



HAL
open science

Holocene climate change promoted allopatric divergence and disjunct geographic distribution in a bee orchid species

Anaïs Gibert, Roselyne Buscail, Michel Baguette, Christelle Fraïsse, Camille Roux, Bertrand Schatz, Joris A. M. Bertrand

► To cite this version:

Anaïs Gibert, Roselyne Buscail, Michel Baguette, Christelle Fraïsse, Camille Roux, et al.. Holocene climate change promoted allopatric divergence and disjunct geographic distribution in a bee orchid species. *Journal of Biogeography*, 2024, 51 (12), pp.2424-2439. <10.1111/jbi.14998>. <hal-04687630>

HAL Id: hal-04687630

<https://hal.science/hal-04687630v1>

Submitted on 4 Sep 2024

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

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



HAL Authorization

1 Holocene climate change promoted allopatric divergence and 2 disjunct geographic distribution in a bee orchid species

3

4 **Running title** Biogeography of *Ophrys aveyronensis*

5

6 Anaïs Gibert¹, Roselyne Buscail², Michel Baguette^{3,4}, Christelle Fraïsse⁵, Camille
7 Roux⁵, Bertrand Schatz⁶ & Joris A. M. Bertrand¹

8

9 ¹ Laboratoire Génome et Développement des Plantes (LGDP), UMR 5096, Université de
10 Perpignan Via Domitia (UPVD) - Centre National de la Recherche Scientifique (CNRS) - Institut
11 de Recherche pour le Développement (IRD), EMR 269 MANGO, F-66860 Perpignan.

12 ² Centre de Formation et de Recherche sur les Environnements Méditerranéens (CEFREM),
13 Université de Perpignan Via Domitia (UPVD) - Centre National de la Recherche Scientifique
14 (CNRS), F-66860 Perpignan.

15 ³ Institut Systématique, Évolution, Biodiversité (ISEB), UMR 7205, Museum National d'Histoire
16 Naturelle (MNHN) - Centre National de la Recherche Scientifique (CNRS) - Sorbonne Université
17 - École Pratique des Hautes Études (EPHE) - Université des Antilles, F-75005 Paris.

18 ⁴ Station d'Écologie Théorique et Expérimentale (SETE), UMR 5321, Centre National de la
19 Recherche Scientifique (CNRS) - Université Toulouse III, F-09200 Moulis.

20 ⁵ CNRS, Univ. Lille, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille.

21 ⁶ Centre d'Écologie Fonctionnelle et Évolutive (CEFE) - Université de Montpellier - Centre
22 National de la Recherche Scientifique (CNRS) - École Pratique des Hautes Études (EPHE) -
23 Institut de Recherche pour le Développement (IRD), F-34293 Montpellier.

24

25 Correspondence: anais.gibert@gmail.com

26

27 **Abstract**

28 **Aim**

29 Species with disjunct geographic distributions provide natural opportunities to investigate
30 incipient or recent allopatric divergence. The combination of both genetic and ecological data
31 may be fruitful to decipher the causes of such pattern: i) actual vicariance, ii) successful
32 colonization from one source range to a new range (dispersal, biological introduction) or iii)
33 parallel convergent evolution.

34 **Location**

35 Southern France and Northern Spain.

36 **Taxon**

37 The bee orchid *Ophrys aveyronensis* (and its two recognized subspecies *O. a.* subsp.
38 *aveyronensis* and *O. a.* subsp. *vitorica*) displays a disjunct geographic distribution with two
39 subranges separated by 600 km on both sides of the Pyrenees mountain range.

40 **Methods**

41 As allopatric divergence is often complex to document in the wild, we used a combination of
42 population genomics and Ecological Niche Modelling (ENM) to investigate the causes of this
43 intriguing biogeographic pattern.

44 **Results**

45 The population genomic data demonstrate that all the studied populations exhibit similar
46 patterns of genetic diversity and dramatic decrease in effective size compared to the ancestral

47 population. Significant genetic differentiation and reciprocal monophyly exist between
48 populations of the two subranges of *O. aveyronensis*, despite a very recent divergence time
49 as young as ca. 1500 generations ago. Moreover, (paleo-)ecological Niche Modelling analyses
50 support that the disjunct geographic distribution of *O. aveyronensis* is consistent with a range
51 split of a broad ancestral range, contraction and distinct longitudinal and latitudinal shifts in
52 response to climate warming during the Holocene.

53 **Main Conclusion**

54 The congruence of the results obtained from both population genomics and ENM approaches
55 documents how very recent continental allopatric divergence initiated speciation in this
56 system. *Ophrys aveyronensis* provides a promising opportunity to study the onset of
57 reproductive isolation and parallel evolution following an initial stage of geographic
58 separation in a group with high diversification rate.

