bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Running title:** Saffron evolution

2

3

4 # Saffron (*Crocus sativus*) is an autotriploid that evolved in Attica
5 # (Greece) from wild *Crocus cartwrightianus*

6

7 **Zahra Nemati,[1,a] Dörte Harpke,[1,a] Almila Gemicioglu,[1,2] Helmut Kerndorff,[1] and Frank R.**
8 **Blattner[1,*]**

9

10 [1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben,
11 Germany.

12 [2]Dept. of Biology, University of Istanbul, Istanbul, Turkey.

13

14 [a]These authors contributed equally to the study

15

16 ***For correspondence:** E-mail: blattner@ipk-gatersleben.de
17 Frank R. Blattner
18 Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
19 Corrensstr. 3
20 D-06466 Seeland/Gatersleben
21 +49 39482 5202

22

23

24 **Word counts**
25 Abstract: 149
26 Introduction: 521
27 Results and discussion: 2073
28 Materials and methods: 2039

29

30 Figures: 2
31 Tables: 1
32 Supporting information: 1 PDF

33

34

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35 ## Abstract

36 ***Crocus sativus* is the source of saffron, which is made from dried stigmas of the plant.**
37 **It is a male-sterile triploid that ever since its origin has been propagated vegetatively.**
38 **The mode of evolution and area of origin of saffron are matters of long-lasting**
39 **debates. Here we analyzed chloroplast genomes, genotyping-by-sequencing (GBS)**
40 **data, nuclear single-copy genes, and genome sizes to solve these controversial**
41 **issues. We could place 99.3% of saffron GBS alleles in *Crocus cartwrightianus*, a**
42 **species occurring in southern mainland Greece and on Aegean islands, identifying it**
43 **as the sole progenitor of saffron. Phylogenetic and population assignment analyses**
44 **together with chloroplast polymorphisms indicated the wild *C. cartwrightianus***
45 **population south of Athens as most similar to *C. sativus*. We conclude that the crop is**
46 **an autotriploid that evolved in Attica by combining two different genotypes of *C.***
47 ***cartwrightianus*. Vegetative propagation prevented afterwards segregation of the**
48 **favorable traits of saffron.**

49

50 ***Key words:*** **autotriploidy, *Crocus*, crop evolution, domestication, genotyping-by-**
51 **sequencing, saffron**

52

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

53 ## Introduction

54 Saffron is the most expensive spice worldwide and is used since ancient times for its aroma
55 and the ability to give dishes and textiles a golden-yellow hue (Negbi, 1999). During the last
56 decades also medicinal properties of the plant became of interest (Abdullaev and Frenkel,
57 1999). The use of saffron was already documented in 3600-year-old Minoan frescos from the
58 southern Aegean islands Crete and Santorini. At that time the plants were probably not *C.*
59 *sativus* but belonged to wild saffron, *C. cartwrightianus* (Negbi and Negbi, 2002). While this
60 latter species is an obligate outbreeding diploid, cultivated saffron is a male-sterile triploid.
61 Both species are morphologically similar, but the stigma of the saffron crocus is thought to be
62 longer and of darker color than that of wild saffron and the aroma is more pronounced in the
63 cultivar. The overall similarity between both taxa was sometimes the reason to assume that
64 *C. cartwrightianus* is the progenitor of the saffron crocus (Matthew, 1999; Negbi and Negbi,
65 2002). However, there are other similar species and molecular data were inconsistent
66 regarding the contribution of possible parental taxa. Although *C. cartwrightianus* has been
67 postulated several times as parent or one of the parental species, also *C. almehensis*, *C.*
68 *hadriaticus*, *C. haussknechtii*, *C. mathewii*, *C. michelsonii*, *C. pallasii*, *C. thomasii* and *C.*
69 *serotinus* have been proposed as possible parents (see Nemati et al., 2018). However, none
70 of the analyses conducted to date have allowed for the safe inference of the parent(s), partly
71 due to study designs that did not include the relevant species, and partly due to
72 methodological shortcomings (Nemati et al., 2018). Additionally, the mode of evolution of
73 triploid *C. sativus*, i.e. if it originated through autopolyploidization from a single progenitor
74 (Brighton, 1977; Ghaffari, 1986) or allopolyploidization involving two parental species
75 (Tsaftaris et al., 2011; Harpke et al., 2013) has been debated. In a recent study inferring
76 phylogenetic relationships of the species of *Crocus* series *Crocus*, that is the taxonomic
77 group to which *C. sativus* belongs, we found that *C. cartwrightianus* is the closest relative of
78 the saffron crocus and hypothesized that no other species might have contributed to the
79 formation of the triploid (Nemati et al., 2018).

80 Here we follow up on this hypothesis by analyzing genome-wide single-nucleotide
81 polymorphisms (SNP) to see *(i)* if all alleles present in *C. sativus* can be detected within *C.*
82 *cartwrightianus* or if a fraction of the saffron alleles might not be derived from this species.
83 This should allow us to discern an auto- from an allopolyploid origin of saffron. *(ii)* SNP data
84 are also used to find populations or areas where genetic similarity of the progenitor(s)
85 towards *C. sativus* is highest. Depending on the species' genetic structure, this should allow
86 identifying the region where *C. sativus* originated. *(iii)* We test the results of SNP data by
87 analyzing diversity of chloroplast genomes, which provide an independent source of
88 population data. (*iv*) To understand why analyses of potential progenitor species of the
89 saffron crocus resulted up to now in widely contradicting results we analyze allele diversity at
90 five nuclear single-copy genes in a group of species closely related to *C. sativus*.

91

92

93 ## Results and discussion

94 ### Analyses of genome-wide SNP data

95 To see if alleles present in *C. sativus* can all be traced back to its putative progenitor *C.*
96 *cartwrightianus* and to infer the geographic location of the origin of saffron, we analyzed
97 genome-wide SNP data obtained via genotyping-by-sequencing (GBS; Elshire et al., 2011).
98 We first processed 22 saffron individuals within the IPYRAD analysis pipeline (Eaton, 2014) to
99 detect the loci common for 85% of the individuals, i.e. assembling a GBS reference of saffron

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

100 that omits the majority of somatic present/absent mutations that might have accumulated
101 through time within this clonal lineage as well as loci with too low or high coverage or
102 indications that loci are not occurring in single-copy state. This reference, consisting of 6512
103 GBS loci (of which 1768 were heterozygous in saffron), was then used to evaluate how many
104 of the saffron GBS alleles occur in *C. cartwrightianus* and *C. oreocreticus*. Here we
105 consistently used 7–10 individuals out of ten *C. cartwrightianus* populations from the species'
106 entire distribution area (Figure 1A) and four individuals each out of two *C. oreocreticus*
107 populations. We could place up to 90.47% of the GBS alleles occurring in saffron to the *C.*
108 *cartwrightianus* populations analyzed, while 76.83% occur in *C. oreocreticus.* For the set of
109 all 96 *C. cartwrightianus* individuals the proportion of saffron alleles was 98.58% (table 1).

110 As *C. sativus* individuals are genetically very similar (Busconi et al., 2015) and this is also the
111 case for the individuals we studied, we included ten of them in phylogenetic analyses based
112 on the GBS dataset. In all analyses the samples of *C. sativus* grouped within *C.*
113 *cartwrightianus* populations from the northern part of the species' distribution range and
114 formed the sister group of the Attic individuals (Figure 1B, Figure S1 and S2; figures and
115 tables indicated by "S" are available as Supplementary Files online). GBS-based genetic
116 distances in *C. cartwrightianus* are an order of magnitude higher than in saffron (up to 0.41%
117 vs. 0.032%), and also within the *C. cartwrightianus* populations genetic diversity is
118 comparatively high (e.g., for Attica 0.29%), This is also reflected in the respective branch
119 lengths of the phylogenetic trees (Figure S1 and S2). Differences in amount of genetic
120 diversity between both species were also reported by Larsen et al. (2015). For a Bayesian
121 population assignment analysis (Pearse and Crandall, 2004) we identified an optimum of
122 nine groups of genotypes (K = 9) including *C. oreocreticus*. Here saffron is assigned to the
123 individuals from Attica (Figure 1C, Figure S3). However, we did not find a single *C.*
124 *cartwrightianus* individual that is identical, or even most similar, to saffron but could only
125 identify the entire Attica population as closest relatives of the cultivar, possessing 90.47% of
126 the saffron GBS alleles. This correlation also holds after correction for sample size (Table 1).

127 To see if all alleles occurring in *C. sativus* can be found in Attic *C. cartwrightianus* individuals
128 we increased our sample from Attica to 71 individuals, which we genotyped by GBS. We
129 could place 97.60% of the saffron GBS alleles in this set of Attic samples and together with
130 the individuals from the other areas they accounted for 99.31% of the saffron alleles (Table
131 1). Thus, only 0.69% of them could not be assigned to any of the included *C. cartwrightianus*
132 individuals. This clearly indicates that only wild saffron contributed genetic material to saffron.
133 Otherwise a much higher proportion of saffron GBS alleles should occur that are different
134 from the ones found in *C. cartwrightianus*.

