

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 Chollet ¹, Françoise Gilard¹, Edita M. Šolić²,
6 Michel Dron ¹, Abdelhafid Bendahmane¹, Adnane Boualem¹, Amandine Cornille³

7
8 1. Institute of Plant Sciences Paris- Saclay (IPS2), UMR 9213/UMR 1403, Université

9 Paris- Saclay, 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**

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

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

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

39

40 **Introduction**

41

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

49 Despite the increasing use of medicinal aromatherapy, genetic and phenotypic variation
50 among wild and cultivated perfume, aromatic and medicinal plants (PAMP) are still little
51 explored (but see [4,5]). Plant specialized metabolites, such as alkaloids or terpenes, are the
52 characteristics for which PAMP are valued. Traditional phytochemical studies have invested
53 massive effort to isolate and characterize these chemical components, but only on a handful
54 of elite samples. In nature, PAMP chemical composition showed a huge variation which
55 represents a potential untapped allelic variation for breeding programs. However, the plant
56 specialized metabolite composition of a plant depends on both environmental and genetic
57 factors [6,7]. For instance, altitude potentially influences plant specialized metabolism as a
58 combination of abiotic and biotic factors: temperature, hygrometry, wind speed and water
59 availability change with altitude, promoting the development of distinct microbial and insect
60 communities [8,9]. For targeted genetic enhancement of PAMP performances, breeders need
61 to therefore tease apart the environment and genetic effects at the origin of PAMP chemical
62 variation. Common garden experiments including several genotypes sampled along an
63 altitudinal gradient can help identifying interesting phenotypic variation that has only a
64 genetic origin. Interesting phenotypes can then be selected for PAMP breeding programs in

65 order to increase specialized metabolites yield. Common garden experiments can also be
66 coupled with the characterization of the genetic variation at candidate genes involved in the
67 specialized metabolite production pathways to point out interesting genetic variation, or lack
68 of variation due to recent selection imposed by human. The chemical variation in crop and
69 wild PAMP, and the underlying genes involved, are therefore good targets for breeding
70 programs, but have generally not received a great deal of attention. Genetic study coupled
71 with a phenotypic characterization of chemical variation of the crop and wild PAMP
72 populations are therefore needed.

73 The clary sage, *Salvia sclarea* L., is a biennial or short-lived perennial with a native
74 distribution in Northern Mediterranean, but has a vast geographic distribution towards
75 Central Asia and the Middle East [10,11], and also become a weed outside its native range
76 including North America. Clary sage has been traditionally used as a medicinal plant for its
77 antispasmodic, carminative and estrogenic properties [12], and it is now mainly used for its
78 fragrance. Clary sage essential oil, appreciated for its herbaceous and earthy scent, indeed
79 enters in the composition of numerous high-end fragrances. However, there has been very
80 few breeding concerning clary sage. Only a handful of clary sage cultivars exist for different
81 uses, *i.e.*, oil production or ornamental purposes, and result from traditional or recent
82 breeding methods. Traditional clary sage breeding methods consist of successive cycles of
83 selection and inbreed crosses among individuals from the same population sharing close
84 candidate phenotypes. For oil production, candidate phenotypes show high sclareol and
85 linalool content and, when possible, early flowering. Indeed, the most valued compounds of
86 clary sage is Sclareol, a diterpene extracted from clary sage straw, used as starting material
87 for the synthesis of Ambroxide[®], characterized by a musky odor and scent fixative ability,
88 and linalool, a monoterpene and its acetylated derivative linalyl acetate [12–16]. Early
89 flowering is also a good candidate breeding trait in clary sage. It is widely assumed that the
90 productivity of late-flowering clary sage plants is negatively impacted by summer drought.