59

60 **Keywords** Biogeography, Ecological Niche Modeling (ENM), Evolutionary divergence,
61 Disjunct geographic distribution, *Ophrys aveyronensis*, Orchidaceae, Population genomics,
62 RAD sequencing, Species Distribution Modeling (SDM), Vicariance

63 Introduction

64 Allopatric speciation (also referred as geographic speciation, or vicariant speciation) occurs
65 when populations of a given ancestral species become reproductively isolated from each
66 other because of geographic separation. Aside from being conceptually intuitive, allopatric
67 speciation has been evidenced by theoretical models, laboratory experiments and empirical
68 studies to the point that it is generally regarded as the “default mode” of speciation (see Coyne
69 & Orr, 2004). In allopatric speciation, spatial and reproductive isolation are often assumed to
70 originate from geologic-caused topographic changes (e.g. mountain formation) and/or
71 climate-induced changes (e.g. sea level modifications in continental island systems) that form
72 the geographic barrier. However, not only the fore-mentioned barriers but any kind of
73 fragmentation of species distribution range, including those caused by human activities, may
74 also initiate and be responsible for evolutionary divergence.

75 The putative complexity of the speciation process makes it difficult to elucidate its
76 ecological and evolutionary causes, first because current spatial patterns of distribution only
77 provide a snapshot from the speciation continuum (Nosil, Feder, Flaxman & Gompert, 2017;
78 Stankowski & Ravinet, 2021). Allopatric divergence can be a spatio-temporally dynamic
79 process as geographic and reproductive isolation is prone to be only transient. For example,
80 climate change-induced oscillations during the Quaternary (Hewitt, 2000) have been shown
81 to be an important driver of plant speciation and radiations, particularly in mountainous and
82 arid environments (Kadereit & Abbott, 2021). In this context, populations may experience at
83 least one phase of isolation and secondary contact which may alter our ability to properly infer
84 the chronology of the divergence (see Ravinet et al., 2017) and separate the ‘historical-
85 geographical’ from ‘ecogeographic’ causes of reproductive isolation (sensu Sobel, Chen, Watt
86 & Schemske, 2010, see also Sobel, 2014).

87 In this regard, taxa with disjunct geographic distribution may provide particularly
88 insightful opportunities to study populations that became recently allopatric, *i.e.* with
89 insufficient time to accumulate enough divergence to be considered as distinct species yet, as
90 promising candidates to investigate incipient parallel evolution (see James et al., 2023).
91 However, reviews of empirical cases of speciation tend to point out the relative scarcity of
92 such candidates having started to diverge as recently as during the Holocene, especially in
93 plants (see Kadereit & Abbott, 2021; de Queiroz et al., 2022). Aside, although such natural
94 examples may help to understand the consequences of physical barriers in early stages of
95 evolutionary divergence, disjunct geographic distributions do not necessary result from the
96 fragmentation of an ancestral range. Indeed, dispersal and successful colonization of a new
97 range from the historical one may also lead to a similar biogeographic pattern.

98 Species of the genus *Ophrys* have long been known for their unusual pollination
99 syndrome (called ‘sexual swindling’) that involves insect-like flowers mimicking their female
100 pollinator to lure conspecific males and ensure pollination but are now also increasingly
101 considered as a promising biological model to investigate adaptive radiations. The *Ophrys*
102 flower, and in particular one petal, the labellum but also additional floral traits (e.g. Lussu, De
103 Agostini, Cogoni, Marigani & Cortis, 2019), have evolved dimensions that match the features
104 of their insect pollinator. The degree of plant-pollinator specialization (Joffard, Massol, Grenié,
105 Montgellard & Schatz, 2019) required to ensure the success of the *Ophrys* pollination strategy
106 was shown to be the driver of the impressive rate of diversification of the genus within the
107 Mediterranean basin, and to a lesser extent in other parts of Western Europe (see Baguette,
108 Bertrand, Stevens & Schatz, 2020, for a recent review).

109 On the one hand, the explosive diversification that *Ophrys* spp. experience makes it a
110 promising system to investigate speciation as an evolutionary process. On the other hand,
111 such rates of diversification question the use of traditional species concepts. Most *Ophrys* taxa
112 display phenotypic differences and attract distinct pollinator species, which induce pre-zygotic
113 barriers, but are still able to produce fertile hybrids in the wild (*i.e.* show only incomplete post-
114 zygotic barriers) and thus do not conform to the biological species concept (e.g. *sensu* Mayr,
115 1942). In addition, most phylogenetic and population genetic approaches have failed to assign
116 closely related *Ophrys* taxa to reciprocally monophyletic groups, as required to conform to the
117 phylogenetic concept of species. The combination of typological criteria with multi-omic
118 approaches however offers the promising opportunity to better understand the onset of
119 reproductive isolation, even when still incomplete.