135

136 **Analysis of chloroplast genome diversity**

137 BLAST searches revealed that none of the 6512 GBS loci of saffron we used were derived
138 from the chloroplast or mitochondrial genome. Therefore, we analyzed DNA differences in
139 the maternally inherited chloroplast to base our conclusions also on a marker type that is
140 independent from the nuclear GBS data. To obtain initial information about potentially
141 informative loci we used genome skimming (Straub et al., 2012), which is based on low-
142 coverage whole-genome shotgun (WGS) sequencing, and assembled the chloroplast
143 genomes (Figure S4) of two *C. cartwrightianus* individuals from the southern (Crete) and
144 northern (Attica) borders of the species' distribution area. The alignment of both chloroplast
145 genomes had a length of 150,942 nucleotides and showed 99.97% identity. Among the
146 differences was an 84 base pair (bp) deletion in the *trn*S–*trn*G intergenic spacer in the Attic

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 individual that we found also in an individual of *C. sativus*. We used this marker as sequence
148 characterized amplified region (SCAR) to screen for chloroplast differences in *C. sativus* and
149 *C. cartwrightianus* individuals from all populations. We found that all *C. sativus* individuals
150 possess the short allele while in *C. cartwrightianus* it occurs only in Attica and in one out of
151 two populations from the island of Kea (Figure 2). Kea is directly adjacent of Attica (Figure
152 1A) and was connected to the mainland repeatedly when Quaternary sea levels dropped
153 (Lambeck, 1996). This may have allowed gene flow between the crocus stands of these
154 areas. In all other populations of *C. cartwrightianus* only the longer allele was detected
155 (Figure 2).

156 To further compare the chloroplast sequences of saffron and Attic *C. cartwrightianus*, the
157 chloroplast genome of a *C. sativus* individual was assembled. Comparing both sequences
158 we found them to be identical except for an additional adenine occurring at position 1509 of
159 the gene for the beta subunit of the acetyl-CoA carboxylase (*acc*D) in *C. sativus*. This
160 mutation results in a stop codon terminating the *acc*D coding region three amino acids earlier
161 than in the wild type. We re-sequenced the relevant region of the *acc*D gene in 60 individuals
162 of Attic *C. cartwrightianus* but could not detect the mutated variant in these samples. Still, the
163 sole occurrence of the deletion in the *trn*S–*trn*G region in the chloroplast genome in Attic and
164 Kean *C. cartwrightianus* individuals clearly places the maternal parent of the saffron crocus
165 in this area of Greece and, thus, is in accord with the GBS data.

166

167 **Analysis of genome size in *C. cartwrightianus***

168 To be able to infer the mode of origin of the triploid, i.e. if it evolved through a cross between
169 a di- and a tetraploid parent or through the combination of a reduced with an unreduced
170 gamete within diploids, we collected leaves from 100 *C. cartwrightianus* individuals in the
171 Attica area, dried them in the field in silica gel, and analyzed genome sizes for these
172 individuals in the lab by flow cytometry. We were able to obtain results for 91 individuals. We
173 found that all of them have a 2C genome size of 7.06 ±0.09 pg (SD). This value was also
174 observed in individuals with chromosome counts of $2n = 2x = 16$. From this we conclude that
175 *C. cartwrightianus* in Attica is diploid and that tetraploid plants are not frequent. This is in
176 accord with the observation that *C. sativus* is a clonal lineage that originated most probably
177 only once, indicating that it was a rare event. If tetraploid individuals would occur regularly
178 within the diploid *C. cartwrightianus* populations, triploid plants should arise over and over
179 again through crosses between both ploidy levels. Over time such continuous input of
180 triploids into the saffron gene pool should have broadened the genetic diversity occurring in
181 saffron. This is, however, not the case as saffron is genetically rather uniform (Nemati et al.,
182 2014; Busconi et al., 2015).

183

184 **Analysis of nuclear single-copy loci**

185 Taking into account the high genetic diversity in *C. cartwrightianus*, we hypothesized that this
186 could have influenced the outcome of earlier phylogenetic studies where *C. cartwrightianus*
187 and *C. sativus* were included but did not result as sister species. We used DNA sequences
188 of five single-copy genes that were amplified from *C. cartwrightianus*, *C. sativus* and their
189 four closest relatives *C. hadriaticus*, *C. oreocreticus*, *C. pallasii* and *C. thomasii* (Nemati et
190 al., 2018), which all share the same chromosome base number of $x = 8$. Where initial direct
191 sequencing provided evidence for the presence of more than one copy of a gene within an
192 individual, amplicons were cloned and six clones per individual were sequenced.
193 Phylogenetic analyses of the DNA sequences of the five genes (Figure S5) revealed in all

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

194 cases that alleles occurring in different species were not completely sorted according to their
195 species affiliation. This phenomenon, referred to as incomplete lineage sorting (Maddison,
196 1997), is often found among closely related species (Jakob and Blattner, 2006; Brassac and
197 Blattner, 2015). Even the different alleles or homeologs detected within the *C. sativus*
198 individual were rather diverse and could group in different clades in the gene trees. Thus, the
199 high genetic diversity in *C. cartwrightianus* is not restricted to non-coding parts of the
200 genome (Larsen et al., 2015) but concerns also the gene space of this species, which could
201 have rather diverse allelic constitutions.

202

203 **Conclusions**

204 Earlier molecular studies of saffron evolution did not arrive at clear results regarding the
205 parental species of *C. sativus* or the area of origin of saffron. The main reason seems that
206 they did not take into account the high intra-specific genetic diversity present in *C.*
207 *cartwrightianus*. Depending on the individual(s) studied and the marker region used, the
208 resulting phylogenetic trees might reflect nearly arbitrary relationships (Figure S5). In
209 contrast, our GBS data were based on an exhaustive collection of *C. cartwrightianus*
210 populations and clearly place the *C. sativus* individuals as sister of Attic *C. cartwrightianus*.
211 Possible reasons for the sister group position instead of grouping within the Attic population
212 is most probably the triploid and clonal nature of *C. sativus* that, as a group, has therefore a
213 unique character combination that is in this way not present in any individual of *C.*
214 *cartwrightianus*. Still, overall frequencies of GBS alleles (Table 1) and also chloroplast data
215 (Figure 2) support the saffron origin within the Attica/Kea region with closest similarity of the
216 saffron crocus to the *C. cartwrightianus* plants occurring in Attica (Figure 1B, 1C).

217 When collecting leaves of Attic *C. cartwrightianus*, we evaluated the populations for the
218 presence of the important traits typical for the saffron crocus. We recognized the bunchy
219 growing habit, very long stigmas, stigmas of rather dark red color (Figure 1D, Figure S6), and
220 also stigmas with the specific taste and aroma of saffron, particularly in the southern part of
221 this area. However, we did not find plants combining all these traits within single individuals
222 in the same way as saffron. As genetic diversity is high in *C. cartwrightianus* and the species
223 is an obligate outbreeder, it is unlikely to find in today's individuals regularly the exact allele
224 combination characteristic for triploid *C. sativus*, as allele composition is constantly jumbled
225 by genetic recombination. This is also apparent in the diverse karyotpyes of *C.*
226 *cartwrightianus*, where saffron-specific chromosomes can be found, although they are not
227 united within single *C. cartwrightianus* individuals (Schmidt et al., 2019).

228 From Minoan frescos it is clear that more than 3600 years ago humans already used wild
229 saffron in the southern Aegean. The first clear indication for the cultivation of triploid saffron
230 can be found in *Historia Plantarum* (350 BC–287 BC) where Theophrastus described the
231 plant as being propagated by corms (Negbi and Negbi, 2002). Our GBS data point to the
232 small Greek region south of Athens as the place where saffron evolved. We assume that
233 sometime in between 1600 BC and 350 BC a triploid *C. cartwrightianus* cytotype originated
234 in Attica and was selected by humans. They must have realized that they have a highly
235 aromatic and stable type at hands that keeps the valuable properties of saffron through time
236 and (vegetative) generations. A bit surprising is the fact that the main growing regions for
237 saffron are today found clearly outside the distribution area of *C. cartwrightianus*, i.e. in the
238 western Mediterranean (Spain, Morocco) and western Asia (Iran to northern India). While *C.*
239 *cartwrightianus* is restricted to the Mediterranean vegetation zone, saffron cultivation
240 happens mostly in much drier regions and at higher elevation. Thus, we assume differences

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

241 in climate requirements between both species. This could indicate an ecologic niche shift due
242 to polyploidization. We cannot yet determine how much triploidy influences the development
243 of the typical traits of saffron or if the right allele combination in a diploid might provide similar
244 characteristics. Still, the clarification of the mode of evolution of *C. sativus* now opens up a
245 route for overcoming the low genetic diversity present in the saffron crocus, as it will foster
246 new saffron genotypes to be created from different *C. cartwrightianus* individuals.
247
248

249 ## Materials and methods

250 ### Taxon sampling

251 We included 197 individuals in the genotyping-by-sequencing (GBS) study, 22 belonging to
252 *C. sativus*, 167 to *C. cartwrightianus* and eight to *C. oreocreticus* (Table S1). The sampling of
253 *C. cartwrightianus* covered its entire distribution range (Figure 1A). *Crocus oreocreticus*, the
254 closest relative of *C. cartwrightianus* and *C. sativus* (Nemati et al., 2018), is endemic to
255 Crete, where it was collected in two different populations with four individuals each. In
256 addition we used single individuals of the other *Crocus* series *Crocus* species sharing the
257 chromosome number of $2n = 2x = 16$ with *C. cartwrightianus*, which were often named as
258 parents of triploid *C. sativus* ($2n = 3x = 24$). These are *C. hadriaticus*, *C. pallasii* and *C.*
259 *thomasii*, which were included in screens of allelic diversity at nuclear single-copy genes.
260 Voucher information for the analyzed taxa is provided in Table S1.
261

262 ### DNA extraction and PCR reactions

263 Extraction of genomic DNA was carried out using DNeasy Plant DNA Extraction Kit (Qiagen)
264 from about 10 mg of silica-dried leaf material according to the protocol of the manufacturer.
265 DNA concentration and quality were afterwards checked on 0.8% agarose gels.

