



HAL
open science

Study of the genetic and phenotypic variation among wild and cultivated clary sages provides interesting avenues for breeding programs of a perfume, medicinal and aromatic plant

Camille Chalvin, Stéphanie Drevensek, Christel Chollet, Françoise Gilard, Edita Šolić, Michel Dron, Adnane Boualem, Adnane Boualem, Amandine Cornille

► To cite this version:

Camille Chalvin, Stéphanie Drevensek, Christel Chollet, Françoise Gilard, Edita Šolić, et al.. Study of the genetic and phenotypic variation among wild and cultivated clary sages provides interesting avenues for breeding programs of a perfume, medicinal and aromatic plant. 2020. hal-03081424v1

HAL Id: hal-03081424

<https://hal.science/hal-03081424v1>

Preprint submitted on 6 Jan 2021 (v1), last revised 25 Jul 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Study of the genetic and phenotypic variation among wild and cultivated clary sages**
2 **provides interesting avenues for breeding programs of a perfume, medicinal and**
3 **aromatic plant**

4
5 **Camille Chalvin^{1**}, Stéphanie Drevensek^{1**}, Christel Chollet¹, Françoise Gilard¹, Edita M. Šolić²,**
6 **Michel Dron¹, Abdelhafid Bendahmane¹, Adnane Boualem¹, Amandine Cornille³**

7
8 1. Institute of Plant Sciences Paris-Saclay (IPS2), UMR 9213/UMR 1403, Université Paris-Saclay,
9 Bâtiment 630, 91405 Orsay, France

10 2. Institute “Mountain and Sea”, Franjevački put 1, 21300 Makarska, Croatia

11 3. Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190,
12 Gif-sur-Yvette, France

13 * *Correspondence:* adnane.boualem@inrae.fr

14 ** These authors contributed equally to the manuscript

15

16 **Abstract.** A road-map of the genetic and phenotypic diversities in both crops and their
17 wild-related species can help identifying valuable genetic resources for further crop
18 breeding. The clary sage (*Salvia sclarea* L.), a perfume, medicinal and aromatic plant, is
19 used for sclareol production and ornamental purposes. Despite its wide use in the field of
20 cosmetics, the phenotypic and genetic diversity of wild and cultivated clary sage remains to
21 be explored. We characterized the genetic and phenotypic variation of a collection of six
22 wild *S. sclarea* populations from Croatia, sampled along an altitudinal gradient, and of
23 populations of three *S. sclarea* cultivars. We showed low level of genetic diversity for the
24 two *S. sclarea* traditional cultivars used for essential oil production and for ornamental
25 purposes, respectively. In contrast, a recent cultivar resulting from new breeding methods,
26 which involve hybridizations among several genotypes rather than traditional recurrent
27 selection and self-crosses over time, showed high genetic diversity. We also observed a

28 marked phenotypic differentiation for the ornamental clary sage compared with other
29 cultivated and wild clary sages. Instead, the two cultivars used for essential oil production, a
30 traditional and a recent, respectively, were not phenotypically differentiated from the wild
31 Croatian populations. Our results also featured some wild populations with high sclareol
32 content and early-flowering phenotypes as good candidates for future breeding programs.
33 This study opens up perspectives for basic research aiming at understanding the impact of
34 breeding methods on clary sage evolution, and highlights interesting avenues for clary
35 breeding programs.

36
37 Keywords: breeding, perfume, genetic diversity, hybridization, sclareol, *Salvia sclarea*.

38
39 **Introduction**

40

41 A road-map of the genetic and phenotypic diversities in both crops and their wild-related
42 species can help identifying valuable genetic resources for further crop breeding. Over the
43 past century, agricultural practices caused a dramatic erosion of biodiversity in major crop
44 plants [1]. The constant need for genetic improvement to enhance crop performance
45 prompted the search for new genetic resources. Crop-related wild species, the wild plants
46 closely related to cultivated crops, have a high potential genetic diversity for improvement of
47 future breeding programs [2,3].

48 Despite the increasing use of medicinal aromatherapy, genetic and phenotypic variation
49 among wild and cultivated Perfume, Aromatic and Medicinal Plants (PAMP) are still little
50 explored (but see [4,5]). Plant specialized metabolites, such as alkaloids or terpenes, are the
51 characteristics for which PAMP are valued. Traditional phytochemical studies have invested
52 massive effort to isolate and characterize these chemical components, but only on a handful
53 of elite samples. In nature, PAMP chemical composition showed a huge variation which

54 represents a potential untapped allelic variation for breeding programs. However, the plant
55 specialized metabolite composition of a plant depends on both environmental and genetic
56 factors [8,9]. For instance, altitude potentially influences plant specialized metabolism as a
57 combination of abiotic and biotic factors: temperature, hygrometry, wind speed and water
58 availability change with altitude, promoting the development of distinct microbial and insect
59 communities [10,11]. For targeted genetic enhancement of PAMP performances, breeders
60 need to tease apart the environment and genetic effects at the origin of PAMP chemical
61 variation. Common garden experiments including several genotypes sampled along an
62 altitudinal gradient can help identifying interesting phenotypic variation that has only a
63 genetic origin. Interesting phenotypes can then be selected for PAMP breeding programs in
64 order to increase specialized metabolites yield. Common garden experiments can also be
65 coupled with the characterization of the genetic variation at candidate genes involved in the
66 specialized metabolite production pathways to point out interesting genetic variation, or lack
67 of variation due to recent selection imposed by human. The chemical variation in crop and
68 wild PAMP, and the underlying genes involved, are therefore good targets for breeding
69 programs, but have generally not received a great deal of attention. Genetic study coupled
70 with a phenotypic characterization of chemical variation of the crop and wild PAMP
71 populations are therefore needed.

72 The clary sage, *Salvia sclarea* L., is a biennial or short-lived perennial with a native
73 distribution in Northern Mediterranean, but has a vast geographic distribution towards
74 Central Asia and the Middle East [12,13], and also become a weed outside its native range
75 including North America. Clary sage has been traditionally used as a medicinal plant for its
76 antispasmodic, carminative and estrogenic properties [14], and it is now mainly used for its
77 fragrance. Clary sage essential oil, appreciated for its herbaceous and earthy scent, indeed
78 enters in the composition of numerous high-end fragrances. Currently, several clary sage
79 cultivars exist for different uses, *i.e.*, oil production or ornamental purposes, and result from

80 traditional or recent breeding methods. Traditional clary sage breeding methods consist of
81 successive cycles of selection and inbred crosses among individuals from the same
82 population sharing close candidate phenotypes. For oil production, candidate phenotypes
83 show high sclareol and linalool content and, when possible, early flowering. Indeed, the most
84 valued compounds of clary sage is Sclareol, a diterpene extracted from clary sage straw, used
85 as starting material for the synthesis of Ambroxide[®], characterized by a musky odor and
86 scent fixative ability, and linalool, a monoterpene and its acetylated derivative linalyl acetate
87 [12–16]. Early flowering is also a good candidate breeding trait in clary sage. It is widely
88 assumed that the productivity of late-flowering clary sage plants is negatively impacted by
89 summer drought. This rationale is borrowed from breeding programs of chamomile, a
90 medicinal plant which also produces valuable terpenoids in flowers [5]. Thereby, the
91 traditional clary sage breeding led to the creation of cultivars used for essential oil
92 production, *e.g.*, the “Milly” cultivar. Clary sage used for oil production has pink to purple
93 flowers, a phenotype which stands in stark contrast with cultivars used for ornamental
94 purposes which shows white flowers, *e.g.*, the “Vatican White” cultivar. More recently, the
95 raising global demand for sclareol in countries where clary sage is cultivated, including
96 Hungary, Bulgaria, Moldavia, Romania and France, has encouraged new breeding
97 initiatives. The recent clary sage breeding approach involves recurrent selection with
98 successive cycles of three steps: intermating between best-performing populations,
99 evaluation of traits of interest, and selection of best-performing individuals. This new
100 breeding method that involves the mixing of several best performing populations led to the
101 creation of a clary sage variety with high sclareol yield, named “Scalia”. However, the
102 “Scalia” variety displays a late-flowering phenotype, an unwanted phenotype. Further
103 recurrent selection cycles led to the development of the “Toscalia” variety, which is
104 characterized by a high sclareol yield without a late-flowering phenotype. While the Toscalia
105 has an attractive phenotype for breeders, it is still not flowering early enough. Breeders are

106 therefore still seeking for candidate populations with early-flowering, and if possible higher
107 sclareol and linalool contents. Wild clary sage populations represent a potential source of
108 untapped allelic variation. However, phenotypic variation for flowering time, sclareol and
109 linalyl acetate contents of *S. sclarea* cultivars used for different purposes, as well as of wild
110 clary sage populations, are still lacking. The genetic relationships among cultivated and wild
111 clary sages, and their respective genetic diversity, also remained unknown.

