratio between mean mapping efficiencies (0.33) as normalization
448 factor when comparing *E. cicutarium* read coverage values with *E. cazorlanum*. First, we
449 compared samples from *E. cazorlanum* and *E. cicutarium* before normalization of *E.*
450 *cicutarium* read counts, detecting 10,765 DMCs (2,790 CpG, 2,272 CHG, 5,703 CHH)
451 and 113 DMRs (12 CpG, 24 CHG, 77 CHH). After normalization, we found 13,744
452 DMCs (6,739 CpG, 2,154 CHG, 4,851 CHH, **Supplementary Table S10,**
453 **Supplementary Figure S14**) and 183 DMRs (109 CpG, 23 CHG, 51 CHH), which were
454 reduced to 19 DMR (5 CpG, 2 CHG, 12 CHH) after filtering by MSD (**Supplementary**
455 **Table S11, Supplementary Figures S14, S15**). Similarity in the number of DMCs before
456 and after normalization was significantly greater than expected for all contexts

457 (hypergeometric test, P-value < 0.001), although the direction of the loss of DMCs was
458 not the same for the three sequence contexts: 3,961, 232, and 351 DMCs were gained,
459 while 12, 350 and 1,203 DMCs were lost after normalization in CpG, CHG and CHH
460 contexts, respectively. Before and after normalization, *E. cicutarium* showed more hyper-
461 methylated CpG sites compared to *E. cazorlanum*, while *E. cazorlanum* showed more
462 hyper-methylation in CHG and CHH contexts compared to *E. cicutarium*.

463 Hierarchical clustering based on methylation percentage of the DMCs did not show a
464 perfect divergence of species, but clades showed a species-like pattern rather than a
465 population-like pattern, and the use of "gradual" DMCs did not change this pattern
466 (**Supplementary Figure S16**). Multivariate analyses revealed a separation between
467 centroids of the two species, a higher dispersion in *E. cicutarium* samples, and little
468 within-species population variation, either using all DMCs or "gradual" DMCs
469 (**Supplementary Figure S17**).

470 Among the 19 inter-specific DMRs obtained, we found association with an endochitinase
471 (hydrolyses chitin and chitodextrins), *Cytohesin-2* (involved in regulation of protein
472 sorting and membrane trafficking) and *pre-mRNA-splicing factor CEF1* (involved in cell
473 cycle control) hyper-methylated in the CHH context in *E. cazorlanum*. The only DMR
474 hyper-methylated in *E. cicutarium* was associated to the *LTL1* gene (involved in salt
475 tolerance mechanisms) with increased methylation in the CpG context. Eleven DMRs
476 from the inter-species comparison overlapped with exons, five with introns, two with an
477 intergenic region, one with a 3' UTR, and another one with both an exon and an intron.
478 Ten DMRs from the inter-species comparison overlapped with fourteen TEs. They were
479 mostly unknown repeats (eight out of 14), but also included three *Copia*, one *Gypsy*, one
480 *SINE/tRNA-V*, and one *PIF-Harbinger* elements (**Supplementary Table S12**).

481

482 **4. Discussion**

483 This study analyzes variation in both DNA sequence (genetic) and cytosine methylation
484 levels (epigenetic) at single-base resolution in natural populations of two *Erodium*
485 species, the endemic *E. cazorlanum* and the widespread *E. cicutarium*, using bsRADseq
486 and the structural and functional annotation provided by a draft reference genome
487 developed *ad hoc* for the study species with the smallest genome (*E. cicutarium*). The
488 two species exhibited sound differences in life history (woody perennial *vs*. annual),
489 habitat specialization (dolomite specialist *vs*. ruderal) and geographic distribution range
490 (local narrow endemic *vs*. world-wide widespread) that allowed us to test the hypothesis
491 that population genetic and epigenetic structure within the region of coexistence would
492 differ according to each species ecological features, with stronger epigenetic than genetic
493 population structure in the least specialized (*E. cicutarium*). Besides, the use of methods
494 that provide accurate genomic information of the methylation markers allowed the
495 identification of a set of candidate genes whose epigenetic regulation might vary across
496 the study localities of each species according to the recorded differences in DNA
497 methylation, paving the avenue towards more specific analysis of adaptive strategies in
498 species with contrasting genomic and ecological features. In the following paragraphs,
499 we will first discuss the limitations of the reduced representation methodology used when
500 applied to species with contrasting genome features and resources (see also Paun et al.,
501 2019). Then, we will discuss the results obtained for our study system and how they
502 reinforce the *a priori* prediction mentioned above.

503 **4.1. Challenges and opportunities of RRBS analyses in plant ecological epigenetics**

504 In absence of any high-quality gapless genome for any species in the *Erodium* genus,
505 RRBS was the best available option to simultaneously detect natural genome and

506 methylome variants across populations and explore their genomic location. In
507 retrospective, six samples per population were estimated sufficient to reach an appropriate
508 level of representation of what occurs in wild plant populations in genomics studies
509 (Nazareno et al., 2017), and by using RRBS we expected to generate informative
510 methylation datasets when conjointly used with genomic data for the same individuals
511 (Gawehns et al., 2022; Paun et al., 2019). In fact, several studies have successfully
512 identified differentially methylated loci by optimizing RRBS and its downstream analysis
513 in commercial plant species with available reference genome (e.g., *Brassica rapa,* Chen
514 et al., 2015; maize, Hsu et al., 2017). In this study, we succeeded in obtaining
515 differentially methylated cytosines (DMC) and regions (DMR) between populations in
516 the two *Erodium* species with six individuals per population, despite the limited power
517 for differential methylation analysis presumed for the afforded sample sizes (Lea et al.,
518 2017).

519 Interpretation of the results obtained for the portion of genome here explored should take
520 into account that due to the relatively high GC content in the recognition site of the
521 specific enzyme we used (*Pst*I that cuts at CTGCAG), bsRADseq dataset is enriched for
522 genic regions (Lea et al., 2017; Trucchi et al., 2016) and, thus, quantitative global
523 estimates would be less accurate for genomes with a relative high content of repetitive
524 elements (i.e., in *E. cazorlanum* according to its larger genome size). Nevertheless,
525 percentages of cytosine methylation per sequence context in the two species were
526 consistent among each other and also with those reported for other plant species with
527 complete methylomes, i.e., CpG higher than CHG, CHH being the lowest (Cokus et al.,
528 2008; De Kort et al., 2020; Hsu et al., 2017; Zhao et al., 2020). The higher proportion of
529 methylated to unmethylated cytosines obtained in *E. cazorlanum* is consistent with a
530 marked difference in percentages obtained from independent analyses conducted by

531 HPLC to obtain global DNA methylation estimates for the two species (28.4 % *vs*. 18.2
532 %; Alonso unpublished data). The sign of divergence in estimates of global DNA
533 methylation per species could be associated to the largest genome size of *E. cazorlanum*
534 (Alonso et al., 2015; Vidalis et al., 2016). The higher heterogeneity in overall methylation
535 estimates of study individuals obtained for *E. cazorlanum* could be due to either genome
536 size, and some unmeasured micro-environmental features (Valverde et al., 2024) or a
537 combination of the two effects.