120 If *Ophrys* species are notorious for their relatively high number of likely cases of
121 speciation with gene flow, the number of island endemic taxa suggests that geographic
122 isolation (induced by insularity) is also a key factor to form the large number of recent and
123 emerging *Ophrys* species observable nowadays (see Bertrand, Baguette, Joffard & Schatz et
124 al., 2021). On continents, assessing the role of allopatry in *Ophrys* diversification is more
125 challenging, in particular because clear evidences of taxa displaying subranges without being
126 connected by gene flow are lacking. Yet, systems allowing to study and compare *Ophrys* cases
127 of allopatric divergence with those of speciation with gene flow would be insightful to
128 understand the evolution of such diverse genera. One species, *Ophrys aveyronensis* (J.J.
129 Wood) P. Delforge 1984, initially described as *Ophrys sphegodes* subsp. *aveyronensis* J.J.
130 Wood, 1983, displays an intriguing case of disjunct geographic distribution. This taxon was
131 initially described as endemic to a very restricted area in the Grands Causses region (Southern
132 karstic areas of the Massif Central, France). Observations of morphologically very similar
133 populations were reported later from mid-elevations regions in Northern Spain (La Rioja,
134 Burgos, Alava and Vizcaya provinces) and described as a distinct species (*Ophrys vitorica*,
135 Kreutz, 2007) on the basis of their geographic isolation as well as because of subtle floral
136 characteristics (e.g. smaller flowers with narrower labella which display simpler H-shaped
137 ornamentation). Ever since, the significance of the differences at some floral morphometry
138 traits (e.g. sepal and labellum widths) and coloration has been confirmed (see Gibert et al.,
139 2022). However, *Ophrys aveyronensis* is rather considered as a unique species comprising two
140 subspecies (*O. aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica*) by
141 other authors because the populations share, at least a mutual insect pollinator species (*i.e.*
142 the solitary bee *Andrena hattorfiana*, Paulus & Gack, 1999; Claessens & Kleynen, 2016; Paulus,
143 2017; Benito Ayuso, 2019) across their whole range. *Andrena hattorfiana* is a widespread
144 species occurring across the whole western Palearctic where it quasi-exclusively forages on field
145 scabious (*Knautia arvensis*). Yet, the populations of *Ophrys aveyronensis* occur in two rather
146 restricted ranges separated by a gap of about 600 km (and Pyrenees Mountains), in the Grands
147 Causses region (France) and in the central part of Northern Spain (Fig. 1). Altogether, current
148 data are thus consistent with the existence of a unique species formed by two geographically
149 isolated subspecies (as proposed by Paulus, 2017) with no further evidence of pre-zygotic
150 reproductive barriers.

151 The formation of the Pyrenees during the Eocene (from 45 Myrs ago) cannot explain
152 by itself the split of the ancestral range of the species as all phylogenetic hypotheses published
153 so far conclude that the *Ophrys* genus emerged no more than 5-6 Myrs ago and that the highly
154 diverse *Sphegodes* clade, to which *O. aveyronensis* belongs may have diversified in the last
155 million years (Breitkopf et al., 2015, see also Inda, Pimentel & Chase et al., 2012). This does

156 not rule out allopatry as factor to explain current disjunct geographic distribution but suggests
157 that geographic separation would have occurred recently. According to this ‘vicariance
158 hypothesis’, the current disjunct geographic distribution may correspond to two relict pockets
159 of an ancestral range that got fragmented following a scenario according to which populations
160 from Northern and Southern ranges diverge in allopatry. In this case, the degree of genetic
161 differentiation between Northern and Southern populations is expected to be significant.
162 Namely, genetic differentiation is expected to be relatively higher between Northern and
163 Southern ranges than between populations within subranges. Alternatively, we also envisage
164 a ‘biological introduction hypothesis’: a colonization from one historical range to the other.
165 The notoriously small *Ophrys* seeds are likely to be wind-dispersed over hundreds of
166 kilometers (*i.e.* thus being able to travel across most physical barriers including mountains and
167 vast open water areas) and may have found in the newly colonized area, suitable conditions
168 for their germination, development and reproduction. Although ‘orchid dusty seed rains’ may
169 allow to generate founder populations with non-negligible genetic diversity, such newly
170 established populations typically consist of a subset of the genetic pool of the source
171 population (*i.e.* a relatively little divergent population with lower effective population size and
172 genetic diversity). This could lead to a phylogenetic pattern where the populations of the
173 founded taxon are nested within those of the source taxon, rendering the later paraphyletic
174 with respect to the former.

