

1 **Original Article**

2 **Hybridization and introgression are prevalent in Southern European *Erysimum***
3 **(Brassicaceae) species**

4 Carolina Osuna-Mascaró^{1,2,3*}, Rafael Rubio de Casas^{2,4}, José M. Gómez^{2,5}, João
5 Loureiro⁶, Silvia Castro⁶, Jacob B. Landis^{7,8}, Robin Hopkins^{9,10}, and Francisco
6 Perfectti^{1,2*}

7 ¹Departamento de Genética, Universidad de Granada, Granada, Spain

8 ²Research Unit Modeling Nature, Universidad de Granada, Granada, Spain

9 ³ Present address: Department of Biology, University of Nevada, 1664 N Virginia St, Reno, NV 89557,
10 USA

11 ⁴Departamento de Ecología, Universidad de Granada, Granada, Spain

12 ⁵ Departamento de Ecología Funcional y Evolutiva, Estación Experimental de Zonas Áridas (EEZA-
13 CSIC), Almería, Spain

14 ⁶ Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra,
15 Portugal

16 ⁷BTI Computational Biology Center, Boyce Thompson Institute, Ithaca NY 14853, USA

17 ⁸ School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell
18 University, Ithaca NY, USA

19 ⁹Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

20 ¹⁰The Arnold Arboretum, 1300 Centre Street, Boston, MA, USA

21
22
23
24
25
26
27
28
29
30
31

32 Short title: Hybridization and introgression in *Erysimum* species

33 *Corresponding authors: COM: cosuna@unr.edu; FP: fperfect@ugr.es

34 **ABSTRACT**

35

36 **Background and Aims:** Hybridization is a common and important force in plant evolution. One of its
37 outcomes is introgression - the transfer of small genomic regions from one taxon to another by
38 hybridization and repeated backcrossing. This process is believed to be common in glacial refugia,
39 where range expansions and contractions can lead to cycles of sympatry and isolation, creating
40 conditions for extensive hybridization and introgression. Polyploidization is another genome-wide
41 process with a major influence on plant evolution. Both hybridization and polyploidization can have
42 complex effects on plant evolution. However, these effects are often difficult to understand in recently
43 evolved species complexes.

44 **Methods:** We combined flow cytometry, transcriptomic and genomic analyses, and pollen-tube growth
45 assays to investigate the consequences of polyploidization, hybridization, and introgression on the
46 recent evolution of several *Erysimum* (Brassicaceae) species from the South of the Iberian Peninsula, a
47 well-known glacial refugium. This species complex differentiated in the last 2Myr, and its evolution
48 has been hypothesized to be determined mainly by polyploidization, interspecific hybridization, and
49 introgression.

50 **Key Results:** Our results support a scenario of widespread hybridization involving both extant and
51 “ghost” taxa. Several taxa studied here, most notably those with purple corollas, are polyploids, likely
52 of allopolyploid origin. Moreover, hybridization in this group might be an ongoing phenomenon, as
53 prezygotic barriers appeared weak in many cases.

54 **Conclusions:** The evolution of *Erysimum* spp. has been determined by hybridization to a large extent.
55 The adaptive value of such genomic exchanges remains unclear, but our results indicate the importance
56 of hybridization for plant diversification across evolutionary scales.

57 **Keywords:** Hybridization, Introgression, Polyploidy, Allopolyploidy, Glacial refugium, Brassicaceae,
58 *Erysimum* spp.

59 **INTRODUCTION**

60 Hybridization is widespread across the tree of life, determining the branching and diversification
61 patterns of many taxonomic groups (Rieseberg and Carney, 1998; Coyne and Orr, 2004; Abbott et al.,
62 2013; Arnold, 2016). Because of its pervasiveness, hybridization has been a subject of research for a
63 long time (Stebbins, 1959; Anderson, 1953; Arnold et al., 1999). However, it is only recently, with the
64 advent of next-generation sequencing, that scientists have started to analyze the dynamics of
65 hybridization at the scale of whole genomes, thus rekindling interest in the evolutionary relevance of
66 this phenomenon. Although the patterns of hybridization remain unexplored for many groups, the
67 renewed research efforts have undoubtedly increased our understanding of the role of hybridization in
68 nature (Payseur and Rieseberg, 2016; Goulet et al., 2017; Taylor and Larson, 2019).

69 Hybridization is particularly relevant for plant evolution, with many plant species showing
70 hybrid origins (Mallet, 2005; Soltis and Soltis, 2009). The evolutionary outcomes of hybridization may
71 vary widely. Interspecific hybridization can hinder speciation and therefore diversification (Mayr,
72 1992; Schemske, 2000; Mallet, 2005; Saari and Faeth, 2012; Gómez et al. 2015a), but in other cases,
73 hybridization can actually foster the formation of new species (Rieseberg et al., 2003; Stelkens and
74 Seehausen, 2009) or the introgression of novel genetic variation (by hybridization and repeated
75 backcrossing; Anderson and Hubricht, 1938; Anderson, 1953; Rieseberg and Wendel, 1993). In
76 addition, the fusion of genomes between two hybridizing species can lead to changes in ploidy levels
77 (i.e., allopolyploidization; Soltis et al., 2014). There is evidence that introgression might even span
78 ploidy levels (e.g., gene flow between diploid and tetraploid species of *Senecio*; Chapman and Abbott,
79 2010), which opens intriguing questions about the interplay of introgression and polyploidization.
80 However, the specifics of how hybridization, introgression, and polyploidization interact to affect the
81 evolution of particular plant groups remain poorly understood. Advancements in genomic sequencing
82 technology and analyses are now making the challenges of characterizing these processes far more
83 feasible, even in recently diverged lineages and taxa.

84 *Erysimum* L. is one of the largest genera of the Brassicaceae, comprising more than 200 species
85 (Polatschek, 1986), and has been described as a taxonomically complex genus with a reticulated
86 evolutionary history in which polyploidization may have affected the evolution of some clades
87 (Marhold & Lihová, 2006; Turner, 2006; Abdelaziz, 2013; Muñoz-Pajares, 2013). This genus is
88 distributed mainly in Eurasia, with some species in North America and North Africa (Warwick et al.,
89 2006). Notably, more than a hundred species have been described in the Mediterranean region (Greuter
90 et al., 1986) with particular abundance in the Iberian Peninsula, where twenty-one (Polatschek, 1979;
91 Polatschek, 2014) or twenty-three (Nieto-Feliner, 1993; Mateo et al., 1998) species have been
92 described. Most Iberian *Erysimum* species have yellow flowers, but six have purple corollas (Nieto-
93 Feliner, 1993; Gómez et al., 2015b). Interestingly, previous studies suggested that some purple species
94 may have a recent, hybrid, and allopolyploid origin (Nieto-Feliner, 1992, Nieto-Feliner, 1993;
95 Abdelaziz et al., 2014; Gómez et al. 2014). A history of hybridization could further suggest the
96 possibility that the purple flower color has been transferred across the Iberian clade through
97 hybridization and then maintained by natural selection. This scenario would indicate that introgression
98 and polyploidization are intertwined in this group and might have contributed to the adaptive evolution
99 of *Erysimum* spp.

100 Here we studied signals of hybridization across six species of *Erysimum* (*E. mediohispanicum*,
101 *E. nevadense*, *E. fitzii*, *E. popovii*, *E. baeticum*, *E. bastetanum*) that inhabit the Baetic Mountains, an
102 important and dynamic glacial refugium (Médail and Diadema, 2009). The evolution of several plant
103 species has been hypothesized to have been affected by speciation and secondary contacts in this region
104 (Médail and Diadema, 2009; Nieto-Feliner, 2011). The repeated expansion and contraction of ranges
105 and the subsequent cycles of sympatry and isolation might have created conditions for extensive
106 hybridization, introgression, and allopolyploid formation. This species group appears to have
107 differentiated relatively rapidly within the last 2Myr (Osuna-Mascaró et al., 2021). Previous authors
108 have hypothesized that this rapid evolution has been strongly affected by polyploidization and

109 hybridization, as this group spans several ploidy levels, and some species pairs have been reported to
110 produce fertile hybrids (Abdelaziz et al., 2014; Abdelaziz et al., 2021). Species of this group show
111 characteristics that may facilitate ongoing introgression, such as growing in sympatry in some locations
112 and having a generalist pollination system that renders gene flow among different species possible.

113 The main goal of this study is to disentangle the history of hybridization for the *Erysimum*
114 species complex in the Baetic Mountains. Specifically, we considered both whole-genome effects of
115 hybridization (i.e., the interplay between hybridization and polyploidization) and local, potentially
116 important, introgression of specific genomic regions. Moreover, we also quantified prezygotic barriers
117 among extant taxa to estimate the likelihood of gene flow among them. We test the hypotheses that a)
118 Genomes of this species complex must exhibit signals of multiple hybridization events; b) Some taxa
119 might be allopolyploid, and c) If purple corollas are the product of introgression, hybridization and
120 gene-flow should be detectable, and prezygotic barriers may be weak between (at least some) yellow
121 and purple taxa.

122 MATERIAL AND METHODS

123 Plant samples

124 We studied six species in the genus *Erysimum* collected in the Baetic Mountains, South of Spain (Table
125 1; Figure 1). Specifically, we sampled three different populations for *E. mediohispanicum* (yellow
126 corollas; Em21, Em39, Em71), *E. nevadense* (yellow corollas; En05, En10, En12), *E. popovii* (purple
127 corollas; Ep16, Ep20, Ep27), *E. bastetanum* (purple corollas; Ebt01, Ebt12, Ebt13), and *E. baeticum*
128 (purple corollas; Ebb07, Ebb10, Ebb12), and one population for *E. fitzii* (yellow corollas; Ef01). Some
129 of these species appear in sympatry in some of the sampled localities (e.g., *E. popovii*, Ep20, and *E.*
130 *mediohispanicum*, Em39; Table 1). Additionally, we sampled one population of *E. lagascae* (Ela07), an
131 allopatric diploid species with purple corollas inhabiting Central Spain, posited as one potential
132 parental species of the Baetic Mountain species studied here (Nieto-Feliner, 1993). We collected fully

133 developed flower buds for transcriptomic analyses (five buds from an individual per population) and
134 leaves for flow cytometry (6-10 individuals per population).