266 To obtain nuclear single-copy marker regions with high variability in *Crocus*, we used contigs
267 derived from the assembly of low-coverage next-generation sequencing (Illumina HiSeq
268 platform; see below) of *Crocus cartwrightianus*. Potential nuclear single-copy genes and their
269 intron-exon borders were identified using the PLAZA v2.5 and v3 platform (Van Bel et al.,
270 2012; Vandepoele, 2017).

271 We selected five nuclear single-copy genes (Table S2), which all were heterozygous in *C.*
272 *sativus*, and PCR amplified them in five *Crocus* species closely related to saffron. PCR was
273 performed with 1 U Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in the
274 supplied Phusion GC Buffer, 200 µM of each dNTP, 0.5 µM of each primer, and about 20 ng
275 of total DNA in 50 µl reaction volume in a GeneAmp PCR System 9700 (Perkin-Elmer).
276 Amplification was performed with 3 min initial denaturation at 95 °C and 35 cycles of 30 s at
277 95 °C, 25–60 sec at the marker-specific annealing temperature (table S2) and 30 s at 70 °C,
278 followed by a final extension for 8 min at 70 °C. PCR products were purified on a 1% gel and
279 extracted using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's protocol,
280 and eluted in 30 µl water. Both strands of the PCR products were initially directly sequenced
281 with Applied Biosystems BigDye Terminator technology on an ABI 3730xl automatic DNA
282 sequencer using the primers from PCR amplifications. When direct sequencing revealed
283 polymorphic sequence positions or length differences, PCR products were cloned into the
284 pGEM-T Easy vector (Promega) and six clones per individual were sequenced with Templi-
285 Phi DNA Sequencing Template Ampflication Kit (Amersham Biosciences).
286

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

287 **Library preparation and next-generation sequencing**

288 To obtain genome-wide single-nucleotide polymorphisms (SNPs), genotyping-by-sequencing
289 (GBS) analyses (Elshire et al., 2011) were conducted for 197 individuals. For the library
290 preparation 200 ng of genomic DNA were used and cut with the two restrictions enzymes
291 *Pst*I-HF (NEB) and *Msp*I (NEB). Library preparation, individual barcoding, and single-end
292 sequencing on the Illumina HiSeq 2000 followed Wendler et al. (2014).

293 For WGS sequencing of *C. cartwrightianus* from Crete and Attica 1-2 µg DNA were used.
294 Library preparation was carried out as described by Meyer and Kircher (2010) for *C.*
295 *cartwrightianus* from Crete. Library preparation for *C. cartwrightianus* from Attica was done
296 according to the Illumina TruSeq DNA library preparation protocol following the
297 manufacturer's recommendations. DNA was covarized to generate fragments of on average
298 300-400 bp length for the Crete and 400-600 bp for Attic *C. cartwrightianus*, followed by
299 adaptor and barcode ligation. For *C. cartwrightianus* from Crete an additional 8-kb mate-pair
300 library was generated. The libraries were size-selected with a SYBR Gold stained
301 electrophoresis gel. Fragment size distribution and DNA concentration were evaluated on an
302 Agilent BioAnalyzer High Sensitivity DNA Chip and using the Qubit DNA Assay Kit in a Qubit
303 2.0 Flurometer (Life Technologies). Finally the DNA concentration of the libraries was
304 checked by a quantitative PCR run. Cluster generation on Illumina cBot and paired-end
305 sequencing (*C. cartwrightianus* from Crete: 2x100 bp; *C. cartwrightianus* from Attica: 2x250
306 bp) on the Illumina HiSeq 2000/2500 platform followed Illumina's recommendation and
307 included 1% Illumina PhiX library as internal control. The 8-kb mate-pair library of *C.*
308 *cartwrightianus* from Crete was sequenced using 20% of a lane on the Illumina MiSeq
309 platform generating 2x250 bp paired-end reads.

310 Barcoded reads were de-multiplexed using the CASAVA pipeline 1.8 (Illumina). The obtained
311 raw sequencing reads were quality checked and over-represented, i.e. clonal reads were
312 detected with FASTQC (Andrew, 2010). Adapter trimming of sequence reads was performed
313 with CUTADAPT (Martin, 2011) and reads shorter than 60 bp after adapter removal were
314 discarded.

315

316 **Genome assembly of whole-genome shotgun data and chloroplast genomes**

317 *De novo* assembly of WGS sequences of *C. cartwrightianus* from Crete with 210 million
318 quality-filtered read pairs was performed in CLC v4.3.0 (CLC bio) with a minimum length for
319 assembled contigs of 500 bp. NCBI BLAST v2.2.28+ searches were used to check for
320 bacterial contaminations in the sequence reads and to identify plastid-derived contigs.
321 Scaffolding with SSPACE v3.0 (Boetzer et al., 2011) was performed with a minimum number
322 of 100 linked reads to compute a scaffold.

323 Plastid scaffolds were identified by a BLAST search. The scaffolds were then mapped to the
324 chloroplast genomes of *Beta vulgaris* (GenBank accession number KR230391), *Haloxylon*
325 *persicum* (KF534479) and *Iris gatesii* (KM014691). GAPFILLER v1.10 (Boetzer and Pirovano,
326 2012) was used to fill gaps. Proper pairing of reads was checked by mapping the original
327 reads against the obtained *C. cartwrightianus* plastid genome using GENEIOUS R10.2.3
328 (Biomatters Ltd.) and checked manually.

329 Illumina sequencing with lower coverage produced 123.6 million (61.8 million paired end)
330 quality-filtered reads for *C. cartwrightianus* from Attica. For *C. sativus* 74.4 million (37.2
331 million paired end) quality-filtered reads from a WGS experiment were provided by Thomas
332 Schmidt and Tony Heitkam (Molecular and Cell Biology, University of Dresden, Germany)
333 and assembled as before.

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

334 The chloroplast genome of *C. cartwrightianus* from Crete obtained by *de novo* assembly was
335 used as a reference sequence for chloroplast genomes of *C. cartwrightianus* from Attica and
336 *C. sativus*. Annotation of chloroplast genomes was performed in GENEIOUS and edited
337 manually.

338

339 **Chloroplast polymorphisms screening**

340 A sequence part of the chloroplast *trn*S(GCU)–*trn*G(UCC) intergenic spacer, where an 84-bp
341 deletion was observed in *C. sativus* and *C. cartwrightianus* from Attica, as well as a 1-bp
342 insertion in the *acc*D gene in *C. sativus* at position 1509 were confirmed by Sanger re-
343 sequencing (Table S2). Screening of the distribution of the 84-bp deletion in the *trn*S-*trn*G
344 intergenic spacer was conducted by PCR amplification of the respective chloroplast region in
345 the set of 96 *C. cartwrightianus* individuals covering the populations from the entire
346 distribution area of the species (table 1) plus the *C. sativus* individuals. Size determination of
347 the PCR products was done on 1.4% agarose gels against a 100-bp DNA ladder and always
348 including one saffron amplicon as reference for the short allele.

349

350 **Processing of genotyping-by-sequencing data**

351 To have a particular focus on loci present in *C. sativus* we first generated a loci reference file
352 using IPYRAD v0.7.19 (Eaton, 2014). An assembly of the GBS data of 22 saffron individuals
353 was done *de novo*. The minimal number of samples per locus was set to 18, the clustering
354 threshold of reads within and between individuals was set to 0.9. The maximum ploidy level
355 was appointed as triploid. For the other parameters the default settings of parameter files
356 generated by IPYRAD were used. The locus file generated by IPYRAD was converted into a
357 FASTA file with 6512 sequences corresponding to the number of obtained loci.

358 Additional IPYRAD analyses were run, using either 126 or 197 individuals and the reference-
359 based assembly of IPYRAD. The clustering threshold of reads within and between individuals
360 was set to 0.9. The maximum ploidy level was appointed as triploid. For the other parameters
361 the default settings of parameter files generated by IPYRAD were used except for the maximal
362 number of indels, which was increased to 15. The minimal number of samples per locus was
363 set to 18 to generate the output file used for the determination of the proportion of SNPs *C.*
364 *sativus* shares with *C. cartwrightianus* and *C. oreocreticus*. For Bayesian assignment as well
365 as the phylogenetic analyses, the minimal number of samples per locus was set to 55.
366 VCFTOOLS 0.1.14 (Danacek et al., 2011) were used to filter out the genotypes with a depth
367 below six.