538 Further, using the reference genome from a close relative species generated a homology
539 bias during the alignment of the reads that may weaken the potential for functional
540 annotation of loci with divergent methylation in *E. cazorlanum*. Accordingly, the few
541 DMRs obtained in *E. cazorlanum* were located in exonic regions, likely due to a higher
542 similarity in the coding regions of the two species' genomes and not necessarily to
543 absence of methylation divergence in genomic regions that were not captured (Lea et al.,
544 2017).

545 Altogether, our study exemplifies the challenges of applying NGS markers to species with
546 limited genomic resources and complex genomes. However, conducting epigenomic
547 studies of such species is compulsory to improve our current understanding of ecological
548 factors behind the huge variation in global DNA methylation across plant species, that
549 require a broad phylogenetic perspective (Alonso et al., 2015), and might be useful to
550 better characterize the sources of variation in DNA methylation at plant ecosystem or
551 community levels (Alonso et al., 2019b; Boquete et al., 2021). Our results support the use
552 of RRBS for interspecific comparisons of natural variation in DNA methylation at
553 difference sequence contexts, and a more restricted power to explore candidate genes and
554 transposable elements potentially differentially regulated by DNA methylation when
555 high-quality reference genomes are not available (Paun et al., 2019).

556 **4.2. Genetic and methylation variation in natural populations of *E. cazorlanum* and**
557 ***E. cicutarium***

558 In the small geographic area here explored, selected populations of each of the two study
559 species lack or show limited signs of genetic population structure, suggesting they are
560 well connected to each other, showing most genetic variation between individuals from
561 the same population. Whereas this fact should be taken with caution for extrapolation to
562 broader geographic areas, all study individuals could be associated to a single genetic
563 background eventually, making the investigation of methylation variants more
564 straightforward. Spontaneous epimutations, i.e. random changes in methylation status of
565 cytosines, are estimated to be four to five orders of magnitude higher than mutation rate
566 in *Arabidopsis*, which may facilitate the decoupling of methylation variation from genetic
567 variation and some specific contribution to heritable phenotypic variation (Banta &
568 Richards, 2018; Denkena et al., 2021; van der Graaf et al., 2015; Weng et al., 2019). In
569 this study, we found absence of correlation between genetic and epigenetic distances
570 across individuals in the two study species, supporting it would be worth to
571 experimentally exploring their relative contribution to individual phenotypes and their
572 plasticities under controlled environmental conditions (see e.g. Williams et al., 2023).

573 **4.3. Genomic characterization of DNA methylation variants and functional**
574 **annotation of DMRs**

575 In each of the two species, most methylation changes between individuals with different
576 population of origin were found in CpG and CHH contexts compared to a much lower
577 frequency in CHG. A high level of CHH methylation variation is expected given the
578 sensitivity of this context to environmental changes (Dubin et al., 2015); in fact, there is
579 evidence that epigenetic distance based in CHH methylation differences can be positively

580 correlated with climatic distance in *Thlaspi arvense* (Galanti et al., 2022). In the endemic
581 *E. cazorlanum*, natural variation in methylation at those DMCs did not exhibit any
582 population structure, regardless the analyses considered all or context-specific DMCs. In
583 the widespread *E. cicutarium*, we found some epigenetic differentiation between the three
584 study populations, supporting our expectation that, in absence of genetic population
585 structure, the broader habitat conditions expected in a widespread species could
586 contribute to larger methylation variance between study individuals at different scales
587 (Valverde et al., 2024).

588 We only proceed with functional annotation of DMRs because solid experimental
589 evidence has linked methylation gains and losses in large genomic regions (i.e., DMRs)
590 with transcriptional variation and heritable phenotypic effects, whereas the functional
591 relevance of single methylation changes (i.e., DMCs) might be more variable across plant
592 species (De Kort et al., 2022; Denkena et al., 2021). Furthermore, although our analytical
593 procedure based on *Bismark* concentrates exclusively on uniquely mapping reads that in
594 turn will result in a depletion of representation of repetitive genomic elements, there was
595 also a fair number of DMRs overlapping with TEs in our *E. cicutarium* data. As a rule,
596 methylation levels in TEs are maintained high in plants to silence their expression and
597 maintain genome stability (Matzke & Mosher, 2014). In our TE annotated bsRADseq
598 data, the *Copia* family was the most abundant, and it was found overlapping with DMRs
599 in a higher than expected proportion compared to the total number of TEs identified in
600 the *E. cicutarium* draft genome. The expression of *Copia* TE family seems stress
601 responsive in several plant species (Ito et al., 2011, 2016; Klein & Anderson, 2022),
602 suggesting that *Copia* elements with CHH methylation islands in promoter regions that
603 have shifted methylation between our study locations might be potentially linked to stress
604 genes' expression regulation. Additional studies in *E. cicutarium*, in which changes in

605 DNA methylation caused by 5-azacytidine and the effect of recurrent drought were
606 experimentally addressed, also found that induced DMRs were enriched with *Copia*
607 elements (Balao et al., 2023), supporting the idea that this TE family may have a stress-
608 related regulatory function in *E. cicutarium* too.

609 Finally, deep differences in read coverage obtained for the two species could cause biases
610 in the direct inter-specific comparison of methylation landscapes offered by DMC
611 detection. Following a normalization step to equilibrate coverage inspired in library size
612 normalization for RNA-seq data (Conesa et al., 2016; Zeng & Mortazavi, 2012), we found
613 that the overall number of detected DMCs between the two species was practically the
614 same, although the relative frequency per sequence context changed and more significant
615 methylation changes in CpG were retained after normalization. Further studies comparing
616 two or more related species with a single available reference genome would be required
617 to validate this methodology. Interestingly, main differences in genomic methylation
618 profiles observed by DMC detection between the two study species involved
619 hypermethylation in non-CpG sites in *E. cazorlanum* and hypomethylation of CpG sites,
620 except when included in DMRs that were most frequently hypermethylated in all
621 contexts.

622 **4.4. Conclusions and future steps.**

623 With limited genomic resources typically available for non-model plants of conservation
624 concern, this study compared DNA methylation at single-base resolution in a reduced
625 portion of the genomes of two congeneric species and explored their spatial structure in
626 a small area of coexistence. We found support for (i) an overall higher global DNA
627 methylation in the endemic *E. cazorlanum*, particularly associated to hypermethylation
628 in non-CpG sites, and in agreement with its higher genome size compared to its
629 widespread congener *E. cicutarium*. (ii) The absence of genetic population structure in

630 the endemic species, and (iii) the lack of correlation between genetic and epigenetic
631 variation across individuals of both species, suggest that they could have independent
632 effects on individual phenotypes. Finally, (iv) a slightly higher DNA methylation
633 divergence between study populations in *E. cicutarium* supports the hypothesis that a
634 stronger epigenetic, rather than genetic, population structure could be expected in
635 widespread generalist species (see also Valverde et al. 2024). Future studies should take
636 the advantage of the genomic resources generated here to link phenotypic, genetic and
637 epigenetic variation in experimental settings suitable to identify the epigenetic
638 contribution to plant adaptation to environmental variation.