112 We investigated genetic and phenotypic variation among populations of three clary sage
113 cultivars (two traditional cultivars, the Milly and the White Vatican, and one recent cultivar,
114 the Toscalia) and wild Croatian clary sage populations. The wild clary sage populations were
115 sampled along an altitudinal gradient to screen for potential untapped phenotypic and genetic
116 variation for future breeding programs. We chose Croatia because wild clary sage is
117 widespread in this country and there is no cultivation of clary sages there, avoiding the
118 sampling of feral clary sages that may bias the interpretations of our results. Using a genetic
119 dataset (Internal transcribed Spacer 1 marker, referred as to “ITS” hereafter, and two
120 candidate genes involved in the diterpene production pathways [17]), we answered the
121 following questions: do genes involved in the sclareol production pathways show high
122 diversity among wild and cultivated clary sage populations or signature of recent selection?
123 More generally, what is the genetic diversity of wild and cultivated clary sages? What are the
124 genetic relationships among cultivated and wild clary sages? Second, we grew in a common
125 environment in France the wild and cultivated clary sages and characterized their phenotypic
126 variation for sclareol and linalyl acetate contents, as well as their flowering time. We
127 answered the following questions: Is there natural variation in flowering time and sclareol
128 contents among wild and cultivated populations, and among wild populations? Our study had
129 a dual objective: orientating breeding towards promising strategies to boost sclareol yield,
130 and improving our general understanding of the impact of breeding methods of the evolution
131 of a perfume, aromatic and medicinal plant.

132 **Material and Methods**

133

134 **Plant material**

135

136 During summer 2016, seeds from 60 wild *S. sclarea* individuals were collected at six sites in
137 Croatia (Table 1 and Figure 1). These six sites were chosen to represent an altitudinal
138 gradient: Makarska (Coast, altitude=80 m, $N=10$ individuals), Blato (Coast, 80 m, $N=10$),
139 Zaoztrog (Coast, 40 m, $N=10$), Kljenak (mountains, 360 m, $N=10$), Ravca (mountains,
140 altitude=550 m, $N=10$), Banja (plateau, 80m, $N=10$). We chose Croatia because wild clary
141 sage is widespread in this country and there is no cultivation of clary sages there. We
142 therefore avoided the potential sampling of feral populations. Seeds were stored at 4 °C and
143 40% humidity until sowing. We also obtained seeds of cultivated *S. sclarea* from the
144 “Vatican White”, “Toscalia” and “Milly” cultivars. Note that under the “cultivar” umbrella
145 fall cultivated individuals resulting for several crosses between individuals sharing close
146 phenotypes for sclareol and linalool contents, flowering time and flower color. As explained
147 in the introduction, the Milly cultivar is traditionally used for perfume and aromatherapy
148 purposes, whereas the Vatican White population is propagated and commercialized for
149 ornamental purposes. The Toscalia variety is originated from recent breeding for commercial
150 purpose for essential oil. Toscalia and Milly seeds were kindly donated by the French
151 National Conservatory of Perfume, Medicinal, Aromatic and Industrial Plants (CNPMAI)
152 and the French Interprofessional Institute of Perfume, Medicinal, Aromatic and Industrial
153 Plants (ITEIPMAI), respectively. There, the Milly cultivar is maintained as a population of
154 clary sage that are recurrently inter-bred. *Salvia sclarea* ‘Vatican white’ seeds were
155 purchased from Jelitto (Germany). Croatian wild seeds were collected in 2017 but are not
156 considered a strictly protected species according to the Croatian Ordinance and therefore do

157 not fall under the Nagoya protocol. Note that the Vatican white has white flowers whereas
158 ‘Milly’, ‘Toscalia’ and wild *S. sclarea* studied have different shades of purple-pink.

159

160 **Genetic diversity and differentiation**

161

162 For each sampled site for the wild *S. sclarea*, and each *S. sclarea* cultivar, three to four
163 individuals were grown in growth chamber (16h of day; 27 °C during the day, 21 °C at night;
164 60% hygrometry) for six weeks. DNA was extracted from leaves using the Qiagen DNeasy
165 96 Plant Kit following manufacturer recommendations. Genomic fragments from *DXS2*,
166 *CMK* and *ITS 1* loci were amplified by PCR using primers listed in Table S1. Sequences of
167 PCR products were obtained through Sanger sequencing performed by Eurofins Genomics.
168 Sequences were aligned using the ClustalW algorithm implemented in the MEGA7 software
169 [18]. Heterozygous sites were visualized in Sanger sequencing chromatograms provided by
170 Eurofins Genomics and manually annotated in MEGA7 using the IUPAC code.

171 We checked the neutrality of the markers to be used for population genetic analyses with
172 *Tajima's D* test [19]. When *Tajima's D* test was non-significant, we assumed the sequences
173 were evolving neutrally. However, *Tajima's D* can be significant in case of bottleneck or
174 expansion, regardless selection. We therefore further test for neutrality of the sequences with
175 the McDonald and Kreitman test (MKT) [20]. The MKT was only possible when a sequence
176 of a *Salvia* outgroup species was available on NCBI for a given marker. The resulting
177 neutrally-evolving sequences were then concatenated. Because of the occurrence for each
178 sequenced marker of heterozygous sites encoded with the IUPAC code, haploid sequences
179 were phased for each individual with the PHASE algorithm implemented the DnaSP
180 software [21]. Haploid sequences obtained (two per individual) were then used for further
181 analyses below.

182 We computed *Watterson's* θ [22] and *Tajima's* D [19] with the DnaSP software [21] for
183 each of the six sites for the wild *S. salvia* and each cultivar (*i.e.*, Vatican White, Toscalia and
184 Milly). We investigated genetic relationships among wild and cultivated samples with a
185 splitstree and a principal component analysis (PCA). For the splitstree, we used the
186 NeighborNet distance transformation and equal angle splits transformation implemented in
187 SplitsTree v 4 [23]. For the PCA, we used the *dudi.pca* function from the *ade4* [24] and
188 *ade4genet* [25] R packages (R Development CoreTeam, URL <http://www.R-project.org>).

189

190 **Phenotypic characterization: field trial, flowering time measurement, and sclareol and**
191 **linalyl acetate quantification by Gas chromatography–mass spectrometry (GC-MS)**

192

193 Seeds from three wild clary sage sites, representing the three macro-environments (*i.e.*,
194 Makarska-coast, Ravca-mountain and Banja-plateau, Figure 1a), and from populations of the
195 Milly and Vatican White cultivars were sown on August 3rd, 2017, and then grown in a
196 growth chamber (16h of day; 27 °C during the day, 21 °C at night; 60% hygrometry) for four
197 months. On November 7th, 2017, 93 seedlings (between 15 and 23 per site or cultivar) were
198 transplanted in the field in Chemillé (France, W 43'43" / N 47°12'36") to allow the plants
199 overwintering. Plants were placed every 0.4 m x 0.7 m according to a randomized layout.
200 Flowering time was defined as the date of opening of the first flower and was recorded
201 between May 31st and June 29th, 2018, until all plants have flowered. For calyx sampling,
202 inflorescences were cut on July 19th, 2018 at the beginning of the afternoon. Mature calyces
203 (harboring green achenes) were cut from all over the inflorescence, immediately placed in
204 dry ice and stored at -80 °C at the end of the day until further analysis.