368

369 **Determining the proportion of shared GBS alleles**

370 The number of shared GBS alleles between the species and populations was inferred using
371 PopGenReport (Adamack and Gruber, 2014) in R 3.5.0. In principle, the filtered vcf file was
372 first converted into a genind object, then concatenated with the information about the sample
373 origin and converted into a genpop object. The genpop file directly includes the allelic counts
374 per variable position, which we here simply refer to as SNPs. SNPs not present and SNPs
375 potentially originating by autosomal mutation in saffron (present in less than four counts in
376 saffron) were excluded. Counting and calculating percentages were conducted in Microsoft
377 EXCEL v14 after transposing the genpop matrix in R. In addition to the dataset where we
378 used all individuals per population we also calculated the proportion of shared GBS alleles
379 for seven (lowest number of individuals among our population samples) randomly chosen

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

380 individuals for the populations where higher individual numbers were available. For this
381 normalized dataset the included individuals were permuted ten times and the percentage of
382 shared GBS alleles was averaged over the ten runs.
383
384 **Phylogenetic and population genetic analyses**
385 Forward and reverse sequences for the five nuclear and two chloroplast marker regions were
386 manually checked, edited where necessary, and assembled in one sequence for each locus
387 and individual. In cases where cloning revealed the presence of different alleles, cloned
388 sequences were assigned to different haplotypes or consensus sequences were generated
389 for cloned sequences differing by only one substitution.
390 Phylogenetic analysis using maximum parsimony (MP) was performed in PAUP* 4a161
391 (Swofford, 2002) using either the branch-and-bound algorithm (in cases when fewer than 15
392 sequences of the single-copy loci occurred in a dataset) or the heuristic search algorithm
393 with TBR branch swapping for the concatenated GBS-derived sequences and single-copy
394 loci with 15 and more sequences. Gaps were treated as missing data and we used 1000
395 random-addition-sequences (RAS) to construct the starting trees in the heuristic search for
396 the GBS data to avoid suboptimal tree-islands. Bootstrap support values were obtained by
397 500 bootstrap re-samples using the same settings as before except that we excluded RAS.
398 Maximum-likelihood analyses were conducted in RAxML v. 8.0.0 (Stamatakis, 2014) using
399 GTRGAMMA and 100 parsimony starting trees. The number of bootstrap re-samples was set
400 to 500.
401 Pairwise genetic distances within the GBS dataset were calculated in PAUP* as uncorrected
402 ("p") distances for the *C. sativus* individuals and the *C. cartwrightianus* individuals within and
403 among populations. In both species genetic diversity was rather uniformly distributed among
404 the individuals.
405 Bayesian assignment analysis for the GBS data was carried out using the LEA package
406 (François, 2016) in R 3.5.0 in an initial run for K = 1 to K = 20. The lowest entropy here was
407 observed for K = 9. Therefore, for K = 2 to K = 14 Bayesian assignment analysis was run
408 with 20 repetitions. The run and repetition with lowest cross entropy was identified and
409 selected. Additionally, the populations of the sampling data were read in using data.table
410 (Dowle and Srinivasan, 2018). The *Q*-matrix obtained by LEA, including the ancestral
411 assignment frequencies, was sorted in R with Tidyverse (Wickham, 2017) according to the
412 population the samples belong to. Plots were arranged with the gridExtra R package
413 (Auguie, 2017).
414
415 **Genome size determinations in Attic *C. cartwrightianus***
416 To infer genome sizes and ploidy level of Attic *C. cartwrightianus* individuals we collected a
417 leaf from each of 100 individuals in the field and dried them in silica gel. Afterwards the
418 leaves were transported to the lab and genome sizes were measured with a CyFlow Space
419 (Sysmex Partec) flow cytometer against a *Vicia faba* size standard (26.5 pg 2C) using DAPI
420 as staining reagent and essentially following the procedure described by Jakob et al. (2004).
421
422
423

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

424 ## Acknowledgements

425 We like to thank the Greek authorities for providing permits for plant collections, I.
426 Faustmann, C. Koch, B. Kraenzlin and P. Oswald for help with plant cultivation and lab work,
427 T. Schmidt and T. Heitkam for access to saffron whole-genome shotgun data, A.
428 Himmelbach and S. König for performing Illumina sequencing, and H. Poskar and J. Brassac
429 for critical comments on the manuscript. We acknowledge funding by the Deutsche
430 Forschungsgemeinschaft (grants BL462/15 to F.R.B. and HA7550/2 to D.H.).

431

432

433 ## Additional information

434 ### Competing interests

435 We declare that no competing interests exist.

436

437 ### Funding

438	Deutsche Forschungsgemeinschaft	BL462/15	Frank Blattner
439	Deutsche Forschungsgemeinschaft	HA7550/2	Dörte Harpke

440

441 ### Author contributions

442 Designed study: F.R.B., D.H. Coordinated study: Z.N., D.H., F.R.B. Provided data or
443 materials: H.K. Performed experiments: Z.N. Analyzed data: Z.N., D.H., F.R.B., A.G. The
444 initial manuscript was written by F.R.B. All authors contributed to and approved the final
445 version of the manuscript.

446

447 ### Data Accessibility

448 All sequence data are available through DDBJ/ENA/GenBank for the nuclear single-copy
449 genes and the chloroplast locus (LS975036–LS975118), the annotated chloroplast genomes
450 (MH542231–MH542233), and the GBS data (ERR2740826–ERR2740842, ERR2740845–
451 ERR2741003).

452

453

454 ## Additional files

455 ### Supplementary files

456 **Figure supplement 1.** Parsimony consensus tree based on GBS sequence data.

457 **Figure supplement 2.** Maximum-likelihood phylogenetic tree based on GBS sequence
458 data.

459 **Figure supplement 3.** Panel of Bayesian assignment analyses from K = 2 to K = 9.

460 **Figure supplement 4.** Chloroplast genome map for *C. cartwrightianus*.

461 **Figure supplement 5.** Panel of five single-copy gene trees calculated with the maximum-
462 likelihood algorithm to infer the influence of allele diversity in *C. cartwrightianus* on the
463 phylogenetic position of *C. sativus* and other included *Crocus* species.

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

464 **Figure supplement 6.** Photographs of *C. sativus* and three individuals of Attic *C.*
465 *cartwrightianus*.

466 **Table supplement 1.** Studied plant materials.

467 **Table supplement 2.** Analyzed nuclear and chloroplast loci.

468

469

470 ## References

471 Adamack AT, Gruber B. 2014. PopGenReport: simplifying basic population genetic analyses in R.
472 *Methods Ecol Evol* 5:384-387.

473 Abdullaev FI, Frenkel GD. 1999. Saffron in biological and medical research. In: Negbi M, editor.
474 Saffron: *Crocus sativus* L., Harwood Acad Publ. p. 103-113.

475 Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data.
476 http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

477 Auguie B. 2017. gridExtra: miscellaneous functions for "Grid" graphics. R package version 2.3.
478 https://CRAN.R-project.org/package=gridExtra.

479 Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs
480 using SSPACE. *Bioinformatics* 27:578-579.

481 Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56.

482 Brassac J, Blattner FR. 2015. Species-level phylogeny and polyploid relationships in *Hordeum*
483 (Poaceae) inferred by next-generation sequencing and *in silico* cloning of multiple nuclear loci.
484 *Syst Biol* 64:792-808.

485 Brighton CA. 1977. Cytology of *Crocus sativus* and its allies (Iridaceae). *Plant Syst Evol* 128:137-157.

486 Busconi M, Colli L, Sánchez RA, Santaella M, De-Los-Mozos Pascual M, Santana O, Roldán M,
487 Fernández JA. 2015. AFLP and MS-AFLP analysis of the variation within saffron crocus (*Crocus*
488 *sativus* L.) germplasm. *PLoS One* 10:e0123434.

489 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth
490 GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156-2158.

491 Dowle M, Srinivasan A. 2018. Data.table: extension of 'data.frame'. R package version 1.11.2.
492 https://CRAN.R-project.org/package=data.table.

493 Eaton DAE. 2014. Pyrad: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*
494 30:1844-1849.

495 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust,
496 simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*
497 6:e19379.

498 François O. 2016. Running Structure-like population genetic analysis with R. PDF R tutorials in
499 population genetics. Grenoble: University Grenoble-Alpes. p. 1-8.

500 Ghaffari SM. 1986. Cytogenetic studies of cultivated *Crocus sativus* (Iridaceae). *Plant Syst Evol*
501 153:199-204.

502 Harpke D, Meng S, Kerndorff H, Rutten T, Blattner FR. 2013. Phylogeny of *Crocus* (Iridaceae) based
503 on one chloroplast and two nuclear loci: ancient hybridization and chromosome number
504 evolution. *Mol Phylogenet Evol* 66:617-627.

505 Jakob SS, Blattner FR. 2006. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting
506 haplotypes, incomplete lineage sorting, regional extinction, and the consequences for
507 phylogenetic inference. *Mol Biol Evol* 23:1602-1612.

508 Jakob SS, Meister A, Blattner FR. 2004. The considerable genome size variation of *Hordeum* species
509 (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Mol Biol Evol* 21:860-
510 869.

511 Lambeck K. 1996. Sea-level change and shore-line evolution in Aegean Greece since Upper
512 Palaeolithic time. *Antiquity* 70:588-611.

513 Larsen B, Orabi J, Pedersen C, Ørgaard M. 2015. Large intraspecific genetic variation within the
514 Saffron–Crocus group (*Crocus* L., Series *Crocus*; Iridaceae). *Plant Syst Evol* 301:425-437.