135 **Flow cytometry analyses**

136 We used flow cytometry to assess genome size and estimate DNA ploidy levels. Nuclei were isolated
137 from fresh leaf tissues by simultaneously chopping with a razor blade 0.5 cm² of leaf and 0.5 cm² of
138 an internal reference standard (Galbraith et al., 1983). We used *Solanum lycopersicum* L. 'Stupické'
139 with 2C = 1.96 pg or *Raphanus sativus* L. with 2C = 1.11 pg as internal reference standards (Doležel et
140 al., 1992). The nuclei extraction was made on a Petri dish containing 1 ml of WPB buffer (Loureiro et
141 al., 2007). Then, the nuclear suspension was filtered using a 50 µm nylon mesh, and DNA was stained
142 with 50 µg ml⁻¹ of propidium iodide (PI, Fluka, Buchs, Switzerland). Additionally, 50 µg ml⁻¹ of
143 RNase (Fluka, Buchs, Switzerland) was added to degrade dsRNA. After a 5 min incubation, the
144 samples were analyzed in a Sysmex CyFlow Space flow cytometer (532 nm green solid-state laser,
145 operating at 30 mW). Results were acquired using FloMax software v2.4d (Partec GmbH, Münster,
146 Germany) in the form of four graphics: histogram of fluorescence pulse integral in linear scale (FL);
147 forward light scatter (FS) vs. side light scatter (SS), both in logarithmic (log) scale; FL vs. time; and FL
148 vs. SS in log scale. The FL histogram was gated using a polygonal region defined in the FL vs. SS
149 histogram to avoid debris signals. At least 5,000 particles were analyzed per sample. Only CV values of
150 2C peak of each sample below 5% were accepted; otherwise, a new sample was prepared and analyzed
151 until quality standards were achieved (Greilhuber et al., 2007). In a few cases, samples produced
152 histograms of poorer quality even after repetition due to the presence of cytosolic compounds. Thus, it
153 was impossible to estimate ploidy level and/or genome size for some samples (Table 2).

154 We obtained the genome size in mass units (2C in pg; sensu Greilhuber et al., 2005) using the
155 formula: sample 2C nuclear DNA content (pg) = (sample G1 peak mean/reference standard G1 peak
156 mean)* genome size of the reference. The ploidy levels were inferred for each sample based on
157 chromosome counts and genome size estimates available for the genus and species.

158 **RNA extraction and sequencing**

159 Details of the sampling, RNA extraction, and sequencing appear in Osuna-Mascaró et al. (2021). In
160 summary, we stored collected flower buds of each individual in liquid nitrogen until RNA extraction.
161 Floral buds were ground with a mortar and a pestle in liquid nitrogen. We used the Qiagen RNeasy
162 Plant Mini Kit following the manufacturer's protocol to isolate total RNA from 17 samples (one
163 individual per population; three populations of *E. baeticum*, *E. bastetanum*, *E. mediohispanicum*, *E.*
164 *nevadense*, and *E. popovii*, and one population of *E. fitzii* and *E. lagascae*). Then, we checked the
165 quality and quantity of the RNA using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific,
166 Wilmington, Delaware, United States) and agarose gel electrophoresis. Library preparation and RNA
167 sequencing were conducted at Macrogen Inc. (Seoul, Korea). Before sequencing, the quality of the
168 RNA was analyzed again with the Agilent 2100 Bioanalyzer system (Agilent Technologies Inc., Santa
169 Clara, California, United States), and an rRNA-depletion procedure with Ribo-Zero (Illumina, San
170 Diego, California, United States) was used to enrich mRNA content and to avoid the sequencing of
171 rRNA. Library preparation was performed using the TruSeq Stranded Total RNA LT Sample
172 Preparation Kit (Plant). Sequencing of the 17 libraries (one per individual) was carried out using the
173 HiSeq 3000-4000 sequencing protocol and TruSeq 3000-4000 SBS Kit v 3 reagent, following a paired-
174 end 150 bp strategy on the Illumina HiSeq 4000 platform. A summary of sequencing statistics is shown
175 in Table S1 (Supporting Information).

176 **Transcriptome assembly and annotation**

177 Details of the read quality control, trimming, and de novo transcriptome assembly and annotation can
178 be found in Osuna-Mascaró et al. (2021). Briefly, we used FastQC v0.11.5 (Andrews, 2010) to analyze
179 the quality of each library's raw reads. Then, we trimmed the adapters in the raw reads using cutadapt
180 v1.15 (Martin, 2011), and we quality-filtered the reads using Sickle v1.33 (Joshi and Fass, 2011). After
181 trimming, we used FastQC v0.11.5 (Andrews, 2010) again to verify the trimming efficiency. To
182 assemble the resulting high-quality, cleaned reads into contigs, we followed a de novo approach using

183 Trinity v 2.8.4 (Grabherr et al. 2011). Before assembly, each library was normalized in silico to validate
184 and reduce the number of reads using the "insilico_read_normalization.pl" function in Trinity (Haas et
185 al., 2013). Then we used the parameter 'min_kmer_cov 2' to eliminate single occurrence k-mers heavily
186 enriched in sequencing errors, following Haas et al. (2013). Candidate open reading frames (ORF)
187 within transcript sequences were predicted and translated using TransDecoder v 5.2.0 (Haas et al.,
188 2013). We performed functional annotation of Trinity transcripts with ORFs using Trinotate v 3.0.1
189 (Haas, 2015). Sequences were searched against UniProt (UniProt Consortium, 2014), using SwissProt
190 databases (Bairoch and Apweiler, 2000) (with BLASTX and BLASTP searching and an e-value cutoff
191 of 10⁻⁵). We also used the Pfam database (Bateman et al., 2004) to annotate protein domains for each
192 predicted protein sequence. Transcripts were filtered through the eggnoG (Jensen et al., 2007), GO
193 (Gene Ontology Consortium, 2004), and Kegg (Kanehisa and Goto, 2000) annotation databases.

194 **Orthology inference**

195 To reduce redundancy, we clustered the translated sequences using cd-hit v 4.6 (Li and Godzik, 2006),
196 following the steps of the pipeline described in Yang and Smith (2014). For the inference of orthologs,
197 we excluded UTRs and non-coding transcripts, using only coding DNA sequences (CDS) in order to
198 avoid the inclusion of sequencing errors (Yang and Smith, 2014). We identified ortholog genes using
199 the OrthoFinder v 2.3.3 pipeline (Emms and Kelly, 2015). In brief, this pipeline first made a BLASTP
200 analysis with the protein sequences as input for searching the orthogroups (a set of potentially
201 orthologs protein-coding genes derived from a single gene in the last common ancestor of all the
202 species sampled), then clustered and aligned the orthologous sequences using MAFFT v 7.450 (Katoh
203 and Standley, 2013) with default parameters. Finally, we obtained the maximum-likelihood
204 phylogenetic gene trees for all orthogroups using IQ-Tree v 1.6.1 (Nguyen et al., 2014). Then, each
205 orthogroup that contained sequenced from all sampled species was used to infer a species tree using
206 STAG v 1.0.0 (Emms and Kelly, 2019). Then, we used DLCpar v 1.1 (Wu et al., 2014) to reconcile the

207 species tree with the gene trees, considering gene duplication, losses, and incomplete lineage sorting
208 (ILS) as potential causes of discordance among trees.

209 **Phylogenetic reconstruction**

210 We obtained a coalescent species tree using ASTRAL v 5.6.3 (Mirarab et al., 2014) with default
211 parameters. This method reconstructs a species tree from unrooted gene tree topologies. We used the
212 gene trees previously obtained by maximum likelihood by using IQ-Tree v 1.6.1 as input. We used
213 FigTree v 1.4.0 (Rambaut and Drummond, 2012) to visualize and edit the species tree. Then, we
214 compared the alternative tree topologies with the phylogeny obtained from whole chloroplast genome
215 analyses for the same species (presented in Osuna-Mascaró et al., 2021) using the Shimodaira-
216 Hasegawa Test (SH-Test; Shimodaira and Hasegawa, 1999) from the R package phangorn v 2.5.5
217 (Schliep, 2011). Both phylogenies were also compared visually, plotting them as mirror images with
218 the function cophyloplot, using the R package ape v 5.4 (Paradis et al. 2004).

219 **Variant calling**

220 We first ran a variant calling analysis, using the *E. lagascae* transcriptome as a reference. We indexed
221 the *E. lagascae* transcriptome using BWA v 0.7.17 (Li and Durbin, 2009) to create a reference and then
222 mapped all the trimmed raw reads to it using the BWA v 0.7.17 "mem" option. We used SAMtools v
223 1.7 (Li et al., 2009) to convert and sort the alignment files. We then called SNPs using the SAMtools v
224 1.7 "mpileup" command. Lastly, we used bcftools v 1.9 to filter the SNPs (Narasimhan et al., 2016),
225 running the SAMtools v 1.7 Perl script "vcfutils.pl VarFilter" with default parameters to filter down the
226 candidate variants and to eliminate false positives.

227 **Discriminant Analysis of Principal Components (DAPC)**

228 We conducted a Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010) of the
229 SNP data to group the different genotypes avoiding any prior subjective bias using the R package
230 adegenet v 2.1.3 (Jombart and Ahmed, 2011). DAPC is a multivariate method that identifies and
231 describes clusters of genetically related individuals from large datasets, providing a measure of the

232 optimal number of genetic clusters (K) across a range of K values by using the Bayesian Information
233 Criterion (BIC). We set a range of K values from two to seven since K=7 is the number of different
234 species in our dataset. The existence of significant hybridization and introgression would result in $K <$
235 7. To identify the optimal number of K, we selected the model with the lowest BIC.