91 This rational is borrowed from breeding programs of chamomile, a medicinal plant which
92 also produces valuable terpenoids in flowers [5]. Thereby, the traditional clary sage breeding
93 led to the creation of cultivars used for essential oil production, *e.g.*, the “Milly” cultivar.
94 Clary sage used for oil production has pink to purple flowers, a phenotype which stands in
95 stark contrast with cultivars used for ornamental purposes which shows white flowers, *e.g.*,
96 the “Vatican White” cultivar. More recently, the raising global demand for sclareol in
97 countries where clary sage is cultivated, including Hungary, Bulgaria, Moldavia, Romania
98 and France, has encouraged new breeding initiatives. The recent clary sage breeding
99 approach involves recurrent selection with successive cycles of three steps: intermating
100 between best-performing populations, evaluation of traits of interest, and selection of
101 best-performing individuals. This new breeding method that involves the mixing of several
102 best performing populations led to the creation of a clary sage variety with high sclareol
103 yield, named “Scalia”. However, the “Scalia” variety displays a late-flowering phenotype, an
104 unwanted phenotype. Further recurrent selection cycles led to the development of the
105 “Toscalia” variety, which is characterized by a high sclareol yield without a late-flowering
106 phenotype. While the Toscalia has an attractive phenotype for breeders, it is still not
107 flowering early enough. Breeders are therefore still seeking for candidate populations with
108 early-flowering, and if possible higher sclareol and linalool contents. Wild clary sage
109 populations represent a potential source of untapped allelic variation. However, phenotypic
110 variation for flowering time, sclareol and linalyl acetate contents of *S. sclarea* cultivars used
111 for different purposes, as well as of wild clary sage populations, are still lacking. The genetic
112 relationships among cultivated and wild clary sages, and their respective genetic diversity,
113 also remained unknown.

114 We investigated genetic and phenotypic variation among populations of three clary sage
115 cultivars (two traditional cultivars, the Milly and the White Vatican, and one recent cultivar,
116 the Toscalia) and wild diploid Croatian clary sage populations [15]. The wild clary sage

117 populations were sampled along an altitudinal gradient to screen for potential untapped
118 phenotypic and genetic variation for future breeding programs. Using a genetic dataset
119 (Internal transcribed Spacer 1 marker, referred as to “ITS” hereafter, and two candidate genes
120 involved in the diterpene production pathways [16]), we answered the following questions:
121 do genes involved in the sclareol production pathways show high diversity among wild and
122 cultivated clary sage populations or signature of recent selection? More generally, what is the
123 genetic diversity of wild and cultivated clary sages? What are the genetic relationships
124 among cultivated and wild clary sages? Second, we grew in a common environment in
125 France the wild and cultivated clary sages and characterized their phenotypic variation for
126 sclareol and linalyl acetate contents, as well as their flowering time. We answered the
127 following questions: Is there natural variation in flowering time and sclareol contents among
128 wild and cultivated populations, and among wild populations? Our study had a dual
129 objective: orientating breeding towards promising strategies to boost sclareol yield, and
130 improving our general understanding of the impact of breeding methods of the evolution of a
131 perfume, aromatic and medicinal plant.

132

133 **Materials and Methods**

134

135 **Plant material**

136 During summer 2016, seeds from 60 wild *S. sclarea* individuals were collected at six
137 sites in 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 For each sampled site for the wild *S. sclarea*, and each *S. sclarea* cultivar, three to four
162 individuals were grown in growth chamber (16h of day; 27 °C during the day, 21 °C at night;
163 60% hygrometry) for six weeks. DNA was extracted from leaves using the Qiagen DNeasy
164 96 Plant Kit following manufacturer recommendations. Genomic fragments from *DXS2*,
165 *CMK* and *ITS 1* loci were amplified by PCR using primers listed in Table S1. Sequences of
166 PCR products were obtained through Sanger sequencing performed by Eurofins Genomics.
167 Sequences were aligned using the ClustalW algorithm implemented in the MEGA7 software

168 [17]. Heterozygous sites were visualized in Sanger sequencing chromatograms provided by
169 Eurofins Genomics and manually annotated in MEGA7 using the IUPAC code.

170 We checked the neutrality of the markers to be used for population genetic analyses with
171 *Tajima's D* test [18]. When *Tajima's D* test was non-significant, we assumed the sequences
172 were evolving neutrally. However, *Tajima's D* can be significant in case of bottleneck or
173 expansion, regardless selection. We therefore further test for neutrality of the sequences with
174 the McDonald and Kreitman test (MKT) [19]. The MKT was only possible when a sequence
175 of a *Salvia* outgroup species was available on NCBI for a given marker. The resulting
176 neutrally-evolving sequences were then concatenated. Because of the occurrence for each
177 sequenced marker of heterozygous sites encoded with the IUPAC code, haploid sequences
178 were phased for each individual with the PHASE algorithm implemented the DnaSP
179 software [20]. Haploid sequences obtained (two per individual) were then used for further
180 analyses below.