205 Clary sage calyces were immersed in hexane without grinding to extract metabolites
206 present at the surface. For each individual, 12 mL of solvent were used to extract metabolites
207 from eight mature calyces. The 10-undecen-1-ol was employed as internal standard and was

208 added in the solvent before extraction at a concentration of 0.75 mM. After 2 h of agitation, 1
209 mL of extract was centrifuged at 13,000 rpm for 10 min to eliminate potential dust or debris.
210 Samples were directly analyzed by GC-MS with a 7890B/5977A GC-MS system from
211 Agilent and according to a protocol adapted from Laville et al. [26]. The system was
212 equipped with a 10 m guard column and a Rxi-5Sil MS column (length, 30 m; inner diameter,
213 0.25 mm; film thickness, 0.25 μm) (Restek, Bellefonte, PA, USA). A total of 1 μl of sample
214 was injected with a split ratio of 30:1. Oven temperature was set to 110 $^{\circ}\text{C}$ for 2 min, then
215 increased to 270 $^{\circ}\text{C}$ at a rate of 10 $^{\circ}\text{C}/\text{min}$, and finally set to 270 $^{\circ}\text{C}$ for 2 min. Other
216 temperatures were set as follows: injector, 270 $^{\circ}\text{C}$; transfer line, 270 $^{\circ}\text{C}$; source, 230 $^{\circ}\text{C}$;
217 quadrupole, 150 $^{\circ}\text{C}$. The carrier gas was helium at a constant flow of 1 mL/min. The
218 quadrupole mass spectrometer was switched on after a solvent delay of 5 min and was
219 programmed to scan from 35 u to 350 u. Data analysis was carried out using the MassHunter
220 Quantitative Analysis software from Agilent. Sclareol and linalyl acetate were identified by
221 comparison with analytical standards purchased from Sigma-Aldrich. Absolute
222 quantification of sclareol and linalyl acetate was performed using a calibration curve
223 prepared with analytical standards.

224

225 **Statistical analyses**

226

227 In the presence of a domestication syndrome, we expect differences in phenotypes of
228 cultivated accessions compared to wild accessions. We can also expect phenotypic
229 differences between the cultivated clary sages depending on their uses (ornamental or oil
230 production), and between wild *S. sclarea* plants sampled at sites along an altitudinal gradient
231 in Croatia. Indeed, the selective pressures associated with the three environmental
232 conditions from which the three populations were sampled (Makarska, Coast; Ravca,
233 mountains; Banja, plateau) may have led to natural variation in phenotypes. In order to

234 study the influence of the use of the cultivated sages and the wild status of each plant, we
235 fitted a generalized linear model as follows:

236

$$237 \text{ Traits}_{ij} \sim \mu + \text{wild_site_and_cult_use}_i + e_{ij} \quad (1)$$

238

239 Where μ is the overall mean of the phenotypic trait (flowering time or sclareol or linalyl
240 acetate content), 'wild_and_cult_use' is the fixed effect of the status of each plant (wild from
241 Makarska/Coast, wild from Ravca/mountains, wild from Banja/plateau, cultivated clary sage
242 used for ornamental purpose, *i.e.*, White Vatican cultivar, and cultivated clary sage used for
243 oil production, *i.e.*, Milly cultivar) and e is the residual. We ran a general linear model, with a
244 residual term assumed to be normally distributed. Statistical significance of the difference
245 between status was tested with a pairwise a Kruskal-Wallis rank test.

246

247 **Results**

248

249 **Genetic diversity and relationships among cultivated and wild clary sages**

250

251 Tajima's D and MKT showed that the *ITS*, and the two genes involved in the diterpene
252 production pathway, *DXS2* and *CMK*, evolved neutrally (Table S2). Table 1 gives genetic
253 diversity estimates for each sampling site for the wild clary sages, and for the three cultivars.
254 The map of interpolated pairwise nucleotide diversity, π , by site (Figure 1a) showed that
255 genetic diversity was highest for the two wild coastal sites (Blato and Zaothrop) and one
256 mountain site (Kljenak), and the lowest in the plateau site (Banja) (Table 1 and Figure 1a).
257 The cultivated clary sages showed lower level of genetic diversity than the wild populations,
258 except the recent Toscalia cultivar (Table 1 and Figure 1a).

259 The splitree and the PCA (Figure 1, c and d) revealed a genetic differentiation between
260 wild and cultivated clary sages. Within the cultivated clary sages, the Vatican white showed
261 less variation than the Milly and Toscalia cultivar, and the Toscalia were more differentiated
262 than the Milly and the White Vatican cultivars.

263 **Phenotypic variation between wild and cultivated clary sages, and within each group**

264

265 In total, 20 plants out of the 93 initial transplanted seedlings during the winter of 2017 did not
266 survive the winter. The 73 remaining plants (between 10 and 22 per wild site or cultivar)
267 flowered between May 31st and June 29th, 2018, that is, between 301 and 330 days after
268 sowing. Flowering dates were consistent with data available in the literature; indeed, clary
269 sage is usually reported to flower in June or July [12].

270 Overall, we observed significant differences in flowering time, sclareol and linalyl
271 acetate among wild clary sages, clary sages cultivated for ornamental purposes and for oil
272 production (Figure 2a and Table 2). For wild clary sages, no significant differences in
273 sclareol and linalyl acetate of calyces were observed among plants coming from coast,
274 mountain or plateau (Figure 2, c and d). Significant differences were observed for flowering
275 time, with the Makarska/coastal plants flowering in average 5.044 days later than other wild
276 plants ($P=0.046$, Figure 2b). When comparing the cultivated and wild plants, the Vatican
277 White plants, used for ornamental purposes, produced significantly less linalyl acetate (30%
278 less) than the Milly cultivar and the wild clary sage plants (Figure 2c). The Vatican White
279 plants also produced significantly less sclareol (-0.12 mg/calyx in average, $P<0.0001$, Figure
280 2d) compared to the Milly cultivar and the Makarska/coast and Banja/plateau wild plants.
281 Vatican White plants flowered significantly earlier than the Milly cultivar and the wild
282 Makarska plants (10 days earlier in average) (Figure 2b)

283

284

285 **Discussion**

286

287 We provided a first insight into crop and wild genetic and phenotypic variation of a perfume,
288 medicinal and aromatic plant, the clary sage. Our results showed a lower level of genetic
289 diversity for the traditional clary sage cultivars, whatever their uses, compared with the
290 Croatian wild populations. In contrast, the recent Toscalia cultivar, resulting from new
291 breeding methods, showed a high genetic variation, not significantly different from the
292 Croatian wild populations. Phenotypically, a strong divergence is observed for the
293 ornamental clary sage compared with the two cultivars used for essential oil production.
294 Interestingly, the cultivars used for essential oil production were not significantly
295 differentiated from the Croatian wild populations, whatever the area/altitude of origin of the
296 wild population. We also did not observe phenotypic variation among wild Croatian clary
297 sage populations. Our results finally pinpoint certain Croatian wild populations with
298 interesting sclareol and linalyl acetate contents for future clary sage breeding.

299

300 **A lower genetic diversity for the traditional clary sage cultivars, whatever their uses**

301

302 We did find a lower genetic diversity for the traditional cultivars used for essential oil
303 production (Milly) and for ornamental purposes (Vatican White). Traditional clary sage
304 breeding methods consist of successive cycles of selection and inbreed crosses among
305 individuals from the same population sharing close candidate phenotypes, either for sclareol
306 content and/or early-flowering for oil production, or flower color for ornamental purposes.
307 The recurrent selection and bottlenecks implied have likely played a role in the loss of
308 genetic diversity observed for the two cultivars. In contrast, the recent breeding strategy
309 consists of hybridizations of different genotypes to diversify the genetic basis for *S. sclarea*
310 quality though raising the capacity of essential oil accumulation from the same population.

311 Such new breeding method has resulted in the recent Toscalia cultivar, which, not
312 surprisingly showed high level of genetic diversity, equivalent to the one found in the wild
313 Croatian populations studied here.