236 **Phylogenetic inference of introgression**

237 As a first step to detect introgression events between species pairs, we computed phylogenetic species
238 networks. This approach provides a graphical extension of the phylogenetic tree model, representing
239 the gene flow by edges connecting the OTUS that are likely to be linked by introgression. We used the
240 software PhyloNet v 3.6.9 (Than et al., 2008; Wen et al., 2018), which implements a phylogenetic
241 network method based on the frequencies of rooted trees accounting for incomplete lineage sorting
242 (ILS). To generate the input for PhyloNet, we first ultrametricized the trees obtained previously with
243 IQ-Tree v 1.6.1, using the "nnls" method in the "force.ultrametric" function within the R package
244 phytools v 0.6-99 (Revell, 2012). Due to computational limitations, we inferred the species networks
245 using a maximum pseudo-likelihood method (MPL) (Yu and Nakhleh, 2015). We performed the search
246 five times to avoid getting stuck at local optima. We estimated optimal networks among an optimal
247 computational range of 0 to 15 introgression events, determining the most likely network based on
248 Akaike's Information Criterion (AIC; Bozdogan, 1987) with the generic function for AIC in R package
249 stats v 3.6.1. As AIC may not provide precise values when using pseudo-likelihood phylogenetic
250 networks (Cao et al., 2019), we also estimated the more optimal network by slope heuristic of log-
251 likelihood values. The optimal network was then visualized with Dendroscope v 3.5.10 (Huson and
252 Scornavacca, 2019).

253 **ABBA-BABA statistic**

254 To assess gene flow between species, we calculated D-statistics, also known as the ABBA-BABA
255 statistic (Durand et al., 2011). To evaluate introgression among the seven species, we used the software
256 Dsuite v 0.1 (Malinsky, 2019), which allows the assessment of gene flow across large datasets and

257 directly from a variant call format (VCF) file. This algorithm computes the D statistic by considering
258 multiple groups of four populations: P1, P2, P3, and O, grouped in asymmetric trees of the form (((P1,
259 P2), P3), O). The site patterns are ordered such that the pattern BBAA refers to P1 and P2 sharing the
260 derived allele (B-derived allele, A-ancestral allele), ABBA to P2 and P3 sharing the derived allele, and
261 BABA to P1 and P3 sharing the derived allele. The ABBA and BABA patterns are expected to occur
262 with equal frequencies, assuming no gene flow (null hypothesis), while a significant deviation from
263 that suggests possible introgression. To assess whether D is significantly different from zero, Dsuite
264 uses a standard block-jackknife procedure (Green et al., 2010; Durand et al., 2011), obtaining
265 approximately normally distributed standard errors. As recommended by Malinsky (2019), we used a
266 conservative approach estimating the statistic D_{min}, which gives the lowest D-statistic value in a given
267 trio. We used the ruby script "plot_d.rb" to plot into a heatmap the introgression among all the pairs of
268 samples. To complement these analyses, we computed the F_{branch} statistic implemented in Dsuite v
269 0.1 (Mallinsky et al., 2018, Mallinsky et al., 2019). The statistic allows the identification of gene flow
270 events within specific internal branches of a phylogeny. Thus, evaluating the excess sharing of alleles
271 between one species and the descendant or ancestral species, helping to understand when the gene flow
272 happened. We used the whole chloroplast genomes phylogeny from Osuna-Mascaró et al. (2021) in
273 Newick format to establish a reference phylogeny and specify which species could be more accurately
274 treated as sister species (i.e., as P1 and P2) while always using *E. lagascae* as an outgroup.

275 **Pollen tube growth**

276 The existence of prezygotic barriers can fully impede interspecific hybridization. Therefore, the
277 existence of such barriers may indicate that gene flow across a given set of species is highly unlikely,
278 while the lack of such barriers may indicate plausible hybridization and introgression. To explore the
279 existence of prezygotic barriers, we carried out a preliminary experiment on the growth of pollen tubes
280 on a reduced set of co-occurring species (Table 1). We collected 20 individual plants of each *E.*
281 *mediohispanicum*, *E. bastetanum*, and *E. popovii* from natural populations. We grew the plants in a

282 common garden (University of Granada facilities) and moved them into a greenhouse before flowering
283 to exclude pollinators. When the flowers opened, we performed hand-pollination experiments by
284 tipping the anther with a small stick to remove the pollen and placing it on the stigma of a flower from
285 different species previously emasculated (hybrid crosses) or of a flower from the same species but
286 different populations previously emasculated (intra-specific crosses). Moreover, we emasculated some
287 flowers and hand-pollinated them with their own pollen (forced selfing crosses), and some flowers
288 were not manipulated and left for spontaneous self-pollination (spontaneous selfing crosses).

289 We collected the pistils after 72 hours and preserved them in ethanol at 4°C until staining of
290 pollen tubes, following the Mori et al. (2006) protocol with minor modifications. In brief, each pistil
291 was cleaned in 70% EtOH for ten minutes and then moved to 50% EtOH, 30% EtOH, and finally
292 distilled water. We softened the samples by placing them into a small petri dish of 8 M NaOH for one
293 hour at room temperature (as recommended by Kearns and Inouye, 1993). Then, we transferred the
294 pistils to distilled water for ten minutes, and afterward, the stigmas were incubated with 0.1 % aniline
295 blue in phosphate buffer (pH 8.3) for two hours. The final slide preparations were examined under a
296 fluorescence microscope with blue light (410 nm) to observe and measure pollen tube development.

297 **RESULTS**

298 **Ploidy levels**

299 Flow cytometry revealed a wide variation in genome size and, therefore, in DNA ploidy levels across
300 but also within species (Table 2). We found that all samples of *E. fitzii* and *E. nevadense* were diploid.
301 The other species with yellow corollas, *E. mediohispanicum*, also appeared to be predominantly
302 diploid, although the Em71 population deviated from this pattern being tetraploid. The genome size of
303 *E. lasgacae* also corresponded to that of a diploid, while the other purple corolla species showed ploidy
304 levels higher than diploidy (Table 2). Moreover, ploidy levels differed across populations in two of
305 these species. Populations of *E. bastetanum* varied between 4x and 8x, while in *E. popovii*, the range

306 was even greater, from 4x to 10x. In three cases (Ebb12, En05, and En10; Table 2), it was not possible
307 to establish the ploidy level of the samples, and we used those reported in Blanca et al. (1992).

308 **Transcriptome assembly and orthology inference**

309 The sequencing results and the corresponding summary statistics of the assembled transcriptomes can
310 be found in Osuna-Mascaró et al. (2021). In summary, we obtained between 104K and 382K different
311 Trinity transcripts, producing between 66K and 235K Trinity isogenes. The total assembled bases
312 ranged from 92 Mbp (in the Em21 population of *E. mediohispanicum*) to 319 Mbp (in En10 population
313 of *E. nevadense*). The number of annotated unigenes ranged between 71,606 (*E. nevadense*, En12) and
314 197,069 (*E. baeticum*, Ebb10); mean value 146,314.35. The highest proportion of annotated unigenes
315 was obtained using BLASTX to search against the SwissProt reference database. Details of the
316 annotated unigenes using different protein databases can be found in Osuna-Mascaró et al. (2021).
317 OrthoFinder assigned 1,519,064 protein gene sequences (96.4% of total) to 92,984 gene families
318 (orthogroups) (Table S2). Among them, 16,941 orthogroups were shared by all species, and their
319 corresponding gene trees were used for further analyses.

320 **Phylogenetic trees and population clustering**

321 We inferred a coalescence tree using the 16,941 maximum likelihood gene trees obtained with IQ-Tree
322 as input for ASTRAL (Figures 2 and S1). This species tree was nearly fully resolved with high support,
323 having only four nodes with low quartet scores results (posterior probabilities for these nodes: 0.78,
324 0.77, 0.70, and 0.53; see Figure S1). In this tree, rooted with *E. lagascae*, the 4× population of *E.*
325 *mediohispanicum* (yellow corollas; Em71) appeared as the first branching OTU. Three clades, although
326 with low supports, were evident. A clade formed by *E. bastetanum* and *E. baeticum*, both species were
327 having purple corollas; another clade including *E. fitzii* (yellow corollas) and the three populations of
328 *E. popovii* (purple corollas); and the last clade including the populations of *E. nevadense* and the 2x
329 populations of *E. mediohispanicum*, both species with yellow corolla. Although there is some species
330 clustering, not all species appear to be monophyletic, supporting a history of hybridization. Moreover,

331 when comparing the species tree with the whole chloroplast genomes phylogeny (Figure 2), we find
332 clear cytonuclear discordances resulting in a significant SH test result (Diff $-\ln L = 345426.4$, p-value <
333 0.01). This lack of congruence among both phylogenies also supports the hybridization hypothesis.

334 The discriminant analysis revealed K=4 and K=5 as the most likely number of genetic clusters
335 (Figure 3), both with very similar BIC values (K=4, BIC= 189.99; K=5, BIC= 188.99). The clusters
336 corresponding to K=4 produced the same clusters that appeared in the coalescence tree (Figure 2).
337 However, the clusters corresponding to K=5 included three monospecific groups (for *E. lagascae*
338 -purple corollas-, *E. fitzii* -yellow corollas-, and *E. popovii* -purple corollas-), one for the diploid
339 species with yellow corollas (the three populations of *E. nevadense* and the diploid populations of *E.*
340 *mediohispanicum*), and the last including all the populations of *E. baeticum* (purple corollas), *E.*
341 *bastetanum* (purple corollas), *E. popovii* (purple corollas) and Em71, the 4x population of *E.*
342 *mediohispanicum* (yellow corollas).

343 **Analysis of introgression**

344 The network with 13 reticulation instances appeared as the most reliable based on the AIC values for
345 the log-likelihood of the networks (Table S3). The estimates of slope heuristic of log-likelihood values
346 also supported the network with 13 reticulation instances as the most reliable network estimated. This
347 network shows frequent hybridization events in the genealogy of these populations involving yellow
348 and purple species (Figure 4), as indicated by the edges connecting tree branches between different
349 populations and species. Notably, this network includes edges connecting non-terminal branches (see
350 Figure 4), which indicates reticulations with past extinct taxa (i.e., “ghost species”) or incomplete
351 sampled taxa.