515 Maddison WP. 1997. Gene trees in species trees. *Syst Biol* 46:523-536.

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

516 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
517 *EMBnet.journal* 17:10-12.
518 Mathew B. 1999. Botany, taxonomy and cytology of *C. sativus* L. and its allies. In: Negbi M, editor.
519 Saffron: *Crocus sativus* L. Amsterdam: Harwood Acad Publ. p. 19-30.
520 Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target
521 capture and sequencing. *Cold Spring Harbor Protocols*: pdb.prot5448.
522 Negbi M. 1999. Saffron cultivation: past, present and future prospects. In: Negbi M, editor. Saffron:
523 *Crocus sativus* L., Harwood Acad Publ. p. 1-17.
524 Negbi M, Negbi O. 2002. The painted plaster floor of the Tell Kabri palace: reflections on saffron
525 domestication in the Aegean Bronze Age. In: Oren ED, Ahituv S, editors. Aharon Kempinski
526 Memorial Volume. Beer-Sheva: Ben-Gurion Univ Press. p. 325-340.
527 Nemati Z, Blattner FR, Kerndorff H, Erol O, Harpke D. 2018. Phylogeny of the saffron-crocus species
528 group, *Crocus* series *Crocus* (Iridaceae). *Mol Phylogenet Evol* 127:891-897.
529 Nemati Z, Mardi M, Majidian P, Zeinalabedini M, Pirseyedi SM, Bahadori M. 2014. Saffron (*Crocus*
530 *sativus* L.), a monomorphic or polymorphic species? *Spanish J Agric Res* 12:753-762.
531 Pearse D, Crandall KA. 2004. Beyond F_{ST}: analysis of population genetic data for conservation.
532 *Conservation Genet* 5:585-602.
533 Schmidt T, Heitkam T, Lietke S, Schubert V, Menzel G. 2019. Adding color to a century-old enigma:
534 multi-color chromosome identification unravels the autotriploid nature of saffron (*Crocus sativus*)
535 as a hybrid of wild *Crocus cartwrightianus* cytotypes. *New Phytol*: In press.
536 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
537 phylogenies. *Bioinformatics* 30:1312-1313.
538 Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the
539 genomic iceberg: next-generation sequencing for plant systematics. *Am J Bot* 99:349-364.
540 Swofford DL. 2002. *PAUP*.* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.
541 Sunderland: Sinauer Associates.
542 Tsaftaris A, Pasentsis K, Makris A, Darzentas N, Polidoros A, Kalivas A, Argiriou A. 2011. The study
543 of the E-class SEPALLATA3-like MADS-box genes in wild-type and mutant flowers of cultivated
544 saffron crocus (*Crocus sativus* L.) and its putative progenitors. *Plant Phys* 168:1675-1684.
545 Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K. 2012.
546 Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Phys* 158:590-
547 600.
548 Vandepoele K. 2017. A guide to the PLAZA 3.0 plant comparative genomic database. *Methods Mol*
549 *Biol* 1533:183-200.
550 Wendler N, Mascher M, Nöh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N. 2014. Unlocking
551 the secondary gene-pool of barley with next-generation sequencing. *Plant Biotech J* 12:1122-
552 1131.
553 Wickham H. 2017. Tidyverse: easily install and load the 'Tidyverse'. R package version 1.2.1.
554 https://CRAN.R-project.org/package=tidyverse.
555

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

556

557 **Table**

558

559 **Table 1.** Proportion of *C. sativus* GBS alleles occurring within *C. cartwrightianus* and *C.*
560 *oreocreticus*.

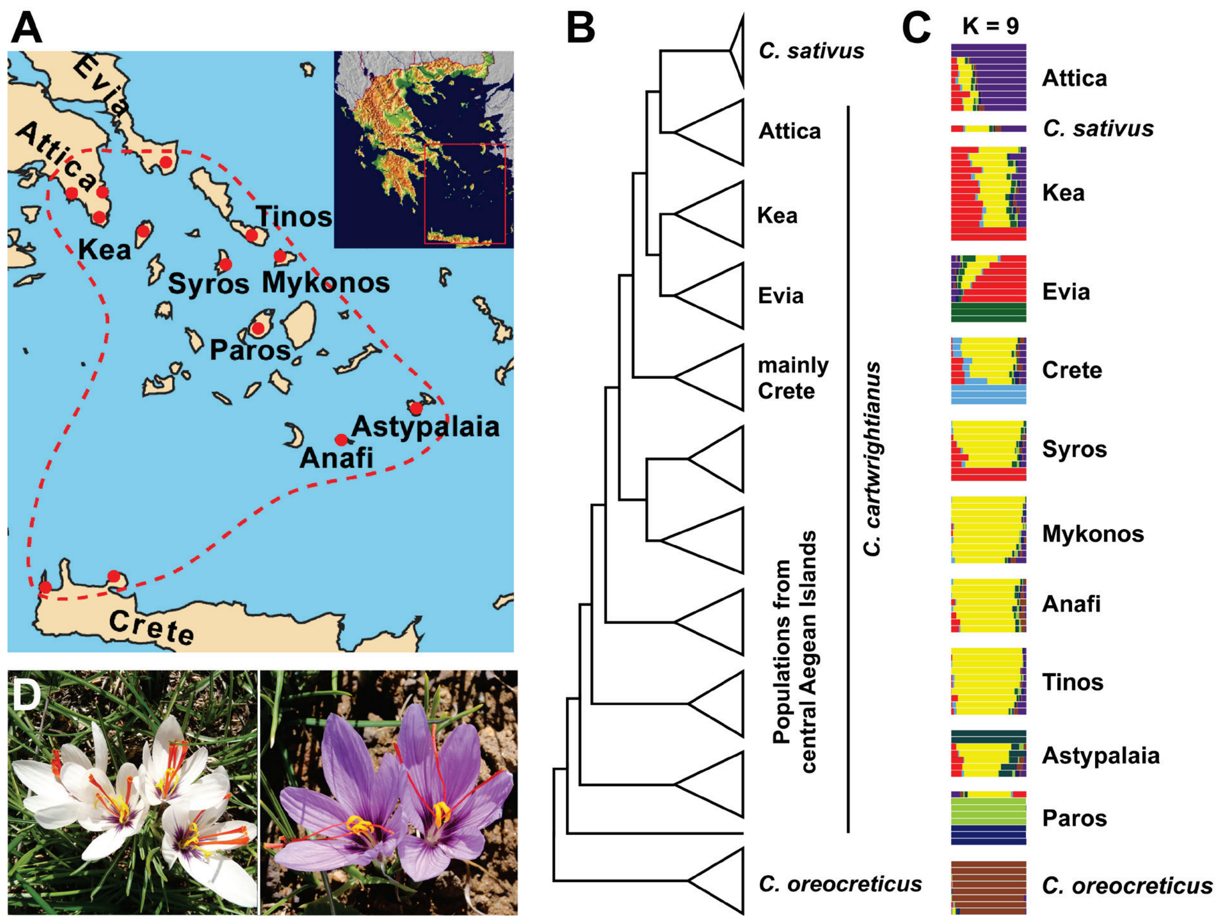
Population/Species	N	Proportion[a]	Normalized proportion[b]
Attica	10	90.47%	87.99% (±0.31)
Kea	14[c]	90.46%	85.44% (±0.98)
Tinos	10	88.34%	85.23% (±0.35)
Mykonos	10	87.96%	84.88% (±0.33)
Evia	10	86.06%	83.33% (±1.10)
Anafi	8	85.96%	84.70% (±0.18)
Crete	10	85.67%	82.14% (±1.65)
Syros	9	84.25%	82.29% (±0.99)
Astypalaia	7	82.13%	82.13%
Paros	8	75.46%	74.77% (±1.39)
Individuals above		98.54%	98.14% (±0.07)
Attica extended	71	97.90%	
All *C. cartwrightianus*	167	99.31%	
C. oreocreticus	8	76.83%	75.81% (±0.09)

561 [a]Including all individuals per location. [b]To correct for different N, proportions were calculated based on
562 seven individuals per location. In brackets standard deviations are provided for 10 runs of randomly
563 sampled individuals. [c]In Kea two populations occur possessing different chloroplast types, which were
564 each included with seven individuals.
565

566

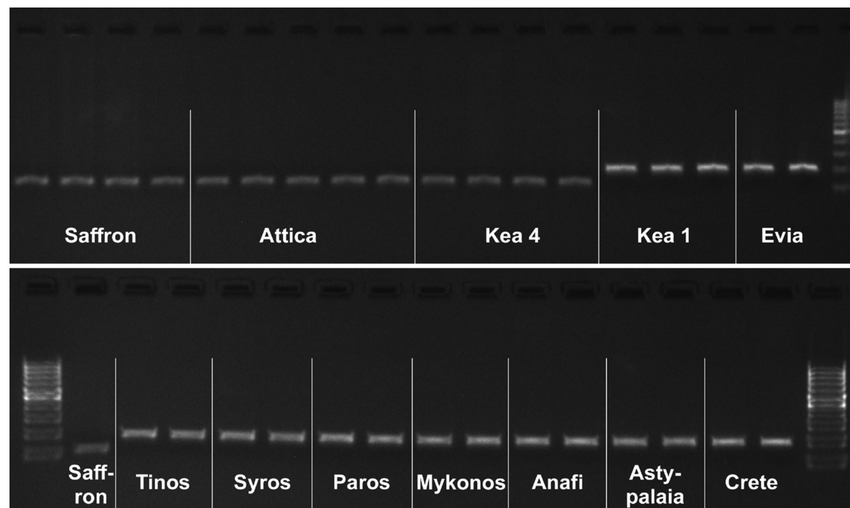
bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

567
568 **Figures**
569



570
571
572 **Figure 1.** Distribution and phylogenetic relationships of *C. cartwrightianus* with regard to *C.*
573 *sativus.* (A) Map providing the distribution area of *C. cartwrightianus* (dashed line). Red dots
574 indicate the collection sites of populations included in the study. (B) Scheme summarizing
575 the topology of two most parsimonious trees based on genome-wide DNA data of a
576 genotyping-by-sequencing analysis of 114 individuals. *Crocus oreocreticus* was defined as
577 outgroup. For clades within *C. cartwrightianus* the geographical origins of the samples are
578 given. (C) Result of a Bayesian population assignment analysis for K = 9. (D) Individuals of
579 *C. cartwrightianus* (left) and *C. sativus* (right).
580 The following figure supplements are available for figure 1:
581 **Figure supplement 1 – 3**
582 **Figure supplement 6**
583

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



584
585
586 **Figure 2.** Gel photos of exemplary samples of PCR amplicons for the distribution of an 84-bp
587 deletion in the chloroplast *trn*S-*trn*G intergenic spacer of *C. cartwrightianus* and *C. sativus*.
588 The shorter allele (119 bp) occurs only in Attica, one population from Kea and in *C. sativus*,
589 while all other populations of *C. cartwrightianus* possess the longer allele (203 bp).