639 **Author contributions**

640 CA obtained funding for the study. MM, FB, OP, CA designed the experimental setting.
641 MM and CA conducted fieldwork, PB performed molecular lab work. RM-B analyzed
642 the data. RM-B wrote the first draft of the paper, with feedback from FB, OP and CA. All
643 co-authors provided comments and suggestions on the final version of the manuscript and
644 approved it.

645 **Acknowledgements**

646 The authors gratefully thank to Noelia Zarza for assisting us during field sampling; to
647 Ben Zonneveld (Netherlands) for determining the genomic sizes of *E. cazorlanum* and *E.*
648 *cicutarium*; to Abelardo Aparicio (University of Sevilla, Spain) for determining the
649 ploidy levels of *E. cazorlanum* and *E. cicutarium*; and to Carlos M. Herrera, Javier
650 Valverde, Javier Puy and Anupoma N. Troyee (Estación Biológica de Doñana, Spain) for
651 providing valuable discussion. Funding for this work has been provided by grant
652 EPIENDEM (CGL2016-76605-P) from the Spanish Ministero de Economía y

Competitividad, and grants EPINTER (PID2019-104365GB-I00) and DISTEPIC (PID2022-141530NB-C22) from the Spanish Ministerio de Ciencia e Innovación.

Data accessibility and Benefit-generation statement

Illumina raw sequence reads are deposited in the SRA (BioProject PRJNA1071520). PacBio long reads are also deposited in the SRA (BioProject PRJNA984161). Draft genome assembly is deposited in NCBI Genome repository (accession JATABU000000000). Scripts are available under personal request to RM-B and CA. Benefits Generated: Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Rubén Martín-Blázquez: https://orcid.org/0000-0002-4400-6514

Mónica Medrano: https://orcid.org/0000-0002-2816-4192

Pilar Bazaga: https://orcid.org/0000-0002-8295-4937

Francisco Balao: https://orcid.org/0000-0003-2104-3846

Ovidiu Paun: https://orcid.org/0000-0002-8295-4937

Conchita Alonso: https://orcid.org/0000-0002-7418-3204

References

Alexa, B., & Rahnenfuhrer, J. (2023). topGO: Enrichment Analysis for Gene Ontology. Retrieved from https://bioconductor.org/packages/topGO doi:10.18129/B9.bioc.topGO

673 Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for
674 individual ancestry estimation. *BMC Bioinformatics,* **12**, 246. doi: 10.1186/1471-
675 2105-12-246
676 Alonso, C., & García-Sevilla, M. (2013). Strong inbreeding depression and individually
677 variable mating system in the narrow endemic *Erodium cazorlanum*
678 (Geraniaceae). *Anales del Jardín Botánico de Madrid,* **70**(1), 72-80.
679 doi:10.3989/ajbm
680 Alonso, C., Medrano, M., Perez, R., Canto, A., Parra-Tabla, V., & Herrera, C. M. (2019a).
681 Interspecific variation across angiosperms in global DNA methylation:
682 phylogeny, ecology and plant features in tropical and Mediterranean
683 communities. *New Phytologist,* **224**(2), 949-960. doi:10.1111/nph.16046
684 Alonso, C., Perez, R., Bazaga, P., & Herrera, C. M. (2015). Global DNA cytosine
685 methylation as an evolving trait: phylogenetic signal and correlated evolution with
686 genome size in angiosperms. *Frontiers in Genetics,* **6**, 4.
687 doi:10.3389/fgene.2015.00004
688 Alonso, C., Ramos-Cruz, D., & Becker, C. (2019b). The role of plant epigenetics in biotic
689 interactions. *New Phytologist,* **221**(2), 731-737. doi:10.1111/nph.15408
690 Anderson, J. T., & Gezon, Z. J. (2015). Plasticity in functional traits in the context of
691 climate change: a case study of the subalpine forb *Boechera stricta*
692 (Brassicaceae). *Global Change Biology l,* **21**(4), 1689-1703.
693 doi:10.1111/gcb.12770
694 Balao, F., Medrano, M., Bazaga, P., Paun, O., & Alonso, C. (2023). Long-term
695 methylome changes after experimental seed demethylation and their interaction
696 with recurrent water stress in *Erodium cicutarium* (Geraniaceae). *bioRxiv*.
697 doi:10.1101/2023.01.19.524693

698 Balao, F., Paun, O., & Alonso, C. (2018). Uncovering the contribution of epigenetics to
699 plant phenotypic variation in Mediterranean ecosystems. *Plant Biology ,* **20** (S1),
700 38-49. doi:10.1111/plb.12594
701 Banta, J. A., & Richards, C. L. (2018). Quantitative epigenetics and evolution. *Heredity*
702 *,* **121**(3), 210-224. doi:10.1038/s41437-018-0114-x
703 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
704 models using *lme4*. *Journal of Statistical Software,* **67**(1).
705 doi:10.18637/jss.v067.i01
706 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical
707 and powerful approach to multiple testing. *Journal of the Royal Statistical*
708 *Society: series B (Methodological),* **57**(1), 289-300. doi:10.1111/j.2517-
709 6161.1995.tb02031.x
710 Bewick, A. J., & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current*
711 *Opinion in Plant Biology,* **36**, 103-110. doi:10.1016/j.pbi.2016.12.007
712 Blanca, G., Cabezudo, B., Cueto, M., Fernández López, C., & Morales Torres, C. (2009).
713 *Flora Vascular de Andalucía Oriental. Volumen 3: Rosaceae–Lentibulariaceae*:
714 Consejería de Medio Ambiente, Junta de Andalucía, Sevilla.
715 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for
716 Illumina sequence data. *Bioinformatics,* **30**(15), 2114-2120.
717 doi:10.1093/bioinformatics/btu170
718 Boquete, M. T., Muyle, A., & Alonso, C. (2021). Plant epigenetics: phenotypic and
719 functional diversity beyond the DNA sequence. *American Journal of Botany,*
720 **108**(4), 553-558. doi:10.1002/ajb2.1645
721 Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology*
722 *Letters,* **11**(2), 106-115. doi:10.1111/j.1461-0248.2007.01130.x