314 The lower diversity of traditional clary sage cultivars may reflect a bottleneck during
315 domestication but also a post-domestication erosion [27]. Further sampling of several
316 cultivars from various geographical origins and representing a gradient of domestication,
317 from old and traditional, to recent cultivars, is now required to widen our conclusions.
318 Besides, small populations that underwent bottlenecks thought recurrent selection during
319 domestication can also be expected to show evidence for recent selection in genes involved in
320 key agronomic traits [28]. Here, we did not find any signatures of selection in two genes
321 involved in chemical production pathways, the CMK and DMSX2 genes. Selection could
322 have played a role in other genes involved in the production pathways that were not
323 sequenced in our study, or/and our sampling was too limited. High-throughput sequencing of
324 several varieties with various phenotypes, and association between phenotypes and
325 genotypes [29,30], or scan for selective sweeps along the cultivated and wild clary sage
326 genomes [31], could provide deeper insights in the genomic processes involved during the
327 domestication of clary sage.

328

329 **Phenotypic divergence of the cultivated clary sages depending on their uses, and lack of**
330 **phenotypic differentiation among wild Croatian clary sages**

331

332 We showed a strong phenotypic divergence for the ornamental clary sage cultivar, with an
333 earlier flowering and much lower sclareol and linalyl acetate content than the wild Croatian
334 clary sage populations and the cultivar used for essential oil production. Indeed, Vatican
335 White plants flowered 10 days earlier and produced 20% less sclareol and 30% less linalyl
336 acetate than the cultivar used for oil production. In addition, Vatican White cultivar has light

337 flowers that stand in stark contrast with the flowers of other cultivars used for oil production
338 (e.g., Milly or Toscalia), or of *S. sclarea* wild populations, that display darker shades of pink
339 and purple. In other plant species, the pink or purple color of flowers is often due to the
340 presence of specialized metabolites belonging to the flavonoid family [32]. Traditional
341 breeding methods have therefore implied simultaneous decrease in flavonoid biosynthesis
342 and terpenoid biosynthesis at the origin of the low oil content and different flower color. If
343 this downregulation is operated at transcriptome level, it would be relevant to conduct a
344 comparative transcriptomic study of Milly and Vatican White flower organs. A differential
345 expression of biosynthetic genes would validate the initial hypothesis, and differentially
346 expressed transcription factors would be good candidates for an involvement in specialized
347 metabolism regulation or flowering time control. Since sclareol and linalyl acetate are mainly
348 produced by glandular trichomes, the lower metabolite content observed in Vatican White
349 plants may also be due to a lower glandular trichome density. In *Artemisia annua*, the amount
350 of artemisinin found in leaves is tightly correlated to glandular trichome density [33]. Careful
351 measurements of glandular trichome density are needed to test this hypothesis.

352 In contrast, we showed a lack of variation in metabolite content between the cultivated
353 clary sage used for oil production and the wild clary sage populations. This lack of crop-wild
354 variation for metabolite content suggests that it may not be a trait to pick up in the Croatian
355 wild populations for future breeding programs. Instead, we did observed variation for
356 flowering time among wild clary sage populations. In particular, the Ravca population
357 showed early-flowering phenotypes. This population could be a good target for future
358 breeding program of clary sage used for essential oil production because of their early
359 flowering and relatively high sclareol content, close to the metabolite content found in the
360 cultivar used for essential oil production (Milly).

361 Finally, planted in a common environment, the six wild populations sampled across an
362 environmental gradient did not show significant differences in metabolite content. This result

363 suggests no genetic difference for metabolic contents among wild populations. Therefore,
364 while plants from the wild Croatian populations can be used for breeding programs as a
365 whole regarding their metabolic content, the Ravca population can be used specifically for its
366 early flowering. Additional common gardens including several wild and cultivated clary sage
367 populations in different sites across an altitudinal gradient are now required to guide future
368 clary sage breeding programs. Specific environmental conditions may induce variation yield
369 and chemical composition in metabolite contents, as previously shown in a related species,
370 *Salvia officinalis* [34].

371

372 **Conclusions**

373

374 Clary sage produces sclareol, a terpenoid widely used in perfume industry for the
375 hemisynthesis of ambroxide, a valued perfume component with remarkable fixative
376 properties. In response to the development of ambroxide use in functional perfumery, the
377 global demand for sclareol currently raises, prompting efforts aiming at increasing sclareol
378 supply. Breeding programs aims at increasing the yield of sclareol production from clary
379 sage through targeted genetic improvement. The purpose of our study was to enhance
380 knowledge on phenotypic and genetic variation of wild and cultivated clary sage to provide a
381 first basis for clary sages breeding programs. A larger number of markers and samples of
382 wild and cultivated sages (used for ornamental and oil purposes) across the clary sage
383 distribution is however now required to better understand the clary sage evolution, and
384 provide further directions for breeding programs.

385

386 **Supplementary Materials**

387

388 Table S1. Primers used to amplify genomic fragments from DXS2, CMK and ITS loci to
389 characterize the genetic diversity of *Salvia sclarea*

390

391 Table S2. Summary statistics obtained for the three markers (ITS, CMK, DXS2) for wild
392 and cultivated *Sclarea salvia*,

393

394 **Author Contributions:** Conceptualization, ABo, ABe and MD; methodology, ABo, AC,
395 SD, MD, FG , CCho and CCha; software, AC; validation, CCha and CCho; formal analysis,
396 CCha, CCho, AC, SD, FG; investigation, AC, CCha, SD, CCho, and FG; resources, MD
397 and EMS; data curation, CCha, SD, FG, AC, ABo; writing—original draft preparation, AC,
398 CCha, SD, MD; writing—review and editing, AC, SD, ABo, ABe, MD; supervision, ABo,
399 ABe, MD and SD; funding acquisition, ABo, ABe and MD. All authors have read and
400 agreed to the published version of the manuscript.

401

402 **Acknowledgments:** We thank Bernard Pasquier, former Director of CNPMAI, France, and
403 Jean-Pierre Bouverat-Bernier, former Director of Iteipmai, France, for kindly giving us
404 'Milly-la-Forêt' and 'Toscalia' accessions respectively. We also thank Clémentine Fayol,
405 Alan Walton and Guillaume Frémondrière for growing and phenotyping plants in
406 ITEIPMAI, France. We are grateful for good advice and connexions of Sonya Jakovlev,
407 Paris-Saclay University, France. To finish, thanks to Dubravka Stepić, from the Croatian
408 Ministry of Environment and Energy, to authorize the sampling of wild clary sage in
409 Croatia and for the fruitful discussions about the Nagoya protocol. This research was
410 funded by Plant Biology and Breeding department of INRAE, the French National
411 Research Agency (ANR), and the grants Program LabEx Saclay Plant Sciences-SPS
412 (ANR-10-LABX-40-SPS).

413

414 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role
415 in the design of the study; in the collection, analyses, or interpretation of data; in the writing
416 of the manuscript, or in the decision to publish the results.

417

418 **Figure legends**

419

420 **Figure 1. Genetic diversity and differentiation of the wild and cultivated clary**
421 **sages (*Salvia sclarea*)** ($N=68$ haploid individuals, nine sites, three concatenated
422 markers CMK, ITS, DXS2). **a.** Spatial genetic diversity (π) per site for wild (six
423 sites) and cultivated (three sites) clary sages interpolated across Croatia. The wild
424 Croatian sites are represented by blue circles. As a matter of comparison, the three
425 cultivar populations (red and green dots depending on their uses, for oil or
426 ornamental purposes, respectively) were added on the map to visualize their genetic
427 diversity relatively to the wild populations, but note they were retrieved from
428 repository in France and Germany, and therefore their X and Y coordinates were
429 defined arbitrarily. **b.** Splitstree of wild and cultivated clary sages. Reticulations
430 indicate the occurrence of recombination. The color corresponds to wild (blue) and
431 cultivated (red and green) clary sages **c.** Principal Component Analysis of the
432 genetic variation observed between wild and cultivated sages, using the phased
433 haploid sequences of the CMK, ITS and DSX2 markers. Dots are colored according
434 to the wild (blue) or cultivated (red and green) individual status. Some dots
435 overlapped as some individuals did not show any variation among each other.