590 The following figure supplement is available for figure 2:

591 **Figure supplement 4**

592

593

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

594 **Supplementary files**

595

596

597 # Saffron (*Crocus sativus*) is an autotriploid that evolved in Attica
598 (Greece) from wild *Crocus cartwrightianus*

599

600 **Zahra Nemati,[1,a] Dörte Harpke,[1,a] Almila Gemicioglu,[1,2] Helmut Kerndorff,[1] and Frank R.
601 Blattner[1,*]**

602

603

604 [1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben,
605 Germany.

606 [2]Dept. of Biology, University of Istanbul, Istanbul, Turkey.

607

608 [a]These authors contributed equally to the study

609 [*]**Corresponding author:** E-mail: blattner@ipk-gatersleben.de

610

611

612

613

614 **Content**

615

616 **Figure supplement 1.** Parsimony consensus tree based on GBS sequence data.

617 **Figure supplement 2.** Maximum-likelihood phylogenetic tree based on GBS sequence
618 data.

619 **Figure supplement 3.** Panel of Bayesian assignment analyses from K = 2 to K = 9.

620 **Figure supplement 4.** Chloroplast genome map for *C. cartwrightianus*.

621 **Figure supplement 5.** Panel of five single-copy gene trees calculated with the maximum-
622 likelihood algorithm to infer the influence of allele diversity in *C. cartwrightianus* on the
623 phylogenetic position of *C. sativus* and other included *Crocus* species.

624 **Figure supplement 6.** Photographs of *C. sativus* and three individuals of Attic *C.*
625 *cartwrightianus*.

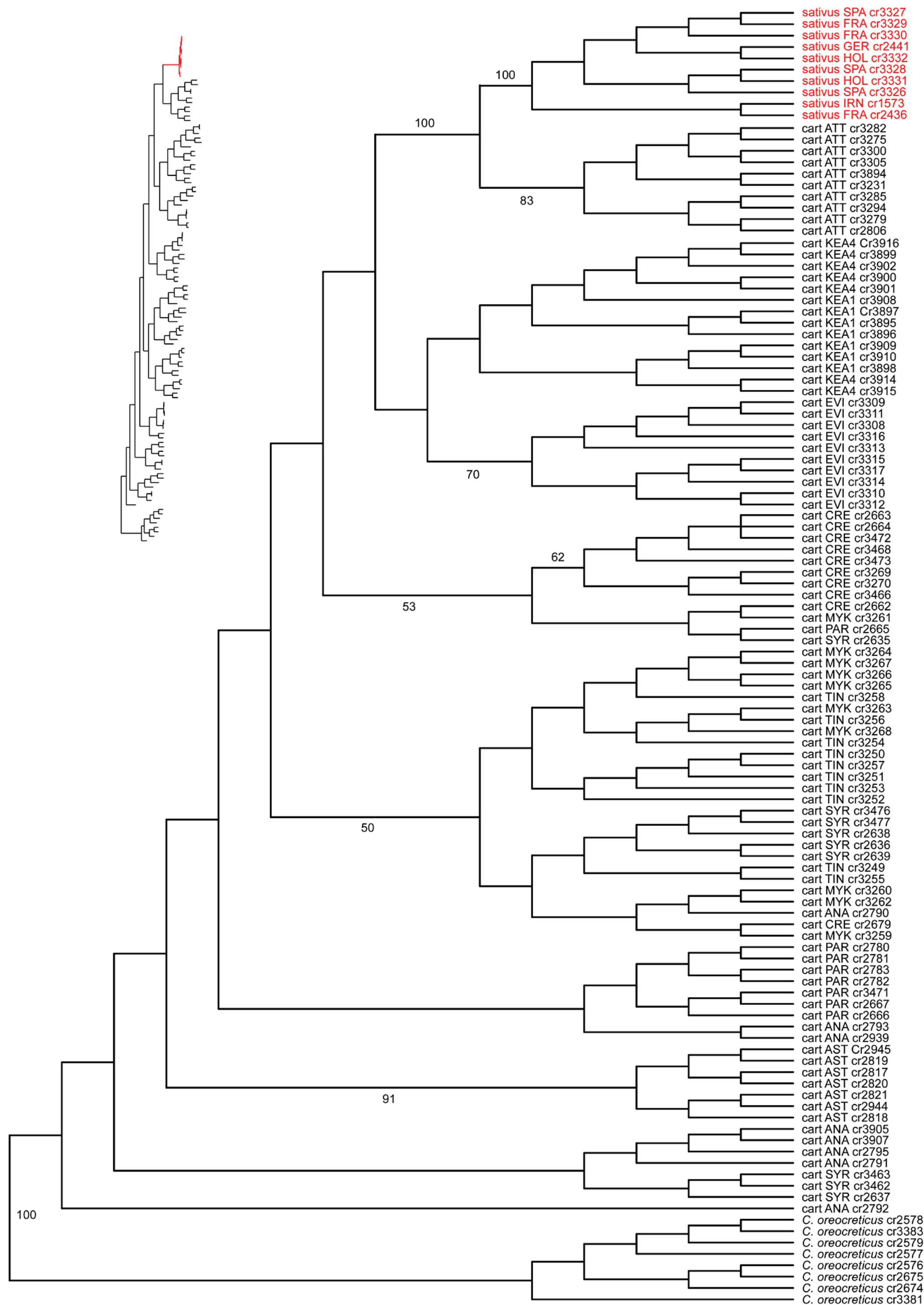
626 **Table supplement 1.** Studied plant materials.

627 **Table supplement 2.** Analyzed nuclear and chloroplast loci.

628

629
630

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



631

632 **Figure S1.** Maximum parsimony consensus tree of two most parsimonious trees obtained
633 through the analysis of the GBS dataset. The inset on the left side provides branch lengths
634 for one of the two trees. Numbers along branches depict bootstrap values ≥50% for major
635 clades.