723 Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., & Colot, V.
724 (2017). DNA methylation dynamics during early plant life. *Genome Biology,*
725 **18**(1), 179. doi:10.1186/s13059-017-1313-0
726 Chen, X., Ge, X., Wang, J., Tan, C., King, G. J., & Liu, K. (2015). Genome-wide DNA
727 methylation profiling by modified reduced representation bisulfite sequencing in
728 *Brassica rapa* suggests that epigenetic modifications play a key role in polyploid
729 genome evolution. *Frontiers in Plant Science,* **6**, 836.
730 doi:10.3389/fpls.2015.00836
731 Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan,
732 S., Nelson, S. F., & Pellegrini, M., & Jacobsen, S. E. (2008). Shotgun bisulphite
733 sequencing of the *Arabidopsis* genome reveals DNA methylation patterning.
734 *Nature,* **452**(7184), 215-219. doi:10.1038/nature06745
735 Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A.,
736 Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016).
737 A survey of best practices for RNA-seq data analysis. *Genome Biology,* **17**, 13.
738 doi:10.1186/s13059-016-0881-8
739 De Kort, H., Panis, B., Deforce, D., Van Nieuwerburgh, F., & Honnay, O. (2020).
740 Ecological divergence of wild strawberry DNA methylation patterns at distinct
741 spatial scales. *Molecular Ecology,* **29**(24), 4871-4881. doi:10.1111/mec.15689
742 De Kort, H., Toivainen, T., Van Nieuwerburgh, F., Andres, J., Hytonen, T. P., & Honnay,
743 O. (2022). Signatures of polygenic adaptation align with genome-wide
744 methylation patterns in wild strawberry plants. *New Phytologist,* **235**(4), 1501-
745 1514. doi:10.1111/nph.18225

746 Denkena, J., Johannes, F., & Colome-Tatche, M. (2021). Region-level epimutation rates
747 in *Arabidopsis thaliana*. *Heredity,* **127**(2), 190-202. doi:10.1038/s41437-021-
748 00441-w
749 Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A.,
750 Terragni, J., Chin, H. G., Tu, A., Hetzel, J., Wohlschlegel, J. A., Pradhan, S., Patel,
751 D.J., & Jacobsen, S. E. (2012). Dual binding of chromomethylase domains to
752 H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell,*
753 **151**(1), 167-180. doi:10.1016/j.cell.2012.07.034
754 Dubin, M. J., Zhang, P., Meng, D., Remigereau, M. S., Osborne, E. J., Paolo Casale, F.,
755 Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., Jagoda, J., Irez, S., Voronin,
756 V., Song, Q., Long, Q., Rätsch, G., ,Stegle, O., Clark, R. M., & Nordborg, M.
757 (2015). DNA methylation in *Arabidopsis* has a genetic basis and shows evidence
758 of local adaptation. *eLife,* **4**, e05255. doi:10.7554/eLife.05255
759 Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect
760 differentially methylated loci from single nucleotide resolution sequencing data.
761 *Nucleic Acids Research,* **42**(8), e69. doi:10.1093/nar/gku154
762 Fiz-Palacios, O., Vargas, P., Vila, R., Papadopulos, A. S., & Aldasoro, J. J. (2010). The
763 uneven phylogeny and biogeography of *Erodium* (Geraniaceae): radiations in the
764 Mediterranean and recent recurrent intercontinental colonization. *Annals of*
765 *Botany* **106**: 871-884. doi:10.1093/aob/mcq184
766 Galanti, D., Ramos-Cruz, D., Nunn, A., Rodríguez-Arevalo, I., Scheepens, J. F., Becker,
767 C., & Bossdorf, O. (2022). Genetic and environmental drivers of large-scale
768 epigenetic variation in *Thlaspi arvense*. *PLoS Genetics,* **18**(10), e1010452.
769 doi:10.1371/journal.pgen.1010452

770 Gawehns, F., Postuma, M., van Antro, M., Nunn, A., Sepers, B., Fatma, S., van Gurp, T.
771 P., Wagemaker, N. C. A. M., Mateman, A. C., Milanovic-Ivanovic, S., Groβe, I.,
772 van Oers, K., Vergeer, P., & Verhoeven, K. J. F. (2022). epiGBS2: Improvements
773 and evaluation of highly multiplexed, epiGBS-based reduced representation
774 bisulfite sequencing. *Molecular Ecology Resources,* **22**(5), 2087-2104.
775 doi:10.1111/1755-0998.13597

776 Hawes, N. A., Fidler, A. E., Tremblay, L. A., Pochon, X., Dunphy, B. J., & Smith, K. F.
777 (2018). Understanding the role of DNA methylation in successful biological
778 invasions: a review. *Biological Invasions,* **20**(9), 2285-2300. doi:10.1007/s10530-
779 018-1703-6

780 Herrera, C. M., Alonso, C., Medrano, M., Perez, R., & Bazaga, P. (2018).
781 Transgenerational epigenetics: Inheritance of global cytosine methylation and
782 methylation-related epigenetic markers in the shrub *Lavandula latifolia*.
783 *American Journal of Botany,* **105**(4), 741-748. doi:10.1002/ajb2.1074

784 Herrera, C. M., Medrano, M., Bazaga, P., & Alonso, C. (2022). Ecological significance
785 of intraplant variation: Epigenetic mosaicism in *Lavandula latifolia* plants
786 predicts extant and transgenerational variability of fecundity-related traits.
787 *Journal of Ecology,* **110**(11), 2555-2567. doi:10.1111/1365-2745.13964

788 Hsu, F. M., Yen, M. R., Wang, C. T., Lin, C. Y., Wang, C. R., & Chen, P. Y. (2017).
789 Optimized reduced representation bisulfite sequencing reveals tissue-specific
790 mCHH islands in maize. *Epigenetics Chromatin,* **10**(1), 42. doi:10.1186/s13072-
791 017-0148-y

792 Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., & Paszkowski, J. (2011). An
793 siRNA pathway prevents transgenerational retrotransposition in plants subjected
794 to stress. *Nature,* **472**(7341), 115-119. doi:10.1038/nature09861