436

437 **Figure 2. Phenotypic variation (for flowering time, sclareol and linalyl acetate**
438 **contents) for the wild and cultivated clary sages (*Salvia sclarea*) measured in**
439 **the common garden at Chemillé in France in 2018.** a. Principal component
440 analysis on phenotypic traits for the wild clary sages (blue), the Milly cultivar (red),
441 and the White Vatican cultivar (green) with the photos of the flowers of
442 representative Vatican White and Milly clary sage plants. b, c, d. Date of flowering
443 (expressed in deviation from the mean), linalyl acetate and sclareol contents
444 (expressed in mg/calyx), respectively, for each wild clary sage population (blue),
445 the Milly cultivar (red), and the White Vatican cultivar (green). Numbers indicate
446 the number of samples for each population, letters indicate the result of a
447 Kruskal-Wallis rank test: values harboring the same letter are not significantly
448 different
449

450 **Tables**

451

452 **Table 1. Genetic polymorphism for the wild and cultivated clary sages, *Salvia slarea*, estimated with the concatenated haploid sequences from**
 453 **the ITS 1, DXS2 and CMK loci ($N=68$ haploid individuals).**

Wild/Crop	Environment/Purposes or Use	Site	Longitude	Latitude	Altitude	Nind	Θ_w (standard deviation)	π	Number of haplotypes	S
Wild	Coast	Makarska	17.042	43.286	80	4	0.17 (0.06)	0.21	3	7
		Blato	17.200	43.174	80	3	0.22 (0.08)	0.27	4	8
		Zaostrog	17.283	43.144	80	4	0.29 (0.08)	0.34	6	12
	Mountain	Ravca	17.145	43.236	550	4	0.19 (0.068)	0.18	3	8
		Kljenak	17.254	43.201	360	3	0.25 (0.08)	0.26	4	9
	Plateau	Banja	17.389	43.220	80	4	0.17 (0.064)	0.11	2	7
TOTAL _(wild)						22	0.17 (0.05)	0.31	8	12
Crop	Traditional cultivar, aromatic and perfume purposes	Milly	CNPMAI		NA	4	0.048 (0.034)	0.03	2	2

	Recent breeding origin, aromatic and perfume purposes	Toscalia	ITEIPMAI	NA	4	0.22 (0.072)	0.21	5	9
	Traditional cultivar, ornamental purposes	Vatican White	Jelitto company	NA	4	0.048 (0.034)	0.07	2	2
TOTAL (crop)					12	0.15 (0.05)	0.18	9	9

454 N_{ind} : number of diploid individuals used for population genetics analyses; S: number of polymorphic sites; $\Theta_{(W)}$: Watterson's theta; π , mean
455 standardized pairwise differences; D: Tajima's D; *All Tajima's D were non-significant. CNPMAI: French National Conservatory of
456 Perfume, Medicinal, Aromatic and Industrial Plants, ITEIPMAI: French Interprofessional Technical Institute for Perfume, Medicinal,
457 Aromatic and Industrial Plants.
458

459 **Table 2.** Summary of the effect of the wild site of origin and the use for cultivation of the cultivated clary sage on phenotypic trait variation

460 (Model 1) in *Salvia sclarea*.

Explanatory variable	wild_site_and_cult_use			
	Trait	χ^2	p-value	df
Linalyl acetate		36.777	2.002e-07	4
Scareol content		29.649	5.77e-06	4
Flowering time		45.424	3.245e-09	4

461 df : degree of freedom

462

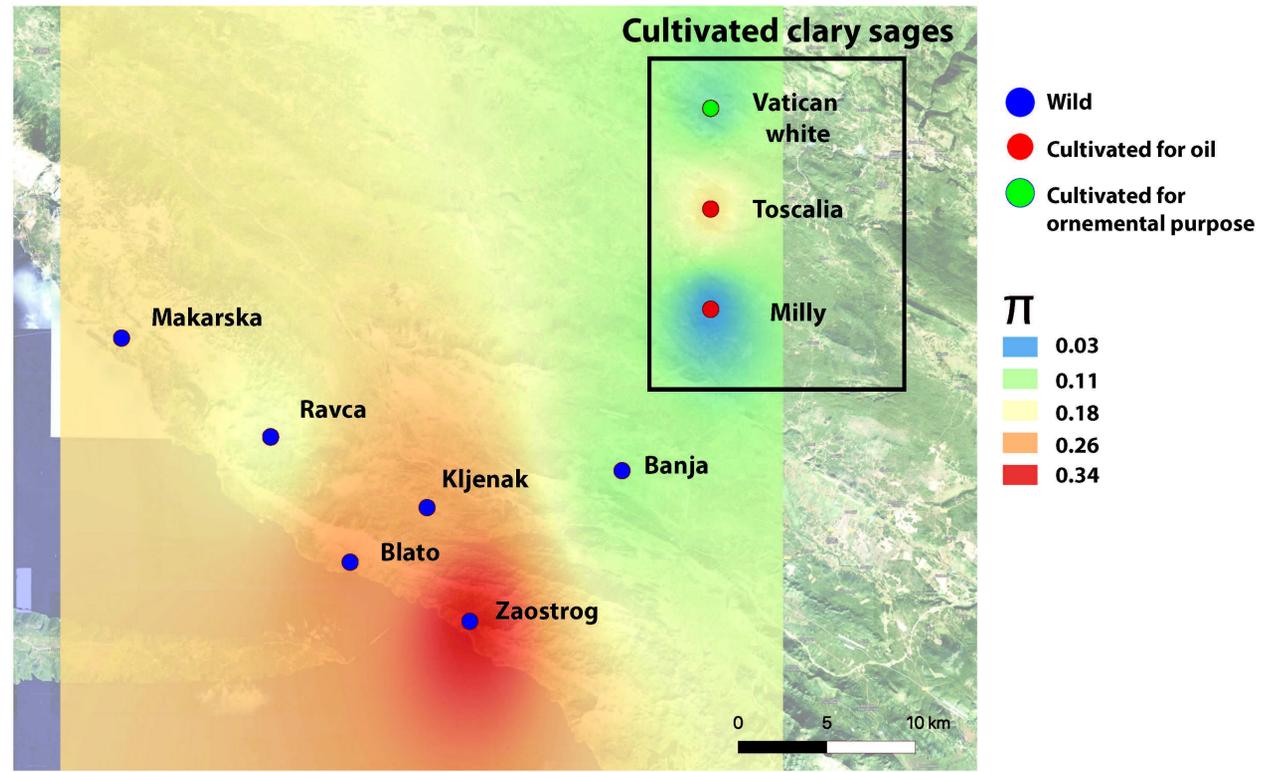
463 **References**

- 464 1. Kilian B, Graner A. NGS technologies for analyzing germplasm diversity in genebanks.
465 Briefings in Functional Genomics. 2012;11: 38–50. doi:10.1093/bfpg/blr046
- 466 2. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H. Back into the wild—Apply
467 untapped genetic diversity of wild relatives for crop improvement. *Evolutionary*
468 *Applications*. 2017;10: 5–24. doi:10.1111/eva.12434
- 469 3. Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H,
470 Eastwood RJ, et al. Global conservation priorities for crop wild relatives. *Nature Plants*.
471 2016;2: 16022. doi:10.1038/nplants.2016.22
- 472 4. Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, Lemainque A, et al. The *Rosa*
473 genome provides new insights into the domestication of modern roses. *Nature Genetics*.
474 2018;50: 772–777. doi:10.1038/s41588-018-0110-3
- 475 5. Otto LG, Mondal P, Brassac J, Preiss S, Degenhardt J, He S, et al. Use of
476 genotyping-by-sequencing to determine the genetic structure in the medicinal plant
477 chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by
478 genome-wide association mapping. *BMC Genomics*. 2017;18: 1–18.
479 doi:10.1186/s12864-017-3991-0
- 480 6. Meyer RS, DuVal AE, Jensen HR. Patterns and processes in crop domestication: an
481 historical review and quantitative analysis of 203 global food crops. *New Phytologist*.
482 2012;196: 29–48. doi:10.1111/j.1469-8137.2012.04253.x
- 483 7. Glémin S, Bataillon T. A comparative view of the evolution of grasses under
484 domestication. *New Phytol*. 2009;183: 273–290.
- 485 8. Figueiredo AC, Barroso JG, Pedro LG, Scheffer JJC. Factors affecting secondary
486 metabolite production in plants: volatile components and essential oils. *Flavour and*
487 *Fragrance Journal*. 2008;23: 213–226. doi:10.1002/ffj.1875
- 488 9. Moore BD, Andrew RL, Külheim C, Foley WJ. Explaining intraspecific diversity in
489 plant secondary metabolites in an ecological context. *New Phytologist*. 2014;201: 733–
490 750. doi:10.1111/nph.12526
- 491 10. Mahzooni-Kachapi SS, Mahdavi M, Jouri MH, Akbarzadeh M, Roozbeh-nasira'ei L.
492 The effects of altitude on chemical compositions and function of essential oils in
493 *stachys lavandulifolia* vahl. (Iran). *International Journal of Medicinal and Aromatic*
494 *Plants*. 2014;4: 107–116.
- 495 11. Foutami IJ, Mariager T, Rinnan R, Barnes CJ, Rønsted N. Hundred Fifty Years of
496 Herbarium Collections Provide a Reliable Resource of Volatile Terpenoid Profiles