636

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

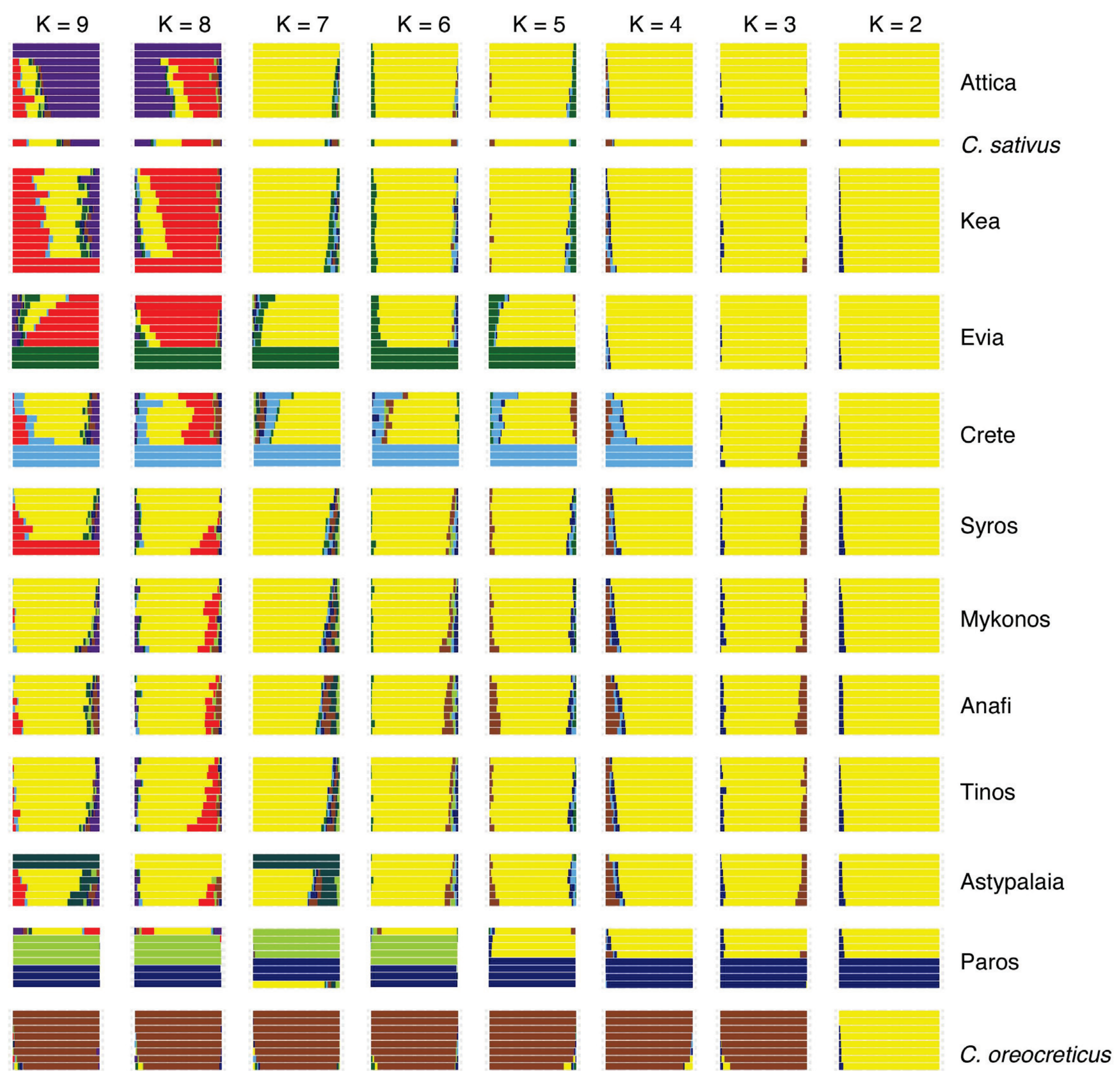
sativus SPA cr3327
sativus FRA cr3329
sativus HOL cr3332
sativus FRA cr3330
sativus GER cr2441
sativus SPA cr3328
sativus HOL cr3331
sativus SPA cr3326
sativus IRN cr1573
sativus FRA cr2436
100
100
cart ATT cr3294
cart ATT cr3231
cart ATT cr3894
cart ATT cr3279
cart ATT cr3282
cart ATT cr2806
83
cart ATT cr3285
cart ATT cr3300
cart ATT cr3305
cart ATT cr3275
cart KEA4 cr3899
cart KEA4 Cr3916
50
cart KEA4 cr3902
cart KEA4 cr3900
cart KEA4 cr3901
cart KEA1 cr3910
cart KEA4 cr3914
cart KEA4 cr3915
98
cart KEA1 cr3898
cart KEA1 cr3908
cart KEA1 cr3896
cart KEA1 cr3909
cart KEA1 cr3895
cart KEA1 Cr3897
cart EVI cr3311
cart EVI cr3309
cart EVI cr3308
cart EVI cr3316
cart EVI cr3310
cart EVI cr3312
53
cart EVI cr3314
cart EVI cr3315
cart EVI cr3317
cart EVI cr3313
cart CRE cr3269
cart CRE cr3270
cart CRE cr3466
cart CRE cr2662
cart CRE cr3473
59
cart CRE cr3472
cart CRE cr2664
cart CRE cr2663
cart CRE cr3468
94
cart SYR cr3463
cart SYR cr3462
cart SYR cr2637
cart ANA cr3907
cart ANA cr2791
65
cart PAR cr2665
cart SYR cr2635
cart TIN cr3251
cart TIN cr3250
63
cart TIN cr3257
cart TIN cr3253
cart TIN cr3252
cart MYK cr3259
cart MYK cr3264
cart MYK cr3267
52
cart MYK cr3266
cart MYK cr3265
cart MYK cr3268
cart TIN cr3254
cart ANA cr2790
cart ANA cr2939
cart MYK cr3260
cart MYK cr3262
83
cart TIN cr3249
cart TIN cr3255
cart SYR cr3476
cart SYR cr3477
cart SYR cr2638
cart SYR cr2639
cart CRE cr2679
cart MYK cr2667
cart PAR cr3471
cart PAR cr2666
cart MYK cr3261
cart MYK cr3263
cart PAR cr2782
cart PAR cr2783
100
cart PAR cr2781
cart PAR cr2780
cart TIN cr3256
cart TIN cr3258
cart SYR cr2636
cart AST cr2821
cart AST cr2944
cart AST cr2818
100
cart AST cr2817
cart AST cr2820
cart AST cr2819
cart AST Cr2945
cart ANA cr2793
100
cart ANA cr2792
cart ANA cr3905
cart ANA cr2795
C. oreocreticus cr2576
C. oreocreticus cr2579
C. oreocreticus cr2675
C. oreocreticus cr3381
C. oreocreticus cr2577
C. oreocreticus cr2674
C. oreocreticus cr2578
C. oreocreticus cr3383

637

638 **Figure S2.** Maximum-likelihood tree obtained through the analysis of the GBS dataset with
639 RAxML. Bootstrap values ≥50% are given for geographically defined clades.

640

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



641

642 **Figure S3.** Bayesian assignment analysis plot for the GBS data with K = 2 to K = 9. *Crocus*
643 *sativus* is assigned to the population from Attica although in some runs it shows also allele
644 patterns similar to individuals from the island of Kea, however, with different proportions.
645

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

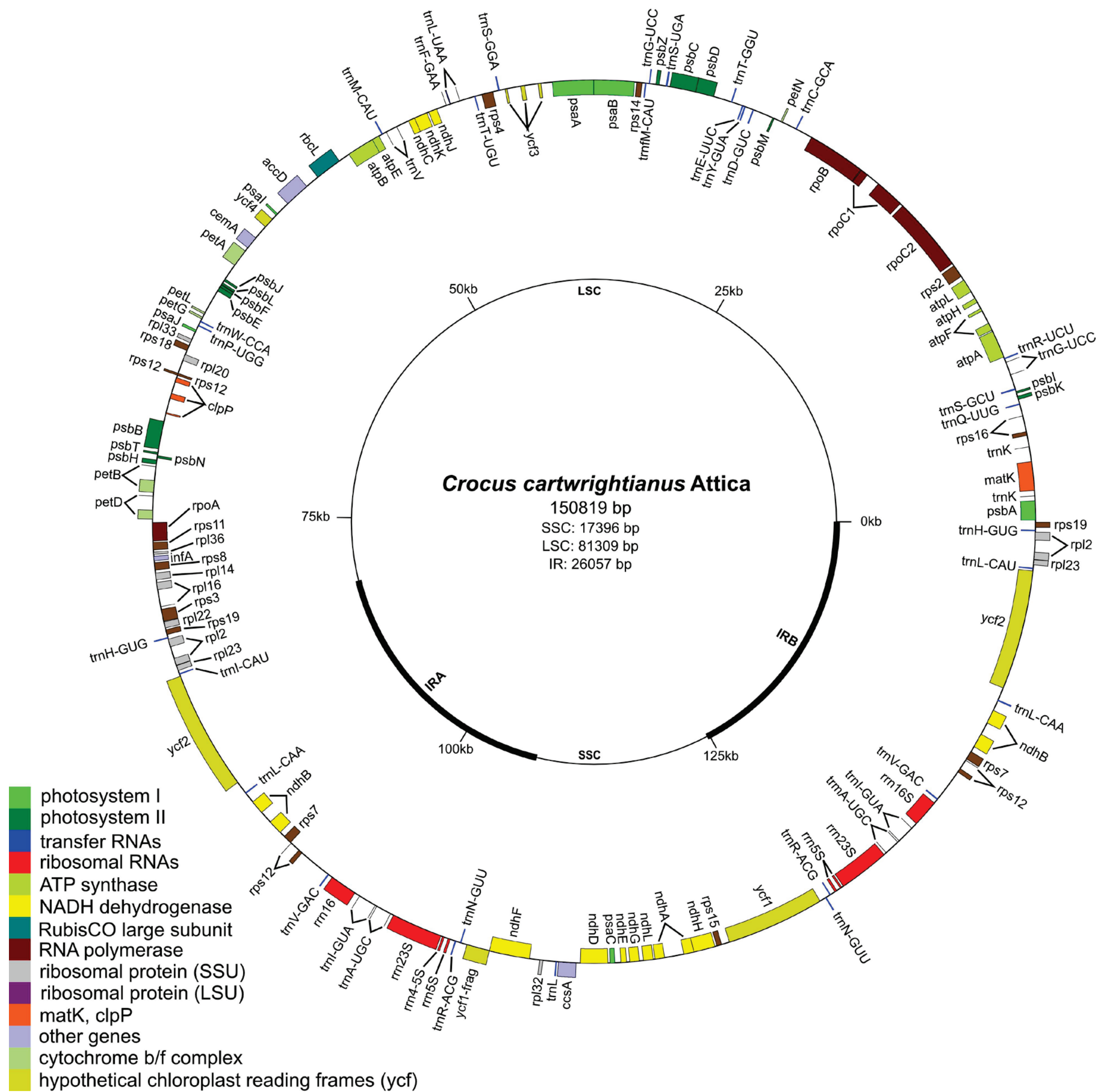
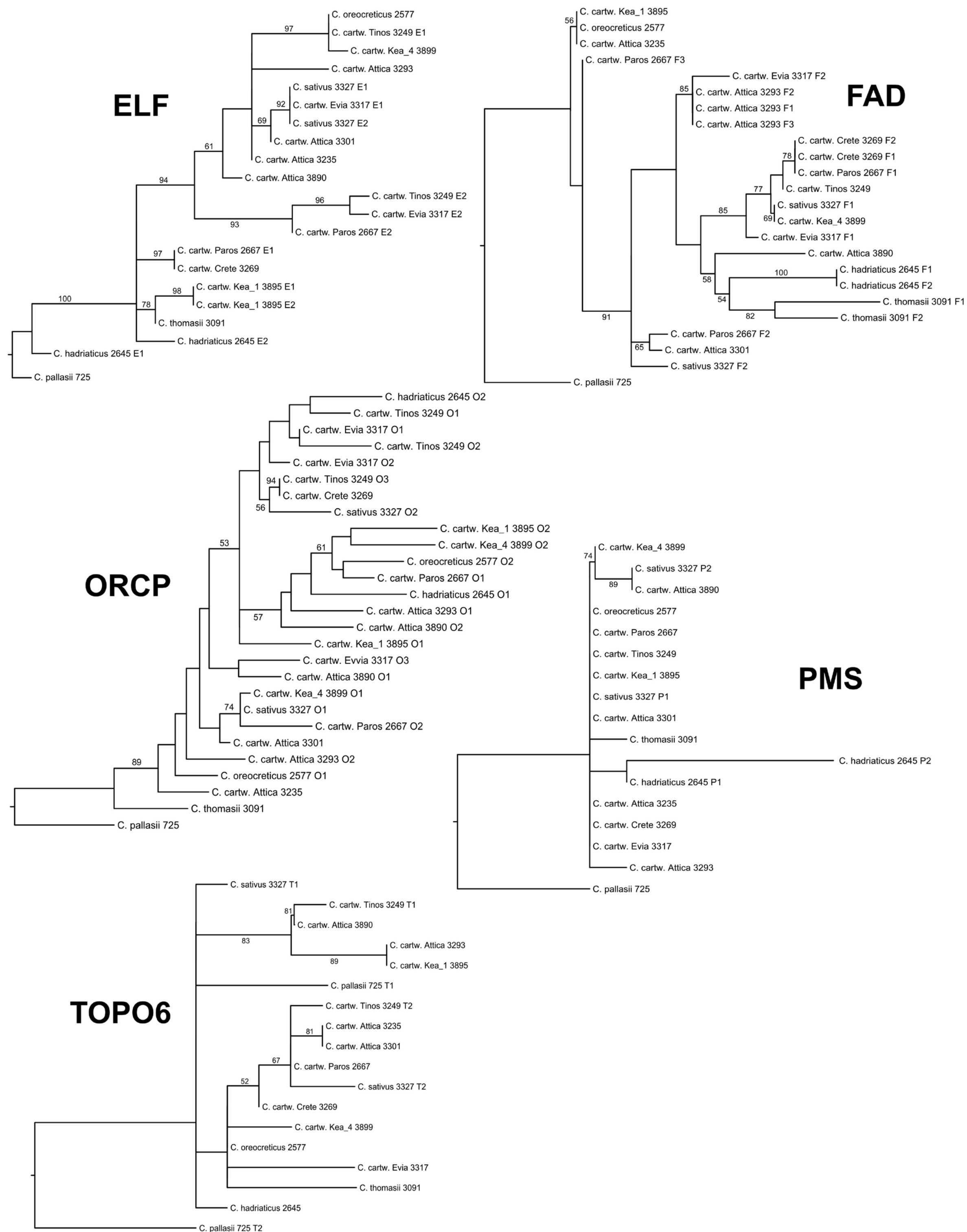