795 Ito, H., Kim, J. M., Matsunaga, W., Saze, H., Matsui, A., Endo, T. A., Harukawa Y.,
796 Takagi H., Yaegashi H., Masuta Y., Masuda S., Ishida J., Tanaka M., Takahashi
797 S., Morosawa T., Toyoda T., Kakutani T., Kato A., & Seki, M. (2016). A stress-
798 activated transposon in *Arabidopsis* induces transgenerational abscisic acid
799 insensitivity. *Scientific Reports,* **6**, 23181. doi:10.1038/srep23181
800 Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in
801 eukaryotes. *Genetica,* **115**, 49-63. doi:10.1023/a:1016072014259.
802 Klein, S. P., & Anderson, S. N. (2022). The evolution and function of transposons in
803 epigenetic regulation in response to the environment. *Current Opinion in Plant*
804 *Biology,* **69**, 102277. doi:10.1016/j.pbi.2022.102277
805 Kolde, R. (2019). pheatmap: pretty heatmaps. R package version 1.0.12. Retrieved from
806 https://CRAN.R-project.org/package=pheatmap
807 Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next
808 Generation Sequencing Data. *BMC Bioinformatics,* **15**(356). doi:10.1186/s12859-
809 014-0356-4
810 Kreutz, C., Can, N. S., Bruening, R. S., Meyberg, R., Merai, Z., Fernandez-Pozo, N., &
811 Rensing, S. A. (2020). A blind and independent benchmark study for detecting
812 differeally methylated regions in plants. *Bioinformatics,* **36**(11), 3314-3321.
813 doi:10.1093/bioinformatics/btaa191
814 Krueger, F. (2013). Reduced Representation Bisulfite-Seq – A Brief Guide to RRBS.
815 Retrieved from
816 https://www.bioinformatics.babraham.ac.uk/projects/bismark/RRBS_Guide.pdf
817 Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller
818 for Bisulfite-Seq applications. *Bioinformatics,* **27**(11), 1571-1572.
819 doi:10.1093/bioinformatics/btr167

820 Lamka, G. F., Harder, A. M., Sundaram, M., Schwartz, T. S., Christie, M. R., DeWoody,
821 J. A., & Willoughby, J. R. (2022). Epigenetics in ecology, evolution, and
822 conservation. *Frontiers in Ecology and Evolution,* **10**, 871791.
823 doi:10.3389/fevo.2022.871791
824 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.
825 *Nature Methods,* **9**(4), 357-359. doi:10.1038/nmeth.1923
826 Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA
827 methylation patterns in plants and animals. *Nature Reviews Genetics,* **11**(3), 204-
828 220. doi:10.1038/nrg2719
829 Lea, A. J., Vilgalys, T. P., Durst, P. A. P., & Tung, J. (2017). Maximizing ecological and
830 evolutionary insight in bisulfite sequencing data sets. *Nature Ecology &*
831 *Evolution,* **1**(8), 1074-1083. doi:10.1038/s41559-017-0229-0
832 Lenth, R. (2023). emmeans: Estimated marginal means, aka least-squares means.
833 Retrieved from https://CRAN.R-project.org/package=emmeans
834 Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic
835 pathway of increasing complexity. *Nature Reviews Genetics,* **15**(6), 394-408.
836 doi:10.1038/nrg3683
837 Medrano, M., Alonso, C., Bazaga, P., López, E., & Herrera, C. M. (2020). Comparative
838 genetic and epigenetic diversity in pairs of sympatric, closely related plants with
839 contrasting distribution ranges in south-eastern Iberian mountains. *AoB Plants,*
840 **12**(3), plaa013. doi:10.1093/aobpla/plaa013
841 Moreno, J. C. (2008). *Lista Roja 2008 de la Flora Vascular Española*: Ministerio de
842 Medio Ambiente y Medio Rural y Marino.
843 Muyle, A. M., Seymour, D. K., Lv, Y., Huettel, B., & Gaut, B. S. (2022). Gene body
844 methylation in plants: Mechanisms, functions, and important implications for

845 understanding evolutionary processes. *Genome Biology and Evolution,* **14**(4).
846 doi:10.1093/gbe/evac038
847 Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum
848 sample sizes for population genomics: an empirical study from an Amazonian
849 plant species. *Molecular Ecology Resources,* **17**(6), 1136-1147.
850 doi:10.1111/1755-0998.12654
851 Noshay, J. M., & Springer, N. M. (2021). Stories that can't be told by SNPs; DNA
852 methylation variation in plant populations. *Current Opinion in Plant Biology,* **61**,
853 101989. doi:10.1016/j.pbi.2020.101989
854 Nunn, A., Can, S. N., Otto, C., Fasold, M., Díez Rodríguez, B., Fernández-Pozo, N.,
855 Rensing, S. A., Stadler, P. F., & Langenberger, D. (2021). EpiDiverse Toolkit: a
856 pipeline suite for the analysis of bisulfite sequencing data in ecological plant
857 epigenetics. *NAR Genomomics Bioinformatics,* **3**(4), lqab106.
858 doi:10.1093/nargab/lqab106
859 Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R.,
860 O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M., Szoecs, E., Wagner, H.,
861 Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M.,
862 De Caceres, M., Durand, S., & Weedon, J. (2022). Vegan: Community ecology
863 package. R package version 2.6-2. Retrieved from https://CRAN.R-
864 project.org/package=vegan
865 Park, Y., & Wu, H. (2016). Differential methylation analysis for BS-seq data under
866 general experimental design. *Bioinformatics,* **32**(10), 1446-1453.
867 doi:10.1093/bioinformatics/btw026

Paun, O., Verhoeven, K. J. F., & Richards, C. L. (2019). Opportunities and limitations of reduced representation bisulfite sequencing in plant ecological epigenomics. *New Phytologist,* **221**(2), 738-742. doi:10.1111/nph.15388

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* **81**(3), 559-575. doi:10.1086/519795

Pustahija, F., Brown, S. C., Bogunić, F., Bašić, N., Muratović, E., Ollier, S., Hidalgo, O., Bourge, M., Stevanović, V., & Siljak-Yakovlev, S. (2013). Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. *Plant and Soil,* **373**(1-2), 427-453. doi:10.1007/s11104-013-1794-x

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. (2015). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics,* **31**(5), 770-772. doi:10.1093/bioinformatics/btu719

Richards, C. L., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colomé-Tatché, M., Durka, W., Engelhardt, J., Gaspar, B., Gogol-Döring, A., Grosse, I., van Gurp, T. P., Heer, K., Kronholm, I., Lampei, C., Latzel, V., Mirouze, M., Opgenoorth, L., Paun, O., & Verhoeven, K. J. F. (2017). Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology Letters,* **20**(12), 1576-1590. doi:10.1111/ele.12858

Richards, C. L., Bossdorf, O., & Verhoeven, K. J. F. (2010). Understanding natural epigenetic variation. *New Phytologist,* **187**(3), 562-564.