181 We computed *Watterson's θ* [21] and *Tajima's D* [18] with the DnaSP software [20] for
182 each of the six sites for the wild *S. salvia* and each cultivar (*i.e.*, Vatican White, Toscalia and
183 Milly). We investigated genetic relationships among wild and cultivated samples with a
184 splitstree and a principal component analysis (PCA). For the splitstree, we used the
185 NeighborNet distance transformation and equal angle splits transformation implemented in
186 SplitsTree v4 [22]. For the PCA, we used the `dudi.pca` function from the `ade4` [23] and
187 `adegenet` [24] R packages (R Development CoreTeam, URL <http://www.R-project.org>).

188

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

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

204

205 **Figure 1. Genetic diversity and differentiation of the wild and cultivated clary**
206 **sages (*Salvia sclarea*)** ($N=68$ haploid individuals, nine sites, three concatenated
207 markers CMK, ITS, DXS2). **a.** Spatial genetic diversity (π) per site for wild (six
208 sites) and cultivated (three sites) clary sages interpolated across Croatia
209 (openstreetmap.org; OpenStreetMap contributors). The wild Croatian sites are
210 represented by blue circles. As a matter of comparison, the three cultivar
211 populations (red and green dots depending on their uses, for oil or ornamental
212 purposes, respectively) were added on the map to visualize their genetic diversity
213 relatively to the wild populations, but note they were retrieved from repository in
214 France and Germany, and therefore their X and Y coordinates were defined
215 arbitrarily. **b.** Splitstree of wild and cultivated clary sages. Reticulations indicate the
216 occurrence of recombination. The color corresponds to wild (blue) and cultivated
217 (red and green) clary sages **c.** Principal Component Analysis of the genetic variation

218 observed between wild and cultivated sages, using the phased haploid sequences of
219 the CMK, ITS and DSX2 markers. Dots are colored according to the wild (blue) or
220 cultivated (red and green) individual status. Some dots overlapped as some
221 individuals did not show any variation among each other.

222

223 Clary sage calyces were immersed in hexane without grinding to extract metabolites
224 present at the surface. For each individual, 12 mL of solvent were used to extract metabolites
225 from eight mature calyces. The 10-undecen-1-ol was employed as internal standard and was
226 added in the solvent before extraction at a concentration of 0.75 mM. After 2 h of agitation, 1
227 mL of extract was centrifuged at 13,000 rpm for 10 min to eliminate potential dust or debris.
228 Samples were directly analyzed by GC-MS with a 7890B/5977A GC-MS system from
229 Agilent and according to a protocol adapted from Laville et al. [25]. The system was
230 equipped with a 10 m guard column and a Rxi-5Sil MS column (length, 30 m; inner diameter,
231 0.25 mm; film thickness, 0.25 μ m) (Restek, Bellefonte, PA, USA). A total of 1 μ l of sample
232 was injected with a split ratio of 30:1. Oven temperature was set to 110 $^{\circ}$ C for 2 min, then
233 increased to 270 $^{\circ}$ C at a rate of 10 $^{\circ}$ C/min, and finally set to 270 $^{\circ}$ C for 2 min. Other
234 temperatures were set as follows: injector, 270 $^{\circ}$ C; transfer line, 270 $^{\circ}$ C; source, 230 $^{\circ}$ C;
235 quadrupole, 150 $^{\circ}$ C. The carrier gas was helium at a constant flow of 1 mL/min. The
236 quadrupole mass spectrometer was switched on after a solvent delay of 5 min and was
237 programmed to scan from 35 u to 350 u. Data analysis was carried out using the MassHunter
238 Quantitative Analysis software from Agilent. Sclareol and linalyl acetate were identified by
239 comparison with analytical standards purchased from Sigma-Aldrich. Absolute
240 quantification of sclareol and linalyl acetate was performed using a calibration curve
241 prepared with analytical standards.