352 The ABBA-BABA analyses support this scenario of frequent hybridization, even using a
353 conservative approach (D-min). We summarized the tested topologies and the inferred D-statistics with
354 corrected p-values for all triplet combinations in Table S4 and Figure S2. The highest signal of
355 introgression occurred between *E. fitzii* (yellow corollas) and *E. baeticum* (purple corollas; populations

356 Ebb12 and Ebb10) and *E. popovii* (purple corollas; Ep16); and between *E. popovii* (purple corollas;
357 Ep16) and *E. bastetanum* (purple corollas; Ebt12) and *E. baeticum* (purple corollas; Ebb07, Ebb10,
358 Ebb12). There was also evidence of interspecific gene flow as manifested by the fbranch statistic
359 (Figure S3) that identifies gene flow events into specific internal branches of the phylogeny while
360 accounting for potential false-positive results due to correlated introgression signatures among closely-
361 related species. Specifically, we found the highest signal of gene flow between *E. bastetanum* (purple
362 corollas; Ebt12) and *E. mediohispanicum* (yellow corollas; Em71, Em21), *E. nevadense* (yellow
363 corollas; En12, En05), and *E. baeticum* (purple corollas; Ebb07, Ebb10); between *E. bastetanum*
364 (purple corollas; Ebt13) and *E. mediohispanicum* (yellow corollas; Em71) and *E. popovii* (purple
365 corollas; Ep20); between *E. bastetanum* (purple corollas; Ebt01) and *E. mediohispanicum* (yellow
366 corollas; Em21, Em39) and *E. nevadense* (yellow corollas; En12). In addition, we have detected other
367 gene flow events with ancestral or non-sampled taxa (Figure S3).

368 **Pollen tube growth**

369 A total of 103 preparations of *Erysimum* pistils were examined: 52 from hybrid crosses, 24 from forced
370 selfing crosses, 16 from spontaneous selfing crosses, and 11 from intra-specific crosses (Table S5). Our
371 results showed full growth of pollen tubes (i.e., reaching the ovary) in 63.33 % of intra-specific crosses,
372 51.92 % of hybrid crosses ($\chi^2= 0.50$, p-value = 0.513 compared with intra-specific crosses), and only in
373 29,16 % of forced selfing crosses ($\chi^2= 3.73$, p-value= 0.074), and in 25,16 % of spontaneous selfing
374 crosses ($\chi^2= 4.03$, p-value=0.057). Although these last values were not significant, when selfing classes
375 were pooled, it showed a significant reduction in the growth of pollen tubes ($\chi^2= 4.93$, p-value= 0.039)
376 (Figure S4). Cases in which pollen tubes grew but did not reach the ovary were treated as non-growing.
377 In these cases, we could not estimate whether tube growth had completely stopped or if it was ongoing
378 but developed too slowly to reach the ovary during the duration of the experiment.

379 **DISCUSSION**

380 Our results suggest that the *Erysimum* species studied here have a strong signature of hybridization and
381 introgression in their genomes. This result is supported by the pollen tube growth experiments that
382 showed that pollen tubes could grow all the way to the ovary in some hybrid crosses, indicating very
383 weak or non-existent prezygotic barriers. Moreover, we found that species with purple flowers are
384 polyploid and have a strong signature of introgression, suggesting an allopolyploid origin. We also
385 found a hybridization signature in the (mostly diploid) yellow species, indicating that hybridization
386 occurred across both colors and ploidy levels.

387 Several phylogenetic reconstructions have been performed for western Mediterranean
388 *Erysimum* species (e.g., Abdelaziz et al. 2014; Gómez et al. (2014, 2015); Züst et al. (2020)) that have
389 used different strategies (several populations per species or only one representative per species; several
390 nuclear and cytoplasmic sequences or NGS transcriptomic data). Abdelaziz et al. (2014) found that
391 populations of *E. mediohispanicum* and *E. bastetanum*, species analyzed here, did not appear as
392 monophyletic (with some populations placed within other branches of the phylogeny), which was
393 indicative that introgression probably produced important reticulation at the population level. Our
394 analyses support this hypothesis. The reticulate nature of these phylogenies imposes some caution in
395 interpreting phylogenies based on only a few nuclear or cytoplasmic sequences, as suggested by Chan
396 & Levin (2005). In these cases, major divisions may reflect the reality of some old phylogenetic splits.
397 However, it will be challenging for more recent speciation events to obtain a clear picture of the
398 phylogeny without interrogating complete genomes or transcriptomes.

399 Overall, our results support a hybrid origin for the purple polyploid *Eysimum* Iberian species, as
400 suggested in previous studies (Nieto-Feliner, 1993; Abdelaziz et al., 2014; Gómez et al., 2015b; Osuna-
401 Mascaró, 2020). In particular, we found support for *E. popovii* (purple corollas and polyploid) and *E.*
402 *fitzii* (yellow corollas and diploid) as sister species. Also, the genome of *E. popovii* exhibited signatures
403 of a hybridization process in which *E. fitzii* may have been involved. The possible hybrid origin of *E.*

404 *popovii* with *E. fitzii* as a potential parental taxon was previously proposed by Nieto-Feliner (1993)
405 based on morphology. Similarly, a hybrid origin of *E. baeticum* (purple corollas and polyploid) had
406 been previously suggested, in this case with *E. nevadense* (yellow corollas and diploid) implicated as a
407 potential parent (Nieto-Feliner, 1992). Our results showed that these two species appear closely related,
408 and *E. baeticum* may have had an introgression signature of *E. nevadense*. Moreover, our results also
409 suggested a complex scenario for *E. bastetanum* (purple corollas and polyploid), which appears closely
410 related to *E. baeticum* (purple corollas and polyploid). In fact, *E. bastetanum* has been considered a
411 subspecies of *E. baeticum* until recently (Lorite et al., 2014). Therefore, the general pattern that
412 emerged from our results is that these purple species are polyploids of hybrid origin, descending from
413 crosses between an unidentified parent and some diploid, often yellow taxon.

414 However, our results also suggested a complex evolutionary history for the mostly diploid
415 yellow species. The contributing lineages also often involve unidentified taxa. This might be
416 attributable to insufficient sampling, as we did not include some *Erysimum* species (*E. ronda*, *E.*
417 *myriophyllum*, -yellow corollas-, and *E. cazorlense* -purple) that also inhabit the Baetic Mountains,
418 although with a limited distribution (Nieto-Feliner, 1993). At this point, it is impossible to establish
419 whether these taxa may have acted as a source of introgression. In any case, our results show that
420 hybridization and introgression are major contributors to the evolutionary history of this species
421 complex, deserving further research.

422 Interestingly, we did not find a consistent, predictable pattern of hybridization for most species.
423 Populations of the same species showed differences in their hybridization history, as shown by the
424 ABBA-BABA test (which detected multiple and diverse introgression events) and the PhyloNet
425 reconstructions (which yielded a tree with 13 reticulations as the most optimal network). In the same
426 vein, the DAPC results did not support a scenario with populations clustered by species. Our results are
427 similar to those of previous studies describing asymmetric hybridization patterns as a consequence of
428 differences in ecological pressures across populations and geographical areas (Payton et al., 2019; Sujii

429 et al., 2019; Wang et al., 2019). At this stage, we cannot unambiguously identify any ecological factor
430 behind the asymmetries we detected. However, we did observe variation in pollinators' preferences and
431 flowering time across populations, which might lead to local differences in gene-flow patterns. Thus, to
432 fully understand this asymmetry in hybridization and why some populations have more introgression
433 signatures than others, future studies considering different ecological pressures for these species and
434 including pollinator censuses of wild populations are required.

435 Evidence of hybridization between at least some of these species has been reported previously
436 (Abdelaziz 2014). Thus, *E. mediohispanicum* and *E. nevadense* show a hybrid zone in a sector of the
437 Spanish Sierra Nevada (Abdelaziz et al., 2021). Pollinators do not appear to constitute strong pre-
438 pollinating barriers since all of these species are extreme generalists and share most pollinators (Gómez
439 et al. 2015b). Moreover, we have found that prezygotic, post-pollination barriers may not be effective
440 since pollen tubes are often growing in hybrid crosses. Contemporary gene flow between different
441 cytotypes of *E. mediohispanicum* seems negligible, as evidenced by an almost complete absence of
442 triploids and other minority cytotypes in the contact zone between tetraploid and diploid populations of
443 this species (Muñoz-Pajares et al. 2018). Historical dynamics of genetic isolation and sympatry might
444 have also played a role (Albaladejo and Aparicio, 2007; Rifkin et al., 2019; Zielinski et al., 2019).
445 These *Erysimum* species are located in a well-known glacial refugium (Médail and Diadema, 2009;
446 Hughes and Woodward, 2017), and thus, the isolation and then re-establishment of gene flow (i.e.,
447 secondary contact zones) among populations of different species may have favored locally specific
448 hybridization patterns (Coyne, 2004; Harrison and Larson, 2014; Arnold, 2015). A better knowledge of
449 the historical dynamics of species and populations and past ranges overlap is required to fully
450 understand the genomic pattern of divergence between closely related species. For instance, combining
451 macroecological methods with niche models and phylogenetic approaches could clarify the opportunity
452 for hybridization through evolutionary time (Folk et al., 2018; Aguirre-Liguori et al., 2021).

453 Furthermore, we detected signatures of ghost introgression, implying that ancestral species have
454 influenced the hybridization history of these *Erysimum* species. This result was first evidenced by
455 cytonuclear discordance, which might be due to past organellar introgression from extinct species
456 (Huang et al., 2014; Folk et al., 2017; Lee-Yaw et al., 2019). We also found a clear signature of
457 ancestral introgression in the phylogenetic species network, in which some of the reticulations appeared
458 from introgression involving "ghost" taxa. Similarly, the fbranch statistic identified gene flow events in
459 internal branches that concurred with introgression with ghost species. Specifically, we observed that
460 some ancestral form of *E. popovii* (purple corollas and polyploid) could have been related to *E. fitzii*
461 (yellow corollas and polyploid). Also, we detected evidence of gene flow between an ancestor of *E.*
462 *mediohispanicum* (yellow corollas and diploid; Em21), *E. bastetanum* (purple corollas and polyploid),
463 and *E. baeticum* (purple corollas and polyploid). Moreover, the results showed that many past gene
464 flow events could have occurred between *E. baeticum* (purple corollas and polyploid), *E. nevadense*
465 (yellow corollas and diploid), and *E. bastetanum* (purple corollas and polyploid). In light of these
466 results, it seems that some unidentified ancestral species played a role as introgression sources for both
467 the purple and yellow species. However, as previously noticed, we include only a subset of the Iberian
468 *Erysimum* species in this study; accordingly, we may be mistaking the signal of the unsampled species
469 for that of ancestral taxa. Further research about the ghost introgression's influence on *Erysimum*
470 evolution, including all the Iberian species and high-quality genome assemblies, would be required to
471 understand the hybridization history thoroughly.