893 Richards, E. J. (2006). Inherited epigenetic variation —revisiting soft inheritance. *Nature*
894 *Reviews,* **7**, 395-401.
895 Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read
896 data using Stacks. *Nature Protocols,* **12(**12), 2640-2659.
897 doi:10.1038/nprot.2017.123
898 Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger,
899 O., Schork, N. J., & Ecker, J. R. (2011). Transgenerational epigenetic instability
900 is a source of novel methylation variants. *Science,* **334**, 369-373. doi:
901 10.1126/science.1212959
902 Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual
903 admixture proportions from next generation sequencing data. *Genetics,* **195**(3),
904 693-702. doi:10.1534/genetics.113.154138
905 Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., & Jacobsen, S.
906 E. (2014). Non-CG methylation patterns shape the epigenetic landscape in
907 *Arabidopsis*. *Nature Structural and Molecular Biology,* **21**(1), 64-72.
908 doi:10.1038/nsmb.2735
909 Szukala, A., Bertel, C., Frajman, B., Schonswetter, P., & Paun, O. (2023). Parallel
910 adaptation to lower altitudes is associated with enhanced plasticity in *Heliosperma*
911 *pusillum* (Caryophyllaceae). *The Plant Journal,* **115**(6), 1619-1632.
912 doi:10.1111/tpj.16342
913 Tan, G., Polychronopoulos, D., & Lenhard, B. (2019). CNEr: a toolkit for exploring
914 extreme noncoding conservation. *PLoS Computational Biology,* **15**(8), e1006940.
915 Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schonswetter, P., & Paun,
916 O. (2016). BsRADseq: screening DNA methylation in natural populations of non-
917 model species. *Molecular Ecology,* **25**(8), 1697-1713. doi:10.1111/mec.13550

Valverde, J., Medrano, M., Herrera, C. M., & Alonso, C. (2024). Comparative epigenetic and genetic spatial structure in Mediterranean mountain plants: a multispecies study. *Heredity,* **132**(2), 106-116. doi:10.1038/s41437-024-00668-3

van der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., Schmitz, R. J., Colomé-Tatché, M., & Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences of the United States of America,* **112**(21), 6676-6681. doi:10.1073/pnas.1424254112

Varotto, S., Krugman, T., Aiese Cigliano, R., Kashkush, K., Kondic-Spika, A., Aravanopoulos, F. A., Pradillo, M., Consiglio, F., Aversano, R., Pecinka, A., & Miladinovic, D. (2022). Exploitation of epigenetic variation of crop wild relatives for crop improvement and agrobiodiversity preservation. *Theoretical and Applied Genetics,* **135**(11), 3987-4003. doi:10.1007/s00122-022-04122-y

Vidalis, A., Zivkovic, D., Wardenaar, R., Roquis, D., Tellier, A., & Johannes, F. (2016). Methylome evolution in plants. *Genome Biology,* **17**(1), 264. doi:10.1186/s13059-016-1127-5

Weng, M. L., Becker, C., Hildebrandt, J., Neumann, M., Rutter, M. T., Shaw, R. G., Weigel, D., & Fenster, C. B. (2019). Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. *Genetics,* **211**(2), 703-714. doi:10.1534/genetics.118.301721

Williams, B. R., Miller, A. J., & Edwards, C. E. (2023). How do threatened plant species with low genetic diversity respond to environmental stress? Insights from comparative conservation epigenomics and phenotypic plasticity. *Molecular Ecology Resources*. doi:10.1111/1755-0998.13897

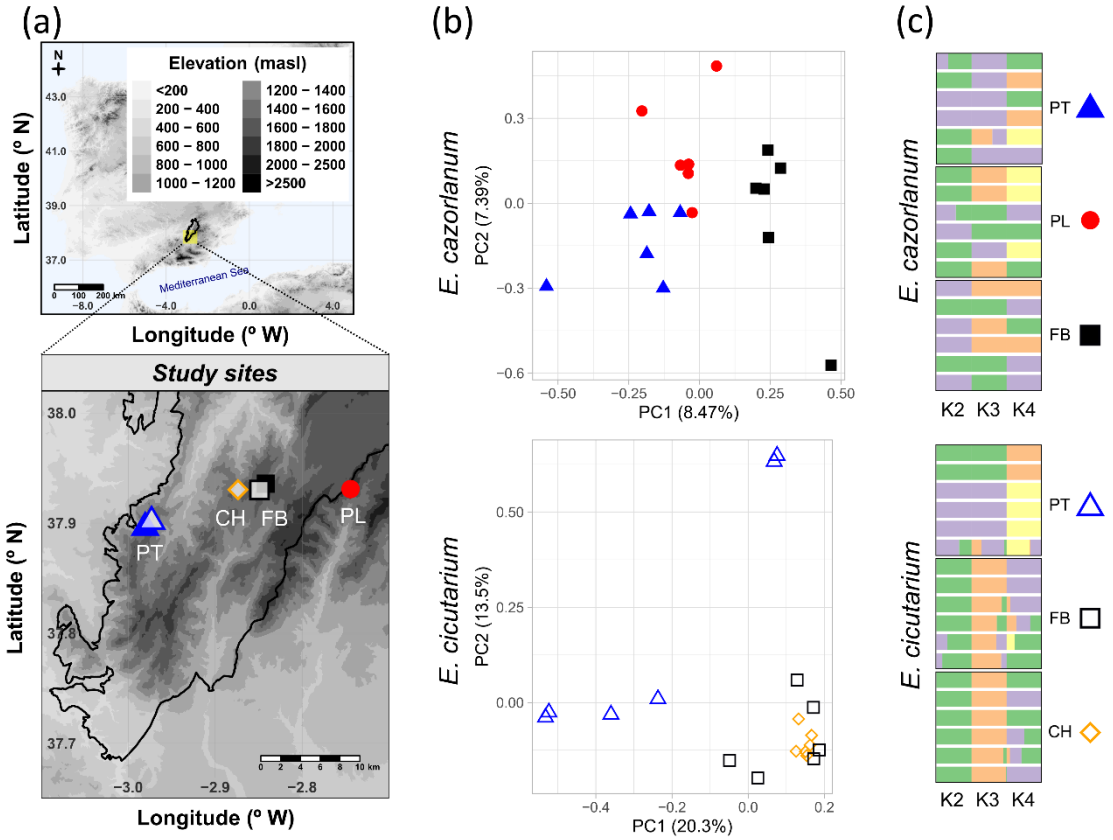
Záveská, E., Maylandt, C., Paun, O., Bertel, C., Frajman, B., The STEPPE Consortium, & Schönswetter, P. (2019). Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Molecular Phylogenetics and Evolution*, **139**. doi:10.1016/j.ympev.2019.106572

Zeng, W., & Mortazavi, A. (2012). Technical considerations for functional sequencing assays. *Nature Immunology,* **13**(9), 802-807. doi:10.1038/ni.2407

Zhao, L., Xie, L., Zhang, Q., Ouyang, W., Deng, L., Guan, P., Ma, M., Li, Y., Zhang, Y., Xiao, Q., Zhang, J., Li, H., Wang, S., Man, J., Cao, Z., Zhang, Q., Zhang, Q., Li, G., & Li, X. (2020). Integrative analysis of reference epigenomes in 20 rice varieties. *Nature Communications,* **11**(1), 2658. doi:10.1038/s41467-020-16457-5

Zonneveld, B. J. M. (2019). The DNA weights per nucleus (genome size) of more than 2350 species of the Flora of The Netherlands, of which 1370 are new to science, including the pattern of their DNA peaks. *Forum geobotanicum,* **8**, 24-78. doi:10.3264/FG.2019.1022