242

243 **Statistical analyses**

244 In the presence of a domestication syndrome, we expect differences in phenotypes of
245 cultivated accessions compared to wild accessions. We can also expect phenotypic
246 differences between the cultivated clary sages depending on their uses (ornamental or oil
247 production), and between wild *S. sclarea* plants sampled at sites along an altitudinal gradient
248 in Croatia. Indeed, the selective pressures associated with the three environmental
249 conditions from which the three populations were sampled (Makarska, Coast; Ravca,
250 mountains; Banja, plateau) may have led to natural variation in phenotypes. In order to
251 study the influence of the use of the cultivated sages and the wild status of each plant, we
252 fitted a generalized linear model as follows:

253

$$254 \text{ Trait}_{ij} \sim \mu + \text{wild_site_or_cult_use}_i + e_{ij} \quad (1)$$

255

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

263

264 **Results**

265

266 **Genetic diversity and relationships among cultivated and wild**

267 **clary sages**

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

276 The splitstree and the PCA (Figure 1, c and d) revealed a genetic differentiation between
277 wild and cultivated clary sages. Within the cultivated clary sages, the Vatican white showed
278 less variation than the Milly and Toscalia cultivar, and, the Toscalia were more differentiated
279 than the Milly and the White Vatican cultivars.

280

281 **Phenotypic variation between wild and cultivated clary sages,** 282 **and within each group**

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

288 Overall, we observed significant differences in flowering time, sclareol and linalyl
289 acetate among wild clary sages, clary sages cultivated for ornamental purposes and for oil
290 production (Figure 2a and Table 2). For wild clary sages, no significant differences in
291 sclareol and linalyl acetate of calyces were observed among plants coming from coast,
292 mountain or plateau (Figure 2, c and d). Significant differences were observed for flowering

293 time: the Makarska/coastal plants flowering in average 5.044 days later than other wild plants
294 ($P=0.046$, Figure 2b). When comparing the cultivated and wild plants, the Vatican White
295 plants, used for ornamental purposes, produced significantly less linalyl acetate (30% less)
296 than the Milly cultivar and the wild clary sage plants (Figure 2c). The Vatican White plants
297 also produced significantly less sclareol (-0.12 mg/calyx in average, $P<0.0001$, Figure 2d)
298 compared to the Milly cultivar and the Makarska/coast and Banja/plateau wild plants.
299 Vatican White plants flowered significantly earlier than the Milly cultivar and the wild
300 Makarska plants (10 days earlier in average) (Figure 2b)

301

302 **Figure 2. Phenotypic variation (for flowering time, sclareol and linalyl acetate**
303 **contents) for the wild and cultivated clary sages (*Salvia sclarea*) measured in**
304 **the common garden at Chemillé in France in 2018.** a. Principal component
305 analysis on phenotypic traits for the wild clary sages (blue), the Milly cultivar (red),
306 and the White Vatican cultivar (green) with the photos of the flowers of
307 representative Vatican White and Milly clary sage plants. b, c, d. Date of flowering
308 (expressed in deviation from the mean), linalyl acetate and sclareol contents
309 (expressed in mg/calyx), respectively, for each wild clary sage population (blue),
310 the Milly cultivar (red), and the White Vatican cultivar (green). Numbers indicate
311 the number of samples for each population, letters indicate the result of a
312 Kruskal-Wallis rank test: values harboring the same letter are not significantly
313 different

314

315 Discussion

316 We provided a first insight into crop and wild genetic and phenotypic variation of a
317 perfume, medicinal and aromatic plant, the clary sage. Our results showed a lower level of

318 genetic diversity for the traditional clary sage cultivars, whatever their uses, compared with
319 the Croatian wild populations. In contrast, the recent Toscalia cultivar, resulting from new
320 breeding methods, showed a high genetic variation, not significantly different from the
321 Croatian wild populations. Phenotypically, a strong divergence is observed for the
322 ornamental clary sage compared with the two cultivars used for essential oil production.
323 Interestingly, the cultivars used for essential oil production were not significantly
324 differentiated from the Croatian wild populations, whatever the area/altitude of origin of the
325 wild population. We also did not observe phenotypic variation among wild Croatian clary
326 sage populations. Our results finally pinpoint certain Croatian wild populations with
327 interesting sclareol and linalyl acetate contents for future clary sage breeding.