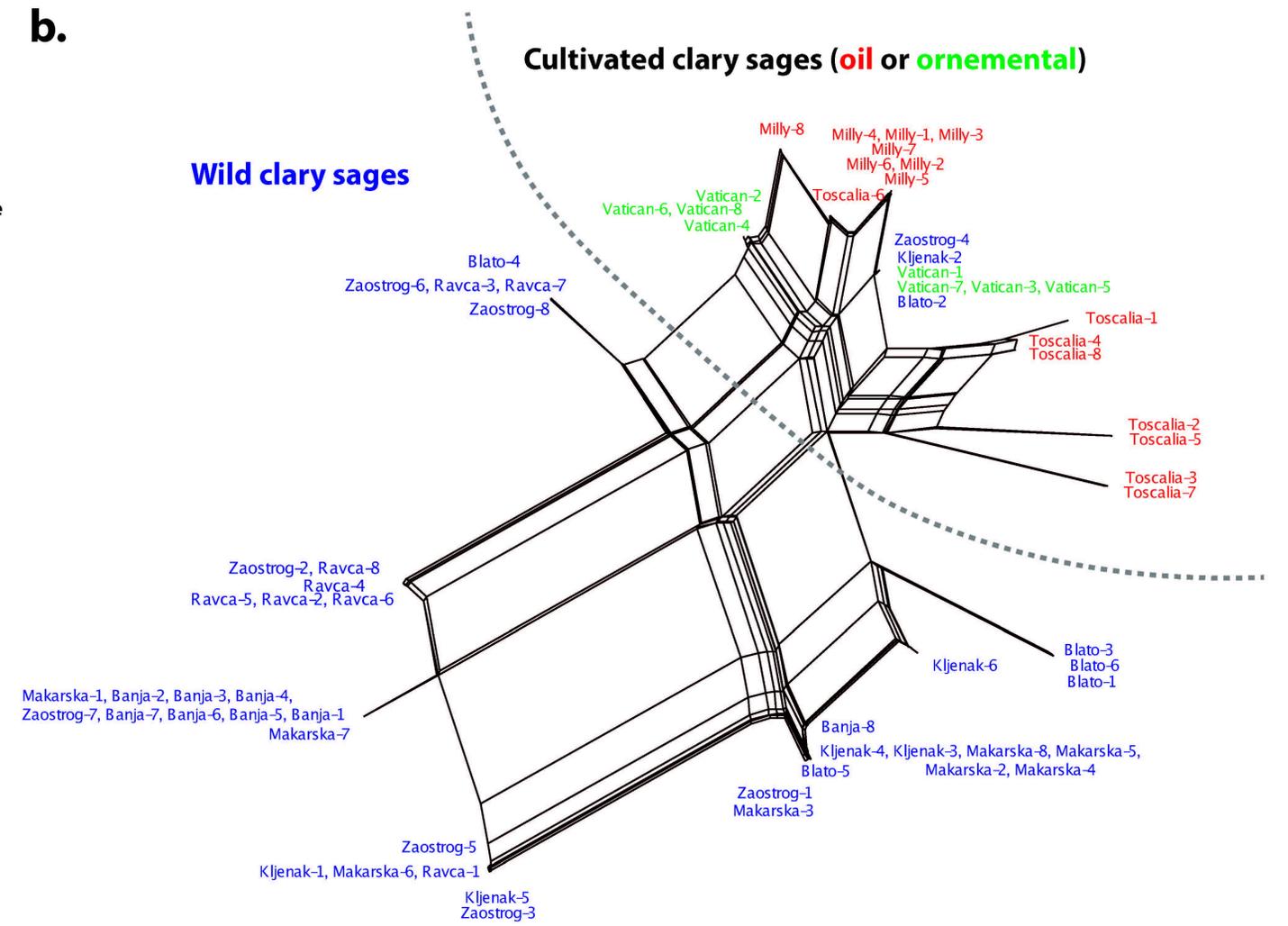
- 497 Showing Strong Species Effect in Four Medicinal Species of *Salvia* Across the
498 Mediterranean. *Frontiers in Plant Science*. 2018;9: 1877. doi:10.3389/fpls.2018.01877
- 499 12. Wagner S, Pflieger A, Mandl M, Böchzelt H. Changes in the Qualitative and
500 Quantitative Composition of Essential Oils of Clary Sage and Roman Chamomile
501 During Steam Distillation in Pilot Plant Scale. *Distillation - Advances from Modeling
502 to Applications*. 2012; 141–158.
- 503 13. Nasermoadeli S, Rowshan V. Comparison of *Salvia sclarea* L . Essential Oil
504 Components in Wild and Field Population. *International Journal of Agriculture and
505 Crop Sciences*. 2013;4: 1997–2000.
- 506 14. Lattoo SK, Dhar RS, Dhar AK, Sharma PR, Agarwal SG. Dynamics of essential oil
507 biosynthesis in relation to inflorescence and glandular ontogeny in *Salvia sclarea*.
508 *Flavour and Fragrance Journal*. 2006;21: 817–821. doi:10.1002/ffj.1733
- 509 15. Schmiderer C, Grassi P, Novak J, Weber M, Franz C. Diversity of essential oil glands of
510 clary sage (*Salvia sclarea* L., Lamiaceae). *Plant Biology*. 2008;10: 433–440.
511 doi:10.1111/j.1438-8677.2008.00053.x
- 512 16. Sharopov FS, Setzer WN. The essential oil of *salvia sclarea* L. from Tajikistan. *Records
513 of Natural Products*. 2012;6: 75–79.
- 514 17. Tholl D. Biosynthesis and Biological Functions of Terpenoids in Plants. *Advances in
515 biochemical engineering/biotechnology*. 2015;123: 127–141. doi:10.1007/10
- 516 18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary
517 Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35: 1547–1549.
518 doi:10.1093/molbev/msy096
- 519 19. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA
520 polymorphism. *Genetics*. 1989;123: 585–595.
- 521 20. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*.
522 *Nature*. 1991;351: 652–654. doi:10.1038/351652a0
- 523 21. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism
524 analyses by the coalescent and other methods. *Bioinformatics*. 2003;19: 2496–2497.
525 doi:10.1093/bioinformatics/btg359
- 526 22. Watterson GA. On the number of segregating sites in genetical models without
527 recombination. *Theor Popul Biol*. 1975;7: 256–276.
528 doi:10.1016/0040-5809(75)90020-9
- 529 23. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies.
530 *Mol Biol Evol*. 2006;23: 254–267. doi:10.1093/molbev/msj030