Figure S4. Annotated chloroplast genome of *C. cartwrightianus.* Two other assembled chloroplast genomes (*C. cartwrightianus* from Crete and *C. sativus*) are identical regarding gene order.

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



651

652 **Figure S5.** Maximum-liklihood analyses of five single-copy loci in close relatives of *Crocus*
653 *sativus*. We analyzed multiple individuals of *C. cartwrightianus* to see how allelic diversity of
654 the five genes influences the outcome of phylogenetic analyses regarding the position of *C.*
655 *sativus* in relation to *C. cartwrightianus* and other included species. *Crocus pallasii* was
656 specified as outgroup in all analyses. The presence of multiple copies of the genes was
657 detected via cloning and sequencing of PCR amplicons. Different copies (paralogs or
658 homeologs) present in single individuals are indicated by the first letter of the gene name
659 followed by numbers 1-3. Numbers along branches depict bootstrap support values derived
660 from 500 bootstrap re-samples. For description of the loci see Table S2.

661

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

662

663

664 **Figure S6.** Comparison of *C. sativus* (A) and different individuals of Attic *C. cartwrightianus*
665 (B-D).

666

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

667 **Table S1.** Plant materials used in the study.

Species	Sample origin and ID	Herbarium: ID
C. cartwrightianus Herb.	Anafi: cr2791, cr2792, cr2793, cr2795, cr3905, cr3907, cr2790, cr2939	GAT: HKEP1601, HKEP1603
	Astypalaia: cr2817, cr2818, cr2819, cr2820, cr2821, cr2944, cr2945	GAT: HKEP1609
	Attica: cr2653, cr2654, cr2655, cr3280, cr3281, cr3282, cr3283, cr3284, cr3285, cr3286, cr3300, cr3301, cr3302, cr3303, cr3304, cr3305, cr3306, cr3307, cr2804, cr2806, cr3226, cr3227, cr3228, cr3229, cr3230, cr3231, cr3232, cr3233, cr2798, cr2799, cr2800, cr3235, cr3236, cr3237, cr3238, cr3240, cr3241, cr2679, cr2811, cr2812, cr2814, cr2815, cr2816, cr3244, cr3245, cr3246, cr3247, cr3248, cr3271, cr3272, cr3273, cr3274, cr3275, cr3276, cr3277, cr3278, cr3279, cr3290, cr3291, cr3292, cr3293, cr3294, cr3295, cr3296, cr3297, cr3298, cr3299, cr2679, cr3890, cr3891, cr3892, cr3893, cr3894	GAT: HKEP1592, HKEP1592b, HKEP1605, HKEP1606, HKEP1607, NB_2017-002A, DH1802
	Crete: cr267, cr306, cr2662, cr2663, cr2664, cr3468, cr3472, cr3473, cr3269, cr3270, cr3466	GAT: HKEP1517, HKEP1525, P9430
	Evia: cr3309, cr3310, cr3311, cr3312, cr3313, cr3314, cr3315, cr3316, cr3317	GAT: HKEP1664
	Kea1: cr3895, cr3896, cr3897, cr3898, cr3908, cr3909, cr3910	GAT: KEA1, KEA4
	Kea4: cr3899, cr3900, cr3901, cr3902, cr3914, cr3915, cr3916	
	Mykonos: cr3259, cr3260, cr3261, cr3262, cr3263, cr3264, cr3265, cr3266, cr3267, cr3268	GAT: ZNMYK
	Paros: cr2665, cr2666, cr2667, cr2780, cr2781, cr2782, cr2783, cr3471	GAT: HKEP1445
	Syros: cr2635, cr2636, cr2637, cr2638, cr2639, cr3462, cr3463, cr3476, cr3477	GAT: HKEP1443
	Tinos: cr3249, cr3250, cr3251, cr3252, cr3253, cr3254, cr3255, cr3256, cr3257, cr3258	GAT: ZNTIN
C. sativus L.	Spain: cr3326, cr3327, cr3328, cr3776, cr3778, cr3781, cr3783, cr3784	GAT: 163
	France: cr2436, cr3329, cr3330, cr3777, cr3780, cr3782, cr3785	GAT: 160, 164
	Iran: cr1573	GAT: CsS
	Germany: cr2441, cr3779	GAT: CsG
	Netherlands: cr3331, cr3332, cr3333, cr3334	s.n.
C. hadriaticus Herb.	Greece, Pindus Mts.: cr2645	GAT: HKEP1580
C. oreocreticus B.L.Burtt	Greece, Crete: cr2576, cr2577, cr2578, cr2579, cr2674, cr2675, cr3381, cr3383	GAT: HKEP1585
C. pallasii Goldb.	Rep. Northern Macedonia, Kumanovo: cr725	Niš, s.n.
C. thomasii Ten.	Italy, Apulia: cr3091	GAT: HKEP1614

668

669

bioRxiv preprint doi: https://doi.org/10.1101/537688; this version posted February 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

670

671 **Table S2.** Analyzed genome regions.
672

Locus	Primer name	Primer sequence (5'–3')	Ampli-con size (bp)	Annealing temperature/ elongation time
Nuclear loci				
ORTHO005659, FAD dependent oxidoreductase; intron 5, exon 6, intron 6	FAD_ex5fwd FAD_ex7rwd	ACGTTTGCTCTCATCATCTGGC CCTAGAGCAATCTCAACACCTGC	~700	68 °C/25 sec
ORTHO008517, FAD/NAD(P)-binding oxidoreductase	orcp_f orcp_r	GGCTTCCAAATCTTCCTCACCGC AAAGCCTCTGCAGCTCTTCTCCC	~1100	68 °C/60 sec
ORTHO006661, mismatch repair endonuclease PMS; intron 3	pms_ex3f pms_ex4r	GAAACCAGAACAAAAGATGAGCC TGGGCATATCAACAGGTCTACC	~750	62 °C/30 sec
ORTHO006442, Topoisomerase 6 subunit B; intron 11	croc_Top6_11f[a] croc_Top6_12r[a]	CAGTTTCTTCAACACGAATTT ATGGATTCTAACTATTTGTTGGG	~750	59 °C/25 sec
ORTHO004935, Elongation factor 2A; intron 2, exon 3, intron 3	eIF2A_Ex1f[b] eIF2A_Ex4r[b]	GATGAATCTATTGCTTGCCGAATG[2] GTCATATATTGTAGCCTTCGCAGG[2]	~750	62 °C/30 sec
Chloroplast loci				
*trn*S(GCU) – *trn*G(UCC) IGS	croc_trnSf croc_trnS-Gr	CCTGGTTAAGTAAAGTACTGG TCTCAAGTATTCAAATAAAGC	119/203	54 °C/20 sec
*acc*D – *psa*L IGS	accD_f psaL_IGS_r	TGCTGAACCCAATGCCTACG TTCACGACTTCCTCGAAAGG	~490	57 °C/15 sec

673 [a]Nemati et al. (2018); [b]Erol et al. (2015, *Phytotaxa* 239:223-232)

674

675