960 **Figures**



961

962 **Figure 1.** Genetic population structure of the two study species. A: Study area and
963 localities. Top: map of the Iberian Peninsula showing the area of the Parque Natural
964 Sierras de Cazorla-Segura-Las Villas (delimited by a solid black line) and the sampling
965 area (yellow square). Altitude (masl = meters above sea level) is represented as a grey
966 scale. Bottom: Detail of the sampling area, showing the study localities of *Erodium*
967 *cazorlanum* and *E. cicutarium*. CH = Nava Correhuelas, FB = Fuente Bermejo, PL =
968 Puerto Lézar, PT= Puerto del Tejo. B: Principal component analysis for *E. cazorlanum*
969 (top) and *E. cicutarium* (bottom) based on SNPs obtained by RADseq. Each biplot shows
970 principal components PC1 and PC2 (together with their percentage of explained variance)
971 in X- and Y-axis, respectively; data points are individuals; populations are shown with
972 different colors and shapes. C: Admixture analysis for *E. cazorlanum* (top) and *E.*
973 *cicutarium* (bottom) based on SNPs from RADseq. X-axis shows the K value chosen for

974 each model. Y-axis shows all study individuals, with populations enclosed in different
975 boxes. Colors represent different modeled genetic lineages according to the selected K
976 value. Each color within a tile indicates the proportion of SNPs from different modeled
977 lineages.

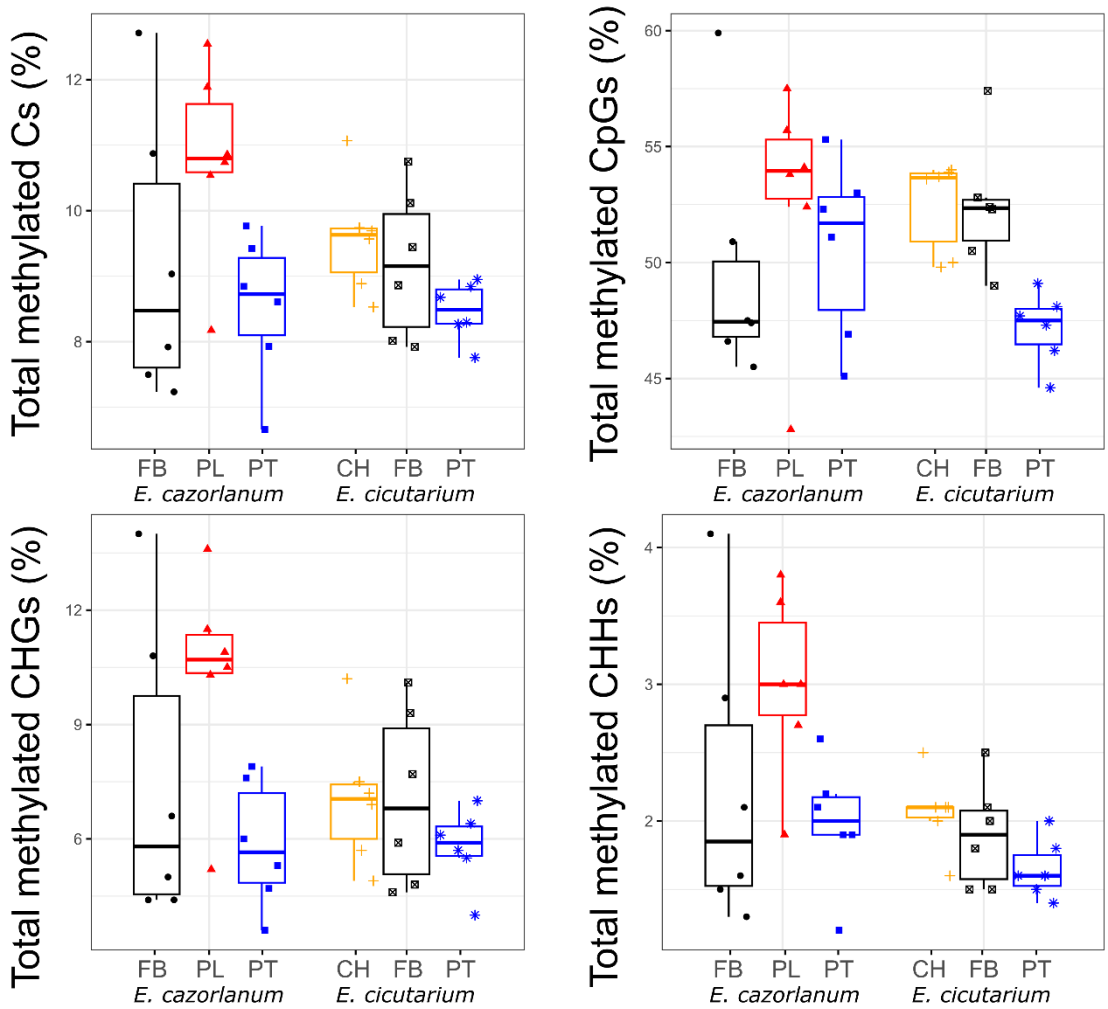
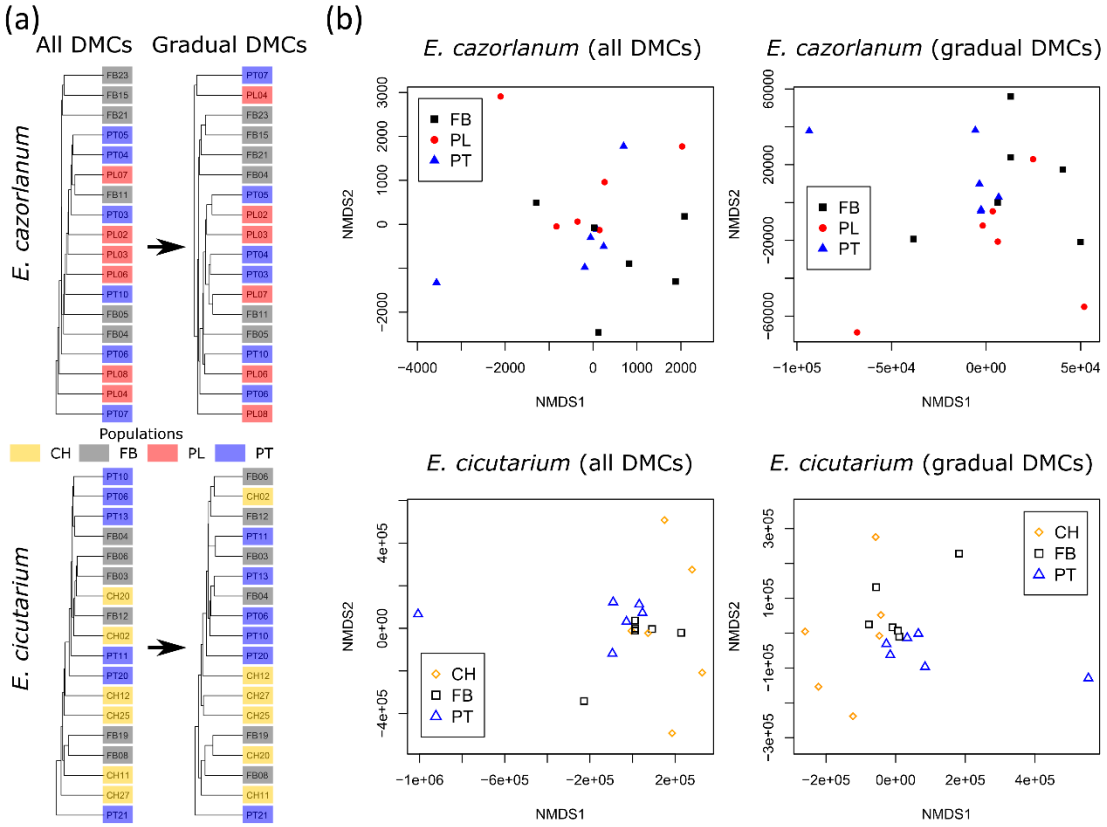