328

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

331 We did find a lower genetic diversity for the traditional cultivars used for essential oil
332 production (Milly) and for ornamental purposes (Vatican White). Traditional clary sage
333 breeding methods consist of successive cycles of selection and inbred crosses among
334 individuals from the same population sharing close candidate phenotypes, either for sclareol
335 content and/or early-flowering for oil production, or flower color for ornamental purposes.
336 The recurrent selection and bottlenecks implied by traditional breeding methods have likely
337 played a role in the loss of genetic diversity observed for the two cultivars. In contrast, the
338 recent breeding strategy consists of hybridizations of different genotypes to diversify the
339 genetic basis for *S. sclarea* quality though raising the capacity of essential oil accumulation
340 from the same population. Such a new breeding method has resulted in the recent Toscalia
341 cultivar, which, not surprisingly showed high level of genetic diversity, equivalent to the one
342 found in the wild Croatian populations studied here.

343 The lower diversity of traditional clary sage cultivars may reflect a bottleneck during
344 domestication but also a post-domestication erosion [26]. Further sampling of several
345 cultivars from various geographical origins and representing a gradient of domestication,
346 from old and traditional, to recent cultivars, is now required to widen our conclusions.
347 Besides, small populations that underwent bottlenecks thought recurrent selection during
348 domestication can also be expected to show evidence for recent selection in genes involved in
349 key agronomic traits [27]. Here, we did not find any signature of selection in two genes
350 involved in chemical production pathways, the CMK and DMSX2 genes. Selection could
351 have played a role in other genes involved in the production pathways that were not
352 sequenced in our study, or/and, our sampling was too limited. High-throughput sequencing
353 of several varieties with various phenotypes, and association between phenotypes and
354 genotypes [28,29], or scan for selective sweeps along the cultivated and wild clary sage
355 genomes [30], could provide deeper insights in the genomic processes involved during the
356 domestication of clary sage.

357

358 **Phenotypic divergence of the cultivated clary sages depends on**
359 **their uses, and lack of phenotypic differentiation among wild**
360 **Croatian clary sages**

361 We showed a strong phenotypic divergence for the ornamental clary sage cultivar, with
362 an earlier flowering and much lower sclareol and linalyl acetate content than the wild
363 Croatian clary sage populations and the cultivar used for essential oil production. Indeed,
364 Vatican White plants flowered 10 days earlier and produced 20% less sclareol and 30% less
365 linalyl acetate than the cultivar used for oil production. In addition, Vatican White cultivar
366 has light flowers that stand in stark contrast with the flowers of other cultivars used for oil
367 production (*e.g.*, Milly or Toscalia), or of *S. sclarea* wild populations, that display darker

368 shades of pink and purple. In other plant species, the pink or purple color of flowers is often
369 due to the presence of specialized metabolites belonging to the flavonoid family [31].
370 Traditional breeding methods have therefore implied simultaneous decrease in flavonoid
371 biosynthesis and terpenoid biosynthesis at the origin of the low oil content and different
372 flower color. If this downregulation is operated at transcriptome level, it would be relevant to
373 conduct a comparative transcriptomic study of Milly and Vatican White flower organs. A
374 differential expression of biosynthetic genes would validate the initial hypothesis, and
375 differentially expressed transcription factors would be good candidates for an involvement in
376 specialized metabolism regulation or flowering time control. Since sclareol and linalyl
377 acetate are mainly produced by glandular trichomes, the lower metabolite content observed
378 in Vatican White plants may also be due to a lower glandular trichome density. In *Artemisia*
379 *annua*, the amount of artemisinin found in leaves is tightly correlated to glandular trichome
380 density [32]. Careful measurements of glandular trichome density are needed to test this
381 hypothesis.