472 CONCLUSIONS

473 Our results indicate that complex evolutionary dynamics have shaped present-day Iberian *Erysimum*
474 diversity. The genomes of extant taxa are the product of multiple polyploidizations, hybridization, and
475 introgression events. Understanding these multi-faceted processes and their interplay is crucial to
476 characterize the evolution of *Erysimum* spp. and probably, of angiosperms in general. Although the
477 evolution of the Iberian *Erysimum* might have been particularly dynamic, this group could be

478 representative of the evolutionary response of multi-species complexes to drastic environmental
479 fluctuations. Further research that incorporates a wider taxonomic sample, whole-genome sequences,
480 and complex demographic and evolutionary statistical methods is needed to precisely characterize the
481 patterns described here.

482 **Author contributions**

483 COM, RR, JMG, and FP conceived and designed the study. COM analyzed the data with the help of FP,
484 JL, and RH. JL and SC performed the flow cytometry analyses. COM wrote the first draft. The final
485 version of the M.S. was redacted with the contribution of all the authors.

486 **Acknowledgments**

487 The authors thank Modesto Berbel Cascales, Tatiana López Pérez, Mercedes Sánchez Cabrera, Raquel
488 Sánchez Fernández, Javier Valverde, and Mohamed Abdelaziz, for their help in the lab and fieldwork.
489 Thanks to Pamela Soltis and Douglas Soltis for their help during the first steps of this work. The
490 authors thank the Sierra Nevada National Park headquarters for providing the permits to work in the
491 National Park.

492 **Funding information**

493 This research is supported by grants from FEDER/Junta de Andalucía-Consejería de Economía y
494 Conocimiento A-RNM-505-UGR18 and P18-FR-3641. This research was also funded by the Spanish
495 Ministry of Science and Innovation (CGL2016-79950-R, CGL2017-86626-C2-2-P), including EU
496 FEDER funds. COM was supported by the Ministry of Economy and Competitiveness (BES-2014-
497 069022).

498 **LITERATURE CITED**

499 Abbott R, Albach D, Ansell S, et al. 2013. Hybridization and speciation. *Journal of Evolutionary*
500 *Biology* 26: 229-246.

- 501 Abdelaziz M. 2013. How species are evolutionarily maintained? Pollinator-mediated divergence and
502 hybridization in *Erysimum mediohispanicum* and *Erysimum nevadense*. PhD Thesis,
503 Universidad de Granada, Spain.
- 504 Abdelaziz M, Muñoz-Pajares AJ, Lorite J, Herrador MB, Perfectti F, Gómez Reyes JM. 2014.
505 Phylogenetic relationships of *Erysimum* (Brassicaceae) from the Baetic Mountains (se Iberian
506 peninsula). *Anales del Jardín Botánico de Madrid* 71: e005.
- 507 Abdelaziz M, Muñoz-Pajares AJ, Berbel M, García-Muñoz A, Gómez JM, Perfectti F. 2021.
508 Asymmetric reproductive barriers and gene flow promote the rise of a stable hybrid zone in the
509 Mediterranean high mountain. *Frontiers in Plant Science* 12.
- 510 Aguirre-Liguori JA, Ramírez-Barahona S, Gaut BS. 2021. The evolutionary genomics of species'
511 responses to climate change. *Nature Ecology & Evolution* 5: 1-11.
- 512 Ai H, Fang X, Yang B, et al. 2015. Adaptation and possible ancient interspecies introgression in pigs
513 identified by whole-genome sequencing. *Nature Genetics* 47: 217-225.
- 514 Albaladejo RG, Aparicio A. 2007. Population genetic structure and hybridization patterns in the
515 Mediterranean endemics *Phlomis lychnitis* and *P. crinita* (Lamiaceae). *Annals of Botany* 100:
516 735-746.
- 517 Anderson E. 1953. Introgressive hybridization. *Biological Reviews* 28: 280-307.
- 518 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
- 519 Arnold ML, Bulger MR, Burke JM, Hempel AL, Williams JH. 1999. Natural hybridization: how low
520 can you go and still be important?. *Ecology* 80: 371-381.
- 521 Arnold ML. 2015. Divergence with genetic exchange. OUP Oxford.

- 522 Arnold BJ, Lahner B, Da Costa JM, et al. 2016. Borrowed alleles and convergence in serpentine
523 adaptation. *Proceedings of the National Academy of Sciences* 113: 8320-8325.
- 524 Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement
525 TrEMBL in 2000. *Nucleic Acids Research* 28: 45-48.
- 526 Barlow A, Cahill JA, Hartmann S, et al. 2018. Partial genomic survival of cave bears in living brown
527 bears. *Nature Ecology & Evolution* 2: 1563-1570.
- 528 Bateman A, Coin L, Durbin R, et al. 2004. The Pfam protein families database. *Nucleic Acids Research*
529 32: D138-D141.
- 530 Bozdogan H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and
531 its analytical extensions. *Psychometrika* 52: 345-370.
- 532 Cao Z, Liu X, Ogilvie HA, Yan Z, Nakhleh L. 2019. Practical aspects of phylogenetic network analysis
533 using phylonet. *BioRxiv* doi: <https://doi.org/10.1101/746362>.
- 534 Chan KM, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of
535 maternally inherited DNA. *Evolution* 59: 720-729.
- 536 Chapman MA, Abbott RJ. 2010. Introgression of fitness genes across a ploidy barrier. *New Phytologist*
537 186: 63-71.
- 538 Coyne JA, Orr HA. 2004. Speciation Sinauer Associates. *Sunderland MA* 276: 281.
- 539 Doležal J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric
540 estimation of nuclear DNA content in plants. *Physiologia Plantarum* 85: 625-631.
- 541 Doležal J, Greilhuber J, Suda J. 2007. Flow cytometry with plant cells: analysis of genes, chromosomes
542 and genomes. John Wiley & Sons.
- 543 Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely
544 related populations. *Molecular Biology and Evolution* 28: 2239-2252.

- 545 Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome
546 comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.
- 547 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics.
548 *Genome Biology* 20: 1-14.
- 549 Folk RA, Mandel JR, Freudenstein JV. 2017. Ancestral gene flow and parallel organellar genome
550 capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology*
551 66: 320-337.
- 552 Folk RA, Visger CJ, Soltis PS, Soltis DE, Guralnick RP. 2018. Geographic range dynamics drove
553 ancient hybridization in a lineage of angiosperms. *The American Naturalist* 192: 171-187.
- 554 Galbraith DW, Harkins KR, Maddox JM, et al. 1983. Rapid flow cytometric analysis of the cell cycle
555 in intact plant tissues. *Science* 220: 1049-1051.
- 556 Gene Ontology Consortium. 2004. The Gene Ontology (GO) database an informatics resource. *Nucleic
557 Acids Research* 32: D258-D261.
- 558 Gokhman D, Mishol N, de Manuel M, et al. 2019. Reconstructing denisovan anatomy using DNA
559 methylation maps. *Cell*, 179: 180-192.
- 560 Gómez JM, González-Mejias A, Lorite J, Abdelaziz M, Perfectti F. 2015a. The silent extinction:
561 Climate change and the potential for hybridization-mediated extinction of endemic high-
562 mountain plants. *Biodiversity and Conservation* 24: 1843–1857.
- 563 Gómez JM, Perfectti F, Lorite J. 2015b. The role of pollinators in floral diversification in a clade of
564 generalist flowers. *Evolution* 69: 863-878.
- 565 Gómez JM, Perfectti F, Klingenberg CP. 2014. The role of pollinator diversity in the evolution of
566 corolla-shape integration in a pollination-generalist plant clade. *Philosophical Transactions B*
567 369: 1–11

- 568 Goulet BE, Roda F, Hopkins R. 2017. Hybridization in plants: old ideas, new techniques. *Plant*
569 *Physiology*, 173: 65-78.
- 570 Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-Seq
571 data without a reference genome. *Nature Biotechnology* 29: 644.
- 572 Green RE, Krause J, Briggs AW, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:
573 710-722.
- 574 Greilhuber J, Doležel J, Lysak MA, Bennett MD. 2005. The origin, evolution and proposed
575 stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Annals*
576 *of Botany* 95: 255-260.
- 577 Greuter W, Burdet HM, Long G. 1986. Dicotyledones (Convolvulaceae-Labiatae). *Med-Checklist* 3:
578 106-116.
- 579 Haas BJ, Papanicolaou A, Yassour D, et al. 2013. *De novo* transcript sequence reconstruction from
580 RNA-Seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:
581 1494.
- 582 Haas BJ. 2015. Trinotate: transcriptome functional annotation and analysis.
- 583 Hanušová K, Ekrt L, Vit P, Kolář F, Urfus T. 2014. Continuous morphological variation correlated with
584 genome size indicates frequent introgressive hybridization among *Diphasiastrum* species
585 (Lycopodiaceae) in Central Europe. *PLoS One* 9: e99552.
- 586 Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries.
587 *Journal of Heredity* 105: 795-809.
- 588 Huang DI, Hefer CA, Kolosova N, Douglas CJ, Cronk QC. 2014. Whole plastome sequencing reveals
589 deep plastid divergence and cytonuclear discordance between closely related balsam poplars,
590 *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytologist* 204: 693-703.