Figure 2. Variation in levels of DNA methylation in populations of *E. cazorlanum* and *E. cicutarium*. The plots show percent of all the sequenced cytosines that were methylated (top left) and estimates obtained for cytosines in the three methylation contexts: CpG (top right), CHG (bottom left) and CHH (bottom right). On each graph, symbols represent individuals, X-axis shows the study populations of the two *Erodium* species, and Y-axis shows the methylation percentage estimated. Each box shows the median (central line), quartiles 1 and 3 (bottom and top edges of the box), and maximum and minimum values (box whiskers) per population (represented by different colors and shapes). Note that Y-axes have different scales.



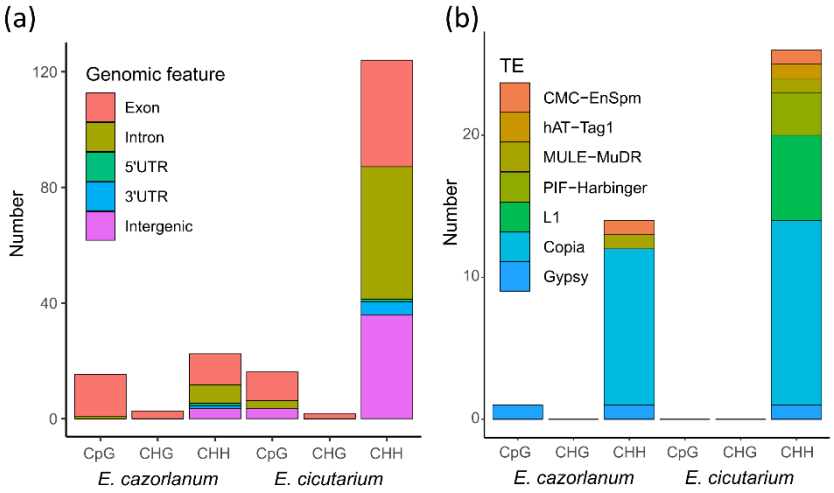
989

990 **Figure 3.** Epigenetic population structure of the two *Erodium* species. A: Hierarchical
991 individual clustering according to obtained differentially methylated cytosines (DMCs)
992 based on percentage of methylation per DMC locus. Top dendrograms correspond to *E.*
993 *cazorlanum* (N = 13,643 loci) and bottom dendrograms to *E. cicutarium* (N = 47,615
994 loci). Each column shows the resulting clustering of all DMCs (left) and gradual DMCs
995 (right), see text for details. Colored tiles attached to each sample ID show the population
996 of origin. B: Multilocus analysis of individual variation in methylation percentage of all
997 DMCs (left graphs) and gradual DMCs (right graphs) based on non-metric
998 multidimensional scaling (NMDS). Top graphs show data for *E. cazorlanum*, bottom
999 graphs for *E. cicutarium*. X and Y axes show non-linear multidimensional scaling
1000 (NMDS) dimensions NMDS1 and NMDS2, respectively. Colors of data points show the
1001 population of origin.

1002

1003

1004



1005 **Figure 4.** Genomic characterization of differentially methylated regions (DMRs)
1006 obtained for the two *Erodium* species. A: Genomic features overlapping with DMRs. X-
1007 axis shows the species and DMR methylation context (*E. cazorlanum*: CpG = 16, CHG
1008 = 3, CHH = 24; *E. cicutarium*: CpG = 18, CHG = 2, CHH = 125). Y-axis shows the
1009 numbers of DMRs found overlapping with each of the five different genomic features
1010 distinguished (exon, intron, 5'UTR, 3'UTR and intergenic regions). B: Transposable
1011 elements overlapping with DMRs. X-axis shows the species and DMR methylation
1012 context (*E. cazorlanum*: CpG = 1, CHG = 0, CHH = 14; *E. cicutarium*: CpG = 0, CHG =
1013 0, CHH = 27). Y-axis shows the number of DMRs found overlapping with each of the
1014 nine families of TEs distinguished.

1015

1016 **Tables**

1017 **Table 1.** Coefficients of variation of global DNA methylation percentages. First column
1018 shows each methylation context. Each set of four columns shows information related to
1019 each species. The first column of each species set shows CV from all populations, the
1020 three remaining columns show CVs for each population.

1021

	E. cazorlanum				*E. cicutarium*			
CVs	Overall	FB	PL	PT	Overall	CH	FB	PT
Total Cs	18.72 %	22.54 %	13.59 %	12.54 %	9.74%	7.55%	12.30 %	5.41%
CpG	9.28%	10.77 %	9.79%	7.65%	6.51%	3.85%	5.42%	3.34%
CHG	42.69 %	52.90 %	26.94 %	28.58 %	27.24 %	25.79 %	32.99 %	17.68 %
CHH	35.50 %	47.68 %	22.61 %	23.31 %	17.93 %	13.91 %	20.25 %	13.14 %

1022

1023

1024 **Table 2.** Global cytosine methylation differences at species and population level. First
1025 column shows the set of cytosines for which a binomial distribution model was run to
1026 test whether the difference between the proportion of methylated to unmethylated
1027 cytosines (mC / uC) was statistically significant. Columns two to four show log-ratio χ2,
1028 degrees of freedom, and p-value for species. Columns five to seven show the same for
1029 populations within species.

	Species			Species:Population		
Proportion of mC / uC	LR χ^2	d.f.	P-value	LR χ^2	d.f.	P-value
Total Cs	74,280	1	< 0.001	918,941	4	< 0.001
CpG	2,118	1	< 0.001	627,781	4	< 0.001
CHG	146,988	1	< 0.001	498,203	4	< 0.001
CHH	340,129	1	< 0.001	426,370	4	< 0.001

1030