382 In contrast, we showed a lack of variation in metabolite content between the cultivated
383 clary sage used for oil production and the wild clary sage populations. This lack of crop-wild
384 variation for metabolite content suggests that it may not be a trait to pick up in the Croatian
385 wild populations for future breeding programs. Instead, we did observed variation for
386 flowering time among wild clary sage populations. In particular, the Ravca population
387 showed early-flowering phenotypes. This population could be a good target for future
388 breeding program of clary sage used for essential oil production because of their early
389 flowering and relatively high sclareol content, close to the metabolite content found in the
390 cultivar used for essential oil production (Milly). Finally, planted in a common environment,
391 the six wild populations sampled across an environmental gradient did not show significant
392 differences in metabolite content. This result suggests no genetic difference for metabolic
393 contents among wild populations. Such a lack of variation can be explained by the narrow

394 geographic area sampled. In conclusion, while plants from the wild Croatian populations can
395 be used for breeding programs as a whole regarding their metabolic content, the Ravca
396 population can be used specifically for its early flowering. Additional common gardens
397 including several wild and cultivated clary sage populations in different sites across an
398 altitudinal gradient are now required to guide future clary sage breeding programs. Specific
399 environmental conditions may induce variation yield and chemical composition in
400 metabolite contents, as previously shown in a related species, *Salvia officinalis* [33].

401

402 **Conclusions**

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

414

415 **Supporting Information**

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

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

420

421 **Author Contributions:** Conceptualization, ABo, ABe and MD; methodology, ABo, AC,
422 SD, MD, FG, CCho and CCha; software, AC; validation, CCha and CCho; formal analysis,
423 CCha, CCho, AC, SD, FG; investigation, AC, CCha, SD, CCho, and FG; resources, MD
424 and EMS; data curation, CCha, SD, FG, AC, ABo; writing—original draft preparation, AC,
425 CCha, SD, MD; writing—review and editing, AC, SD, ABo, ABe, MD; supervision, ABo,
426 ABe, MD and SD; funding acquisition, ABo, ABe and MD. All authors have read and
427 agreed to the published version of the manuscript.

428

429 **Acknowledgments:** We thank Bernard Pasquier, former Director of CNPMAI, France, and
430 Jean-Pierre Bouverat-Bernier, former Director of Iteipmai, France, for kindly giving us
431 'Milly-la-Forêt' and 'Toscalia' accessions respectively. We also thank Clémentine Fayol,
432 Alan Walton and Guillaume Frémondrière for growing and phenotyping plants in
433 ITEIPMAI, France. We are grateful for good advice and connexions of Sonya Jakovlev,
434 Paris-Saclay University, France. We also thank Dubravka Stepić, from the Croatian
435 Ministry of Environment and Energy, to authorize the sampling of wild clary sage in
436 Croatia and for the fruitful discussions about the Nagoya protocol.

437

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

441 **Data Availability**

442 Phenotypic data have been submitted on DRYAD: 10.5061/dryad.n02v6www7, and
443 sequences on NCBI (*rRNA-ITS*: MW345612, *DXS2* : MW431009, *CMK* : MW431010).

444 **Tables**

445

446 **Table 1. Genetic polymorphism for the wild and cultivated clary sages, *Salvia slarea*, estimated with the concatenated haploid sequences from**
 447 **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

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

452

453 **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
454 (Model 1) in *Salvia sclarea*.

Explanatory variable	Effect : Wild sites and cultivar 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