- 591 Hughes PD, Woodward JC. 2017. Quaternary glaciation in the Mediterranean mountains: a new
592 synthesis. *Geological Society London Special Publications* 433: 1-23.
- 593 Huson DH, Scornavacca C. 2019. User Manual for Dendroscope V 3.6.2.
- 594 Jensen LJ, Julien P, Kuhn M, et al. 2007. eggNOG: automated construction and annotation of
595 orthologous groups of genes. *Nucleic Acids Research* 36: D250-D254.
- 596 Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method
597 for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
- 598 Jombart T , Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.
599 *Bioinformatics* 27: 3070-3071.
- 600 Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ
601 files (Version 1.33) [Software].
- 602 Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*
603 28: 27-30.
- 604 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements
605 in performance and usability. *Molecular Biology and Evolution* 30: 772-780.
- 606 Kuhlwilm M, Han S, Sousa VC, Excoffier L, Marques-Bonet T. 2019. Ancient admixture from an
607 extinct ape lineage into bonobos. *Nature Ecology & Evolution* 3: 957-965.
- 608 Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. 2019. An evaluation of alternative
609 explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New*
610 *Phytologist* 221: 515-526.
- 611 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
612 nucleotide sequences. *Bioinformatics* 22: 1658-1659.
- 613 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
614 *Bioinformatics* 25: 1754-1760.

- 615 Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools.
616 *Bioinformatics* 25: 2078-2079.
- 617 Liu L, Bosse M, Megens HJ, et al. 2019. Genomic analysis on pygmy hog reveals extensive
618 interbreeding during wild boar expansion. *Nature Communications* 10: 1-9.
- 619 Loureiro J, Rodriguez E, Doležal J, Santos C. 2007. Two new nuclear isolation buffers for plant
620 DNA flow cytometry: a test with 37 species. *Annals of Botany* 100: 875-888.
- 621 Malinsky M, Matschiner M, Svardal H. 2021. Dsuite-Fast D-statistics and related admixture evidence
622 from VCF files. *Molecular Ecology Resources* 21: 584-595.
- 623 Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20: 229-
624 237.
- 625 Mateo G, Villalba MBC, Udias SL. 1998. Acerca del orófito minusvalorado de la Sierra de Javalambre
626 (Teruel). *Flora Montiberica* 9: 41-45.
- 627 Marhold K, Lihová J. 2006. Polyploidy, hybridization and reticulate evolution: lessons from the
628 Brassicaceae. *Plant Systematics and Evolution* 259: 143-174.
- 629 Martin M. 2011. Cutadapt removes adapter sequences from highthroughput sequencing reads. *EMBnet*
630 *J* 17: 10–12.
- 631 Mayr E. 1992. A local flora and the biological species concept. *American Journal of Botany* 79: 222-
632 238.
- 633 Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean
634 Basin. *Journal of Biogeography* 36: 1333-1345.
- 635 Meyer M, Kircher M, Gansauge MT, et al. 2012. A high-coverage genome sequence from an archaic
636 Denisovan individual. *Science* 338: 222-226.
- 637 Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL:
638 genome-scale coalescent-based species tree estimation. *Bioinformatics* 30: i541-i548.

- 639 Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T. 2006. Generative Cell Specific 1 is essential for
640 angiosperm fertilization. *Nature Cell Biology* 8: 64-71.
- 641 Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden
642 Markov model approach for detecting autozygosity from next-generation sequencing data.
643 *Bioinformatics* 32: 1749-1751.
- 644 Nieto-Feliner G. 1992. Los " *Erysimum*" orófilos nevadenses de flor amarilla y purpúreo-violácea: ¿son
645 coespecíficos?. *Anales del Jardín Botánico de Madrid* 50: 272-274.
- 646 Nieto-Feliner G. 1993. *Erysimum* L. In: Flora iberica. Vol. IV. Cruciferae-Monotropaceae 48–76.
- 647 Nieto-Feliner G. 2011. Southern European glacial refugia: a tale of tales. *Taxon* 60: 365-372.
- 648 Muñoz-Pajares AJ. 2013. *Erysimum mediohispanicum* at the evolutionary crossroad: phylogeography,
649 phenotype, and pollinators. Phd Thesis, Universidad de Granada, Spain.
- 650 Muñoz-Pajares AJ, Perfectti F, Loureiro J, et al. 2018. Niche differences may explain the geographic
651 distribution of cytotypes in *Erysimum mediohispanicum*. *Plant Biology* 20: 139-147.
- 652 Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic
653 algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*
654 32: 268-274.
- 655 Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language.
656 *Bioinformatics* 20: 289-290.
- 657 Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Molecular*
658 *Ecology* 25: 2337-2360.
- 659 Payton AC, Naranjo AA, Judd W, Gitzendanner M, Soltis PS, Soltis DE. 2019. Population genetics,
660 speciation, and hybridization in *Dicerandra* (Lamiaceae), a North American Coastal Plain
661 endemic, and implications for conservation. *Conservation Genetics* 20: 531-543.

- 662 Polatschek A. 1978. Die arten der gattung *Erysimum* auf der Iberischen Halbinsel. *Annalen des*
663 *Naturhistorischen Museums in Wien* 325-362.
- 664 Polatschek A. 1986. *Erysimum*. In Strid A. [ed.], Mountain flora of Greece, 1, 239–247. Cambridge
665 University Press, Cambridge, UK.
- 666 Polatschek A. 2014. Revision der gattung *Erysimum* (Cruciferae): Nachträge zu den bearbeitungen der
667 Iberischen. Halbinsel und Makaronesiens. *Annalen des Naturhistorischen Museums in Wien.*
668 *Serie B für Botanik und Zoologie* 87-105.
- 669 Rambaut A, Drummond AJ. 2012. FigTree version 1. 4. 0.
- 670 Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, et al. 2017. Genomic history of the origin
671 and domestication of common bean unveils its closest sister species. *Genome Biology* 18: 60.
- 672 Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things).
673 *Methods in Ecology and Evolution* 3: 217-223.
- 674 Rieseberg LH, Wendel JF. 1993. Introgression and its consequences in plants. *Hybrid Zones and the*
675 *Evolutionary Process* 70: 109.
- 676 Rieseberg LH, Carney SE. 1998. Plant hybridization. *The New Phytologist* 140: 599-624.
- 677 Rieseberg LH, Raymond O, Rosenthal DM, et al. 2003. Major ecological transitions in wild sunflowers
678 facilitated by hybridization. *Science* 301: 1211-1216.
- 679 Rifkin JL, Castillo AS, Liao IT, Rausher MD. 2019. Gene flow, divergent selection and resistance to
680 introgression in two species of morning glories (Ipomoea). *Molecular Ecology* 28: 1709-1729.
- 681 Saari S, Faeth SH. 2012. Hybridization of Neotyphodium endophytes enhances competitive ability of
682 the host grass. *New Phytologist* 195: 231-236.
- 683 Schemske DW. 2000. Understanding the Origin of Species 1. *Evolution* 54: 1069-1073.
- 684 Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592-593.

- 685 Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to
686 phylogenetic inference. *Molecular Biology and Evolution* 16: 1114-1114.
- 687 Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant*
688 *Biology* 60: 561-588.
- 689 Soltis DE, Visger CJ, Soltis PS. 2014. The polyploidy revolution then and now: Stebbins revisited.
690 *American Journal of Botany* 101: 1057-1078.
- 691 Stebbins GL. 1959. The role of hybridization in evolution. *Proceedings of the American Philosophical*
692 *Society* 103: 231-251.
- 693 Stelkens R, Seehausen O. 2009. Genetic distance between species predicts novel trait expression in
694 their hybrids. *Evolution: International Journal of Organic Evolution* 63: 884-897.
- 695 Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QC, Douglas CJ. 2016. Genomic
696 and functional approaches reveal a case of adaptive introgression from *Populus balsamifera*
697 (balsam poplar) in *P. átrichocarpa* (black cottonwood). *Molecular Ecology* 25: 2427-2442.
- 698 Suarez-Gonzalez A. 2017. Adaptive introgression from *Populus balsamifera* (balsam poplar) in *P.*
699 *trichocarpa* (black cottonwood), PhD Thesis, University of British Columbia, Canada.
- 700 Suarez-Gonzalez A, Lexer C, Cronk QC. 2018. Adaptive introgression: a plant perspective. *Biology*
701 *Letters* 14: 20170688.
- 702 Sujii PS, Cozzolino S, Pinheiro F. 2019. Hybridization and geographic distribution shapes the spatial
703 genetic structure of two co-occurring orchid species. *Heredity* 123: 458-469.
- 704 Sušnik S, Weiss S, Odak T, Delling B, Treer T, Snoj A. 2007. Reticulate evolution: ancient
705 introgression of the Adriatic brown trout mtDNA in softmouth trout *Salmo obtusirostris*
706 (Teleostei: Salmonidae). *Biological Journal of the Linnean Society* 90: 139-152.
- 707 Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence
708 of hybridization in nature. *Nature Ecology & Evolution* 3: 170-177.
- 709 Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing
710 reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

- 711 Turner BL. (2006). Taxonomy and nomenclature of the *Erysimum asperum*- *E.capitatum* complex
712 (Brassicaceae). *Phytologia* 88: 279-287.
- 713 UniProt Consortium. 2014. UniProt: a hub for protein information. *Nucleic Acids Research* 43: D204-
714 D212.
- 715 Wang, D., Wang, Z., Kang, X., & Zhang, J. (2019). Genetic analysis of admixture and hybrid patterns
716 of *Populus hopeiensis* and *P. tomentosa*. *Scientific Reports* 9: 1-13.
- 717 Warwick SI, Francis A, Al-Shehba IA. 2006. Brassicaceae: species checklist and database on CD-Rom.
718 *Plant Systematics and Evolution* 259: 249-258.
- 719 Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Systematic*
720 *Biology* 67: 735-740.
- 721 Wu YC, Rasmussen MD, Bansal MS, Kellis M. 2014. Most parsimonious reconciliation in the presence
722 of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome*
723 *Research* 24: 475-486.
- 724 Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and
725 low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics.
726 *Molecular Biology and Evolution* 31: 3081-3092.
- 727 Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC*
728 *Genomics* 16: S10.
- 729 Zieliński P, Dudek K, Arntzen JW, et al. 2019. Differential introgression across newt hybrid zones:
730 Evidence from replicated transects. *Molecular Ecology* 28: 4811-4824.

731

732

733

734

735

736

737 **List of tables**

738 Table 1. Population code, location, and details of sympatry status for all of the populations sampled.

739 Table 2. Genome size estimates and ploidy levels of the *Erysimum* species samples included in this
740 study. Genome sizes were estimated with flow cytometry analyses.