- 531 24. Dray S, Dufour A-B. The ade4 Package: Implementing the Duality Diagram for
532 Ecologists. *Journal of Statistical Software*; Vol 1, Issue 4 (2007). 2007.
533 doi:10.18637/jss.v022.i04
- 534 25. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
535 data. *Bioinformatics*. 2011;27: 3070–3071. doi:10.1093/bioinformatics/btr521
- 536 26. Laville R, Castel C, Filippi JJ, Delbecq C, Audran A, Garry PP, et al. Amphilectane
537 diterpenes from *Salvia sclarea*: Biosynthetic considerations. *Journal of Natural*
538 *Products*. 2012;75: 121–126. doi:10.1021/np2004177
- 539 27. Allaby RG, Ware RL, Kistler L. A re-evaluation of the domestication bottleneck from
540 archaeogenomic evidence. *Evolutionary Applications*. 2019;12: 29–37.
541 doi:10.1111/eva.12680
- 542 28. Purugganan MD, Fuller DQ. The nature of selection during plant domestication.
543 *Nature*. 2009;457: 843–848.
- 544 29. Brachi B, Morris GP, Borevitz JO. Genome-wide association studies in plants: the
545 missing heritability is in the field. *Genome Biology*. 2011;12: 232.
546 doi:10.1186/gb-2011-12-10-232
- 547 30. Huang X, Han B. Natural Variations and Genome-Wide Association Studies in Crop
548 Plants. *Annual Review of Plant Biology*. 2014;65: 531–551.
549 doi:10.1146/annurev-arplant-050213-035715
- 550 31. Nielsen R. Molecular Signatures of Natural Selection. *Annu Rev Genet*. 2005;39: 197–
551 218. doi:10.1146/annurev.genet.39.073003.112420
- 552 32. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins,
553 betalains and carotenoids. *The Plant Journal*. 2008;54: 733–749.
554 doi:10.1111/j.1365-313X.2008.03447.x
- 555 33. Xiao L, Tan H, Zhang L. *Artemisia annua* glandular secretory trichomes: the biofactory
556 of antimalarial agent artemisinin. *Science Bulletin*. 2016;61: 26–36.
557 doi:10.1007/s11434-015-0980-z
- 558 34. Russo A, Formisano C, Rigano D, Senatore F, Delfino S, Cardile V, et al. Chemical
559 composition and anticancer activity of essential oils of Mediterranean sage (*Salvia*
560 *officinalis* L.) grown in different environmental conditions. *Food and Chemical*
561 *Toxicology*. 2013;55: 42–47. doi:10.1016/j.fct.2012.12.036
- 562 35. He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z. Genotyping-by-sequencing (GBS), an
563 ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in*
564 *Plant Science*. 2014;5: 484. doi:10.3389/fpls.2014.00484
- 565

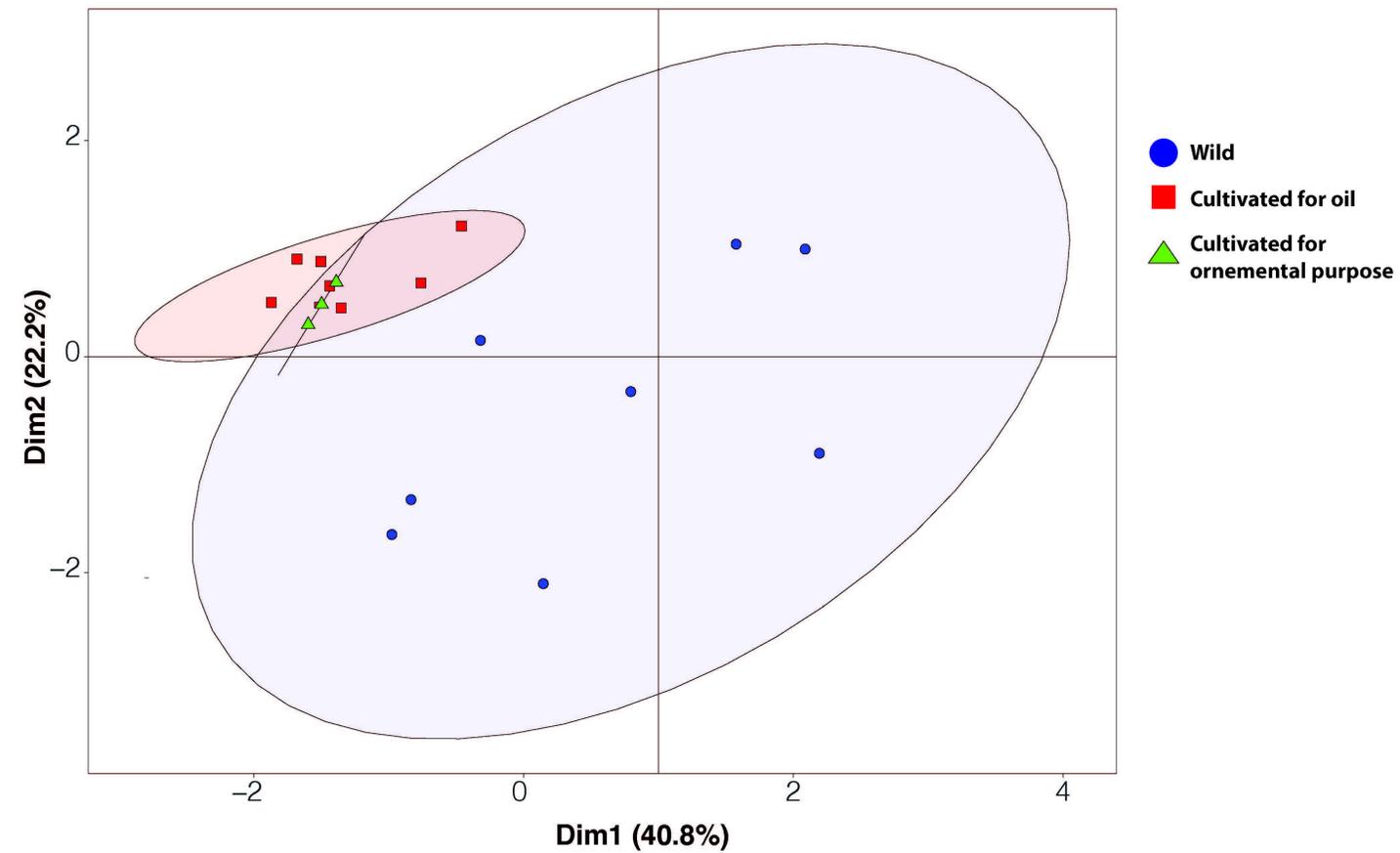
a.



b.