455 df : degree of freedom

456

457 **References**

- 458 1. Kilian B, Graner A. NGS technologies for analyzing germplasm diversity in genebanks.
459 Briefings in Functional Genomics. 2012;11: 38–50. doi:10.1093/bfpg/blr046
- 460 2. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H. Back into the wild—Apply
461 untapped genetic diversity of wild relatives for crop improvement. *Evolutionary*
462 *Applications*. 2017;10: 5–24. doi:10.1111/eva.12434
- 463 3. Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H,
464 Eastwood RJ, et al. Global conservation priorities for crop wild relatives. *Nature Plants*.
465 2016;2: 16022. doi:10.1038/nplants.2016.22
- 466 4. Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, Lemainque A, et al. The *Rosa*
467 genome provides new insights into the domestication of modern roses. *Nature Genetics*.
468 2018;50: 772–777. doi:10.1038/s41588-018-0110-3
- 469 5. Otto LG, Mondal P, Brassac J, Preiss S, Degenhardt J, He S, et al. Use of
470 genotyping-by-sequencing to determine the genetic structure in the medicinal plant
471 chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by
472 genome-wide association mapping. *BMC Genomics*. 2017;18: 1–18.
473 doi:10.1186/s12864-017-3991-0
- 474 6. Figueiredo AC, Barroso JG, Pedro LG, Scheffer JJC. Factors affecting secondary
475 metabolite production in plants: volatile components and essential oils. *Flavour and*
476 *Fragrance Journal*. 2008;23: 213–226. doi:10.1002/ffj.1875
- 477 7. Moore BD, Andrew RL, Külheim C, Foley WJ. Explaining intraspecific diversity in
478 plant secondary metabolites in an ecological context. *New Phytologist*. 2014;201:
479 733–750. doi:10.1111/nph.12526
- 480 8. Mahzooni-Kachapi SS, Mahdavi M, Jouri MH, Akbarzadeh M, Roozbeh-nasira’ei L.
481 The effects of altitude on chemical compositions and function of essential oils in *Sachys*
482 *savandulifolia* vahl. (Iran). *International Journal of Medicinal and Aromatic Plants*.
483 2014;4: 107–116.
- 484 9. Foutami IJ, Mariager T, Rinnan R, Barnes CJ, Rønsted N. Hundred Fifty Years of
485 Herbarium Collections Provide a Reliable Resource of Volatile Terpenoid Profiles
486 Showing Strong Species Effect in Four Medicinal Species of *Salvia* Across the
487 Mediterranean. *Frontiers in Plant Science*. 2018;9: 1877. doi:10.3389/fpls.2018.01877
- 488 10. Wagner S, Pflieger A, Mandl M, Böchzelt H. Changes in the Qualitative and
489 Quantitative Composition of Essential Oils of Clary Sage and Roman Chamomile
490 During Steam Distillation in Pilot Plant Scale. *Distillation - Advances from Modeling*
491 *to Applications*. 2012; 141–158.

- 492 11. Nasermodeli S, Rowshan V. Comparison of *Salvia sclarea* L . Essential Oil
493 Components in Wild and Field Population. International Journal of Agriculture and
494 Crop Sciences. 2013;4: 1997–2000.
- 495 12. Lattoo SK, Dhar RS, Dhar AK, Sharma PR, Agarwal SG. Dynamics of essential oil
496 biosynthesis in relation to inflorescence and glandular ontogeny in *Salvia sclarea*.
497 Flavour and Fragrance Journal. 2006;21: 817–821. doi:10.1002/ffj.1733
- 498 13. Schmiderer C, Grassi P, Novak J, Weber M, Franz C. Diversity of essential oil glands of
499 clary sage (*Salvia sclarea* L., Lamiaceae). Plant Biology. 2008;10: 433–440.
500 doi:10.1111/j.1438-8677.2008.00053.x
- 501 14. Sharopov FS, Setzer WN. The essential oil of *salvia sclarea* L. from Tajikistan. Records
502 of Natural Products. 2012;6: 75–79.
- 503 15. Siljak-Yakovlev S, Pustahija F, Šolić EM, Bogunić F, Muratović E, Bašić N, et al.
504 Towards a Genome Size and Chromosome Number Database of Balkan Flora:
505 C-Values in 343 Taxa with Novel Values for 242. Advanced Science Letters. 2010;3:
506 190–213. doi:10.1166/asl.2010.1115
- 507 16. Tholl D. Biosynthesis and Biological Functions of Terpenoids in Plants. Advances in
508 biochemical engineering/biotechnology. 2015;123: 127–141. doi:10.1007/10
- 509 17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary
510 Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35: 1547–1549.
511 doi:10.1093/molbev/msy096
- 512 18. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA
513 polymorphism. Genetics. 1989;123: 585–595.
- 514 19. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*.
515 Nature. 1991;351: 652–654. doi:10.1038/351652a0
- 516 20. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism
517 analyses by the coalescent and other methods. Bioinformatics. 2003;19: 2496–2497.
518 doi:10.1093/bioinformatics/btg359
- 519 21. Watterson GA. On the number of segregating sites in genetical models without
520 recombination. Theor Popul Biol. 1975;7: 256–276.
521 doi:10.1016/0040-5809(75)90020-9
- 522 22. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies.
523 Mol Biol Evol. 2006;23: 254–267. doi:10.1093/molbev/msj030
- 524 23. Dray S, Dufour A-B. The *ade4* Package: Implementing the Duality Diagram for
525 Ecologists. Journal of Statistical Software; Vol 1, Issue 4 (2007). 2007.
526 doi:10.18637/jss.v022.i04

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