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

Species	Population	Location	Elevation	Geographical coordinates	Flower color	Sympatry with
<i>E. baeticum</i>	Ebb07	Sierra Nevada, Almería, Spain	2128	37°05'46"N, 3°01'01"W	purple	
	Ebb10	Sierra Nevada, Almería, Spain	2140	37°05'32"N, 3°00'40"W	purple	En12
	Ebb12	Sierra Nevada, Almería, Spain	2264	37°05'51"N, 2°58'06"W	purple	
<i>E. bastetanum</i>	Ebt01	Sierra de Baza, Granada, Spain	1990	37°22'52"N, 2°51'49"W	purple	
	Ebt12	Sierra de María, Almería, Spain	1528	37°41'03"N, 2°10'51"W	purple	
	Ebt13	Sierra Jureña, Granada, Spain	1352	37°57'10"N, 2°29'24"W	purple	Em71
<i>E. fitzii</i>	Ef01	Sierra de la Pandera, Jaén, Spain	1804	37°37'56"N, 3°46'46"W	yellow	
<i>E. lagascae</i>	Ela07	Sierra de San Vicente, Toledo, Spain	516	44°05'49"N, 4°40'40"W	purple	
<i>E. mediohispanicum</i>	Em21	Sierra Nevada, Granada, Spain	1723	37°08'04"N, 3°25'43"W	yellow	
	Em39	Sierra de Huétor, Granada, Spain	1272	37°19'08"N, 3°33'11"W	yellow	Ep20
	Em71	Sierra Jureña, Granada, Spain	1352	37°57'10"N, 2°29'24"W	yellow	Ebt13
<i>E. nevadense</i>	En05	Sierra Nevada, Granada, Spain	2074	37°06'35"N, 3°01'32"W	yellow	
	En10	Sierra Nevada, Granada, Spain	2321	37°06'37"N, 3°24'18"W	yellow	
	En12	Sierra Nevada, Granada, Spain	2255	37°05'37"N, 2°56'19"W	yellow	Ebb10
<i>E. popovii</i>	Ep16	Jabalruz, Jaén, Spain	796	37°45'26"N, 3°51'02"W	purple	
	Ep20	Sierra de Huétor, Granada, Spain	1272	37°19'08"N, 3°33'11"W	purple	Em39
	Ep27	Llanos del Purche, Granada, Spain	1470	37°07'46"N, 3°28'48"W	purple	

Table 1. Population code, location, and details of sympatry status for all of the populations sampled.

761

Species	Population	DNA Ploidy level			Genome size (2C, pg)				
		2n	N	Mean	SD	CV	Min	Max	N
<i>E. baeticum</i>	Ebb07	8x	5	2.08	0.08	3.85	1.93	2.17	2
	Ebb10	8x	6	2.07	0.09	4.35	1.93	2.17	5
	Ebb12	8x	-	-	-	-	-	-	-
<i>E. bastetanum</i>	Ebt01	4x	4	1.06	0.06	5.66	0.97	1.10	4
	Ebt12	4x	2	1.06	0.12	11.32	0.97	1.15	2
	Ebt13	8x	64	1.96	0.06	3.06	1.87	2.17	60
<i>E. fitzii</i>	Ef01	2x	3	0.44	0.004	0.91	0.44	0.45	3
<i>E. lagascae</i>	Ela07	2x	10	0.46	0.02	4.35	0.44	0.50	10
<i>E. mediohispanicum</i>	Em21	2x	2	0.44	0.01	2.27	0.43	0.44	2
	Em39	2x	21	0.46	0.02	4.35	0.43	0.49	19
	Em71	4x	59	0.98	0.04	4.08	0.93	1.13	59
<i>E. nevadense</i>	En05	2x	-	-	-	-	-	-	-
	En10	2x	-	-	-	-	-	-	-
	En12	2x	3	0.45	0.03	6.67	0.42	0.47	3
<i>E. popovii</i>	Ep16	4x	3	0.98	0.02	2.041.86	0.95	1.00	3
	Ep20	10x	15	2.49	0.06	2.416	2.40	2.60	10
	Ep27	4x	39	0.96	0.04	4.17	0.92	1.05	9

762

763 Table 2. Genome size estimates and DNA ploidy levels obtained in populations of *Erysimum*. The
764 following data are given for each population and ploidy level: mean, the standard deviation of the mean
765 (SD), coefficient of variation (CV, %), minimum (Min) and maximum values (Max) of the holoploid
766 genome size (2C, pg) followed by sample size for genome size estimates (N); DNA ploidy level (2n)
767 and respective sample size (N) for ploidy estimates. DNA ploidy levels: 2x, diploid; 4x, tetraploid; 8x,
768 octoploid; 10x, decaploid. For Ebb12, En05, and En10 samples were not possible to estimate the ploidy
769 levels, and we have used the described in Blanca et al. (1992).

770

771

772

773 **List of figures**

774 Figure 1. Map of the Iberian Peninsula showing the location of the sampled populations.

775 Figure 2. Cyto-nuclear discordance in *Erysimum* spp.. The phylogeny on the left was obtained using
776 whole plastid genomes in Osuna-Mascaró et al. (2021). The phylogeny on the right is a representation
777 of nuclear genome evolution and was generated from the 16,941 maximum likelihood gene trees
778 computed using the SNP data described in the present paper (see text for details).

779 Figure 3. Membership probability plot showing the DAPC results representing the populations grouped
780 into predetermined different clusters.

781 Figure 4. Optimal species network. The graph represents a maximum pseudo-likelihood (MPL) tree
782 with 13 reticulations computed using PhyloNet. These events are represented by edges connecting the
783 tree branches between different individuals and indicate likely hybridization between different taxa.

784 Note that in some instances, introgression appears to involved ancestral or extinct taxa (i.e., ghost
785 species, dotted lines).