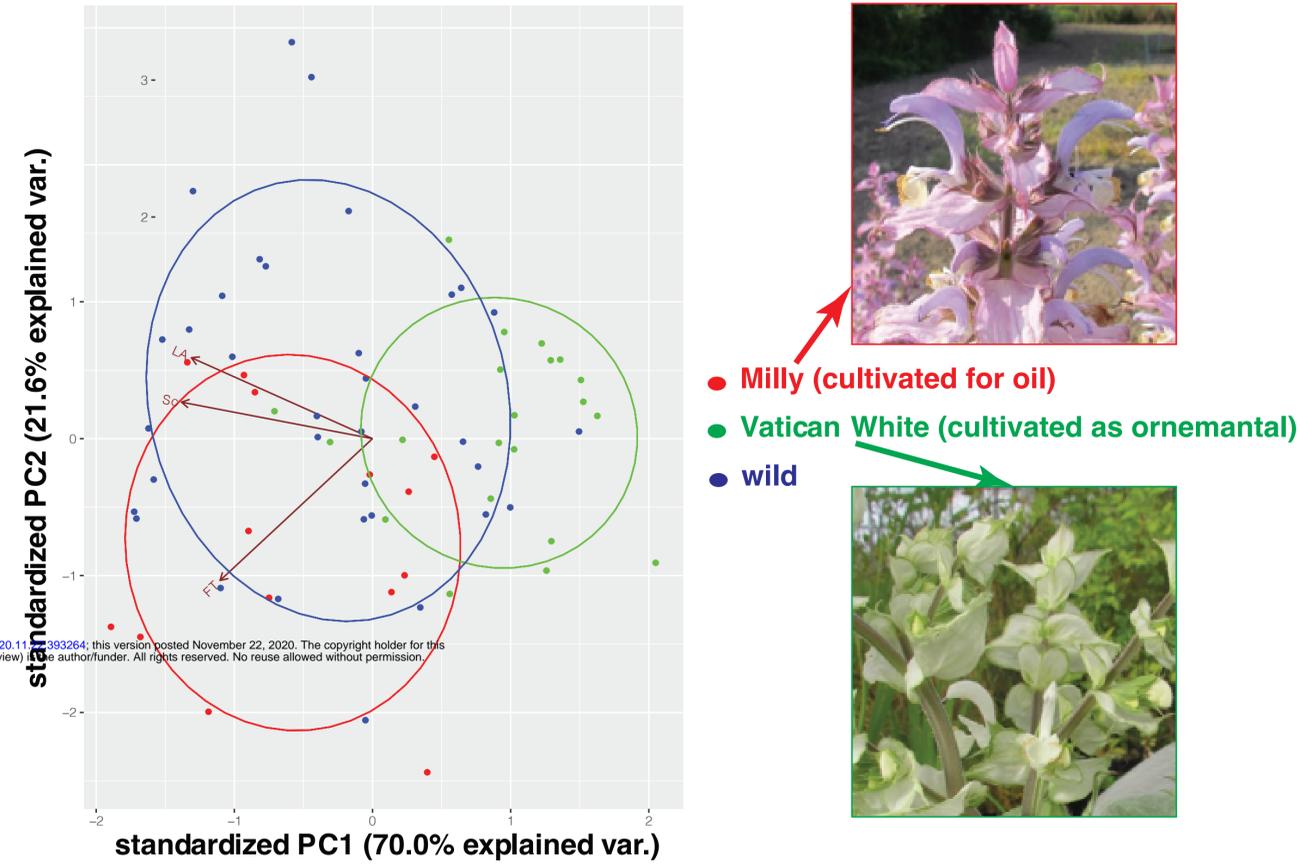


c.

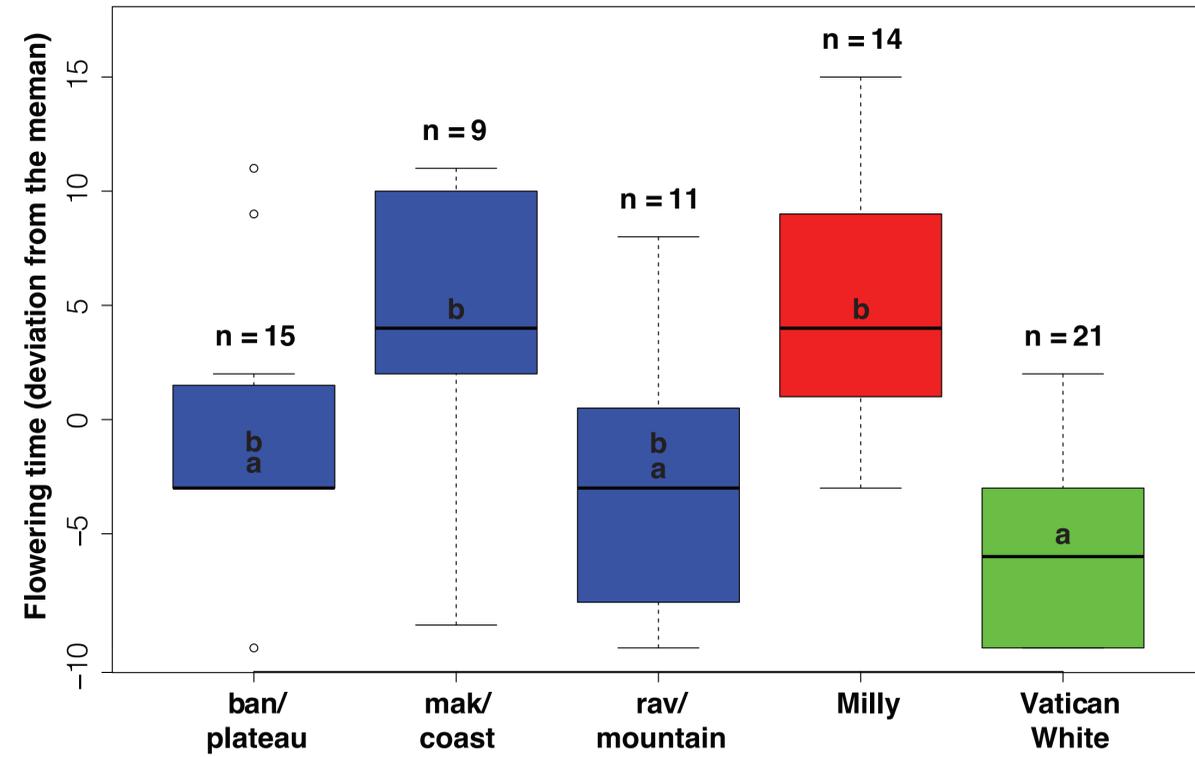


bioRxiv preprint doi: <https://doi.org/10.1101/111111>; this version posted November 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

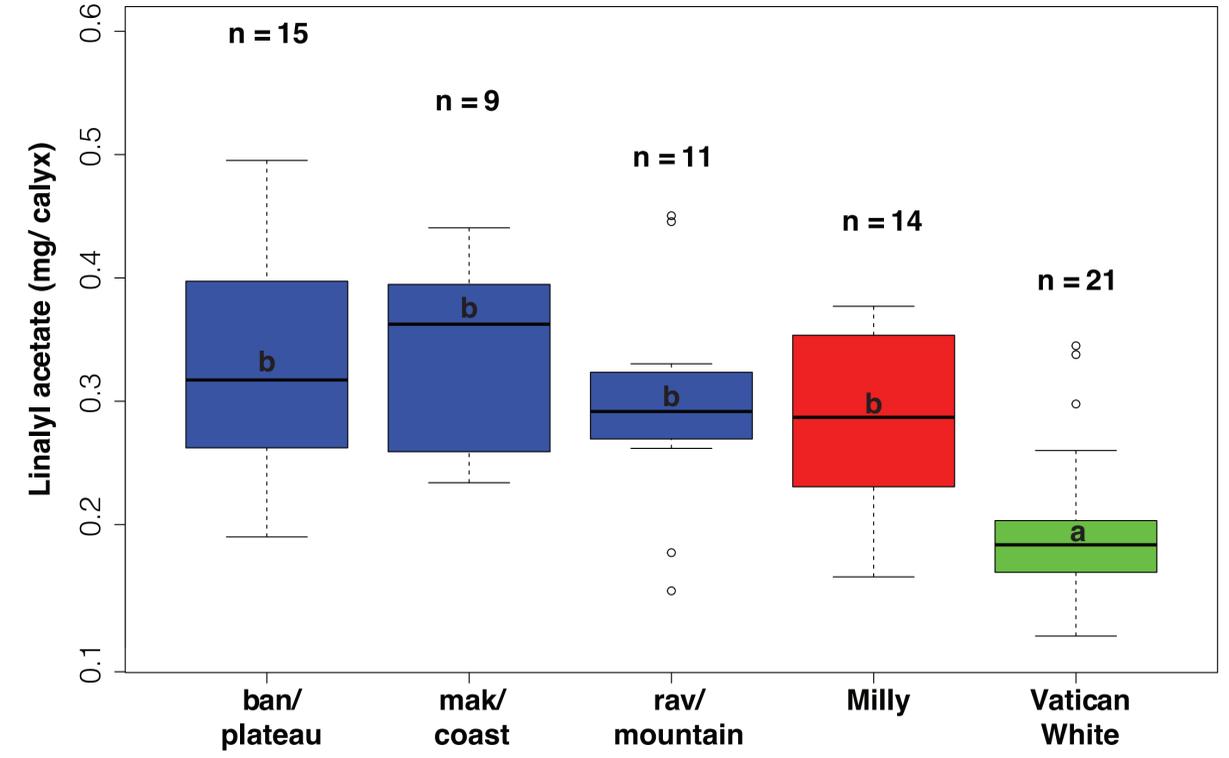
a.



b.



c.



d.