786

787

788

789

790

791

792

793

794

795

796

797

798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821

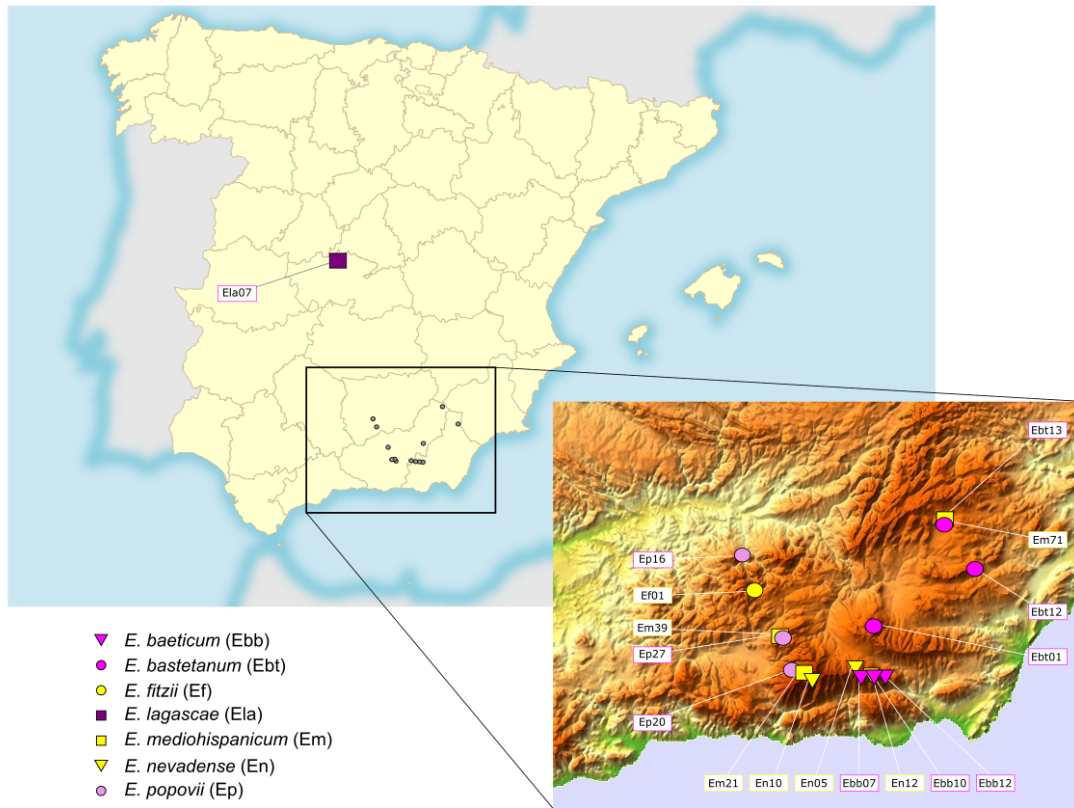
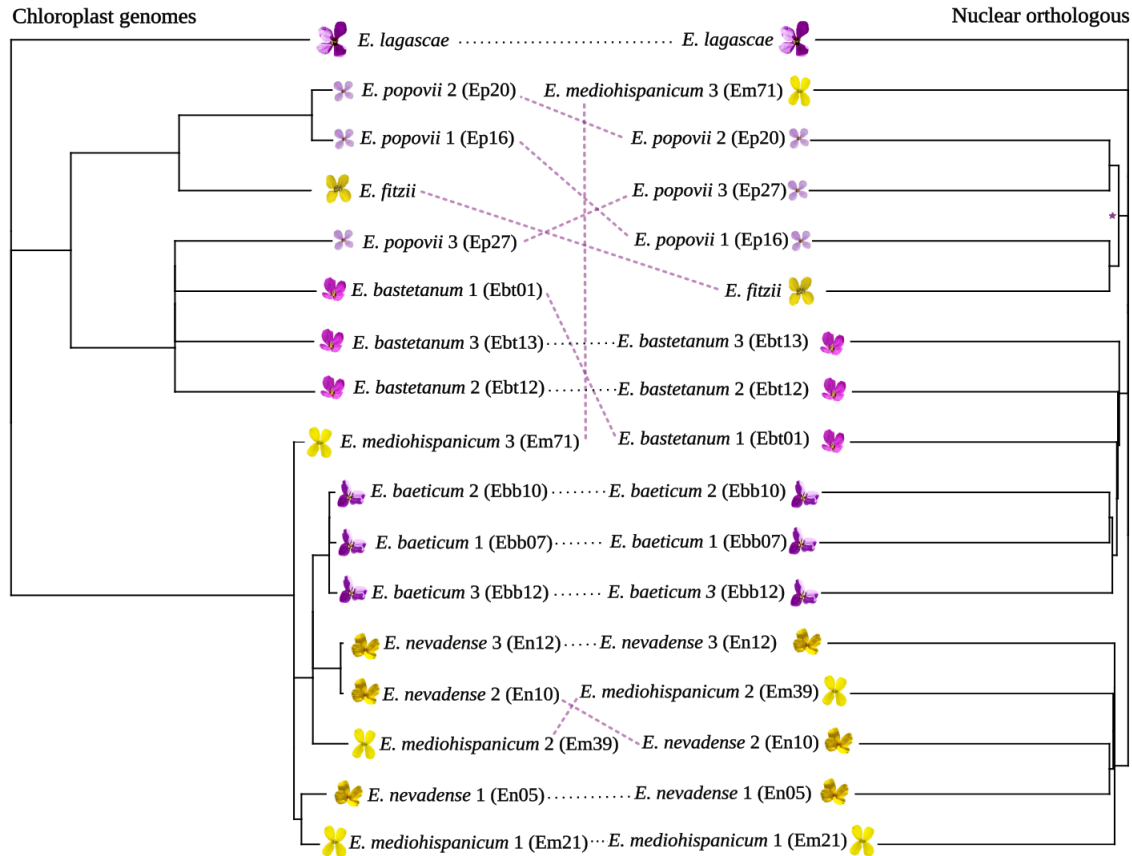
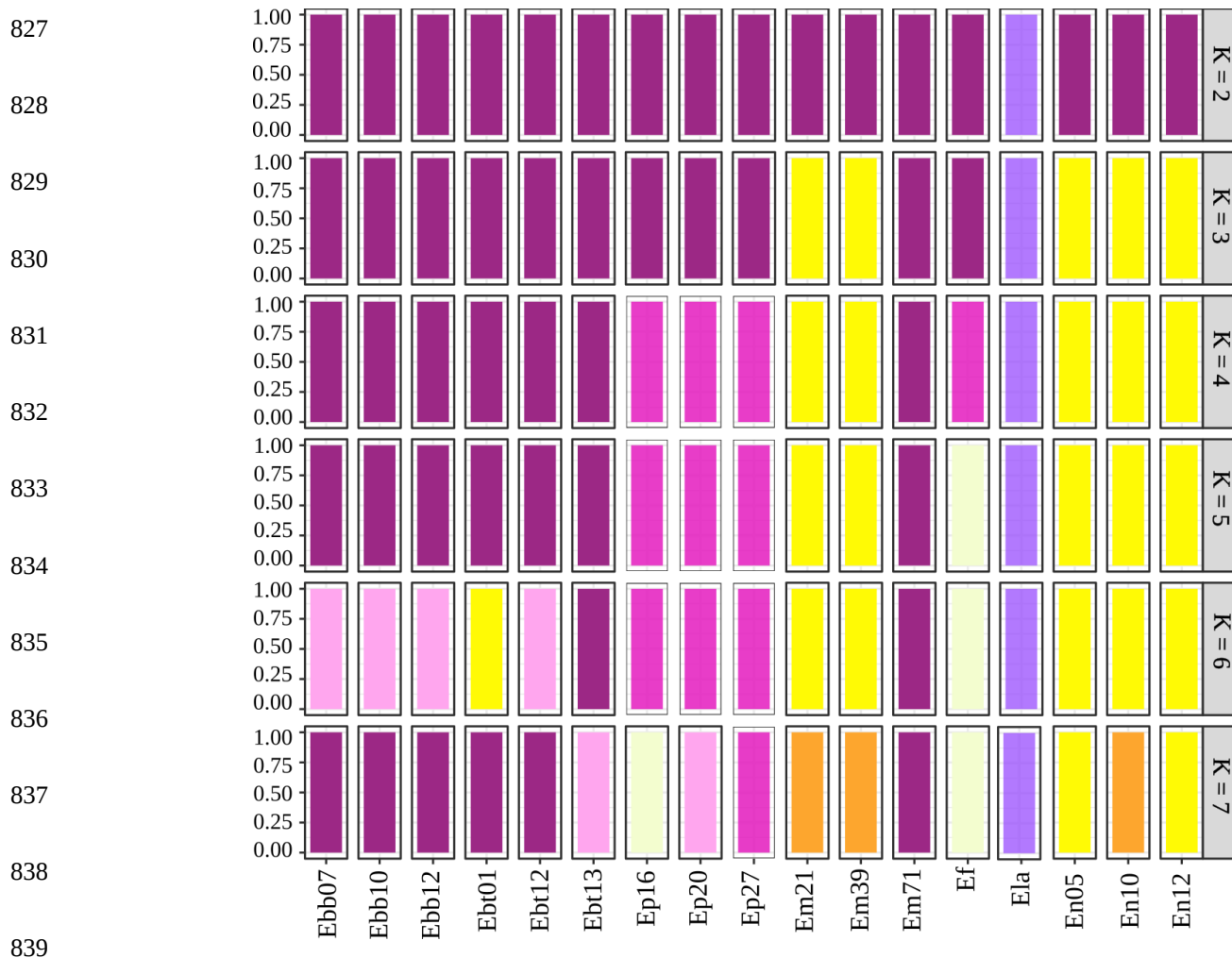


Figure 1. Map of the Iberian Peninsula showing the location of the sampled populations. The insert shows a more detailed map of the Baetic mountains. Purple species are represented with purple symbols and yellow species are represented with yellow symbols. The populations Ebt13 – Em71, Ebb10 – En12, and Em39 – Ep27 represent population pairs located in sympatry.



822 Figure 2. Cyto-nuclear discordance in *Erysimum* spp.. The phylogeny on the left was obtained using
823 whole plastid genomes in Osuna-Mascaró et al. (2021). The phylogeny on the right is a representation
824 of nuclear genome evolution and was generated from the 16,941 maximum likelihood gene trees
825 computed using the SNP data described in the present paper (see text for details).

826

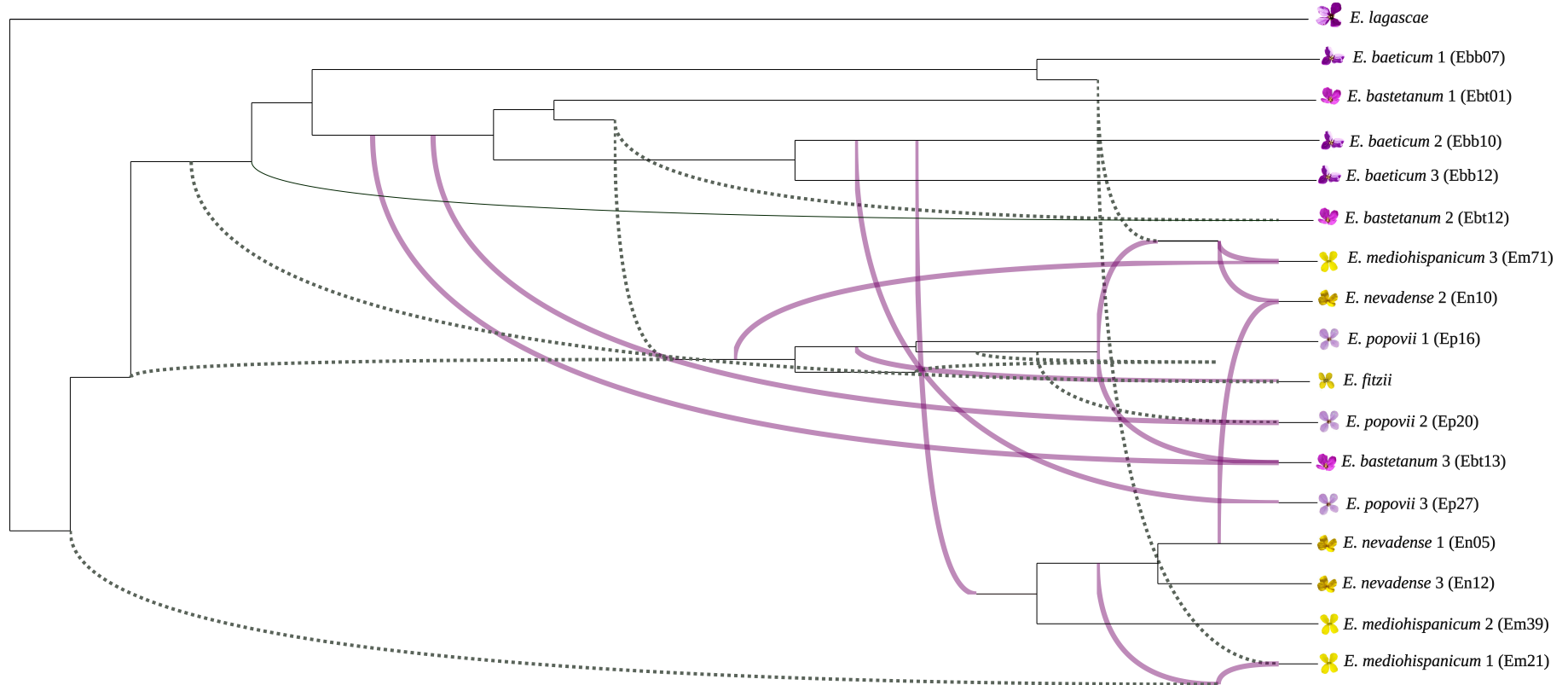


840 Figure 3. Membership probability plot showing the DAPC results representing the populations grouped
841 into predetermined different clusters (ranging from K=2 to K=7, where each color represents a cluster).

842 The Bayesian analyses (BIC) revealed K = 5 as the most likely number of genetic clusters.

843

844



848 Figure 4. Optimal species network. The graph represents a maximum pseudo-likelihood (MPL) tree with 13 reticulations computed using
 849 PhyloNet. These events are represented by edges connecting the tree branches between different individuals and indicate likely hybridization
 850 between different taxa. Note that in some instances, introgression appears to involved ancestral or extinct taxa (i.e., ghost species, dotted
 851 lines).