175 In this study, we used a population genomic approach (*i.e.* through a double digest
176 RAD-seq like protocol called ‘nGBS’) to first investigate patterns of genetic diversity and
177 differentiation among *Ophrys aveyronensis* populations from both part of its disjunct
178 geographic range. We then couple demographic inference analyses and Ecological Niche
179 Modelling approaches to test whether the current disjunct geographic distribution of this
180 species is rather consistent with either vicariance or recent dispersal. Specifically, we aimed
181 at determining how recent the divergence between Northern Spain and Southern France
182 populations of *O. aveyronensis* may be and whether or not this disjunct geographic
183 distribution may have originated with a colonization event from an ancestral range to a new
184 one.

185

186 **Materials and Methods**

187 **Sampling and study sites**

188 We used a minimally destructive sampling method to collect a total of 86 plant tissue samples
189 from six localities that span the entire geographic range of *O. aveyronensis*: three in Southern
190 France (Grands Causses region): 1-Guilhaumard, 2-Lapanouse-de-Cernon and 3-Saint-
191 Affrique, and three in Northern Spain: 4-Valgañón, 5-Larraona and 6-Bercedo, in June 2019,
192 Fig. 1). All the specimens were photographed and phenotyped at a suite of morphological
193 traits (see Gibert et al., 2022). At each sampling site, we also collected 6 to 10, 0.6 mm
194 diameter leaf punches from 13 to 15 individuals which were stored in 100% ethanol until DNA
195 extraction (see Table 1 and Supplementary Table S1). Even though this species is known to be
196 strictly allogamous and producing seeds able to disperse over long distances, we sampled
197 specimens that where at least 1 m apart from each other. As *Ophrys aveyronensis* is a globally
198 rare and nationally protected plant species in France, sampling was done under permit ‘Arrêté
199 préfectoral n°2019-s-16’ issued by the ‘Direction Régionale de l’Environnement de
200 l’Aménagement et du Logement (DREAL)’ from the ‘Région Occitanie’, on 07-May-2019).
201 *Ophrys aveyronensis* is not legally protected in Spain.

202

203 **DNA extraction, library preparation and genotyping**

204 Genomic DNA extraction as well as library preparation and genotyping were subcontracted to
205 LGC Genomics GmbH (Berlin, Germany). A fractional genome sequencing strategy called
206 normalized Genotyping-by-Sequencing (nGBS) was used to subsample the *Ophrys*
207 *aveyronensis* genome. Most *Ophrys* species are diploid: $2n = 36$ (Turco, Albano, Medagli,
208 Wagensommer & D'Emérico, 2023) and *Ophrys aveyronensis*, for which we estimated a
209 relatively large size of ~5-6 Gbp (1C-value = 5.44 to 6.02 picograms, $n = 3$ for *O. a.* subsp.
210 *aveyronensis* and 1C-value = 5.88 pg, $n = 1$, for *O. a.* subsp. *vtorica*) based on flow cytometry
211 techniques (Bertrand, unpublished data) conforms to previous knowledge. The nGBS protocol
212 makes use of a combination of two restriction enzymes (PstI and ApeKI, in our case) to produce
213 a reproducible set of fragments across samples/individuals. It also comprises a normalization
214 step whose aim is to avoid further sequencing of highly repetitive regions. Illumina technology
215 was then used to aim at obtaining a minimum of 1.5 million reads per sample (with 2 x 150 bp
216 read length) for each of the 86 individually barcoded libraries.

217

218 **Population genomic dataset processing**

219 We used different scripts included in *Stacks* v.2.60 (Catchen Amores, Creskp & Postlewait,
220 2011; Catchen, Hohenlohe, Bassham, Amores & Cresko, 2013) to build loci from Illumina
221 reads, *de novo* (*i.e.* without aligning reads to a reference genome). We first used
222 *process_radtags* to demultiplex and clean reads, then *denovo_map.pl* to build loci within
223 individuals, create a catalog and match all samples against it and finally *populations* to further
224 filter the SNPs obtained at the population level and compute basic population genetic
225 statistics. Before running the pipeline on the complete dataset, we first optimized several key
226 parameters: *-m* (the minimum number of identical raw reads required to form a putative
227 allele), *-M* (the number of mismatches allowed between alleles to form a locus) and *-n* (the
228 number of mismatches allowed between loci during construction of the catalog) by running it
229 on a subset of 12 individuals representative of the whole dataset (*i.e.* geographic origin and
230 coverage). As recommended by several authors, we varied *M* and *n* from 1 to 9 (fixing $M = n$)
231 while keeping $m = 3$ (see Paris, Stevens, & Catchen, 2017). The combination *-m* 3, *-M* 5 and *-n*
232 5 was found to be the most suitable to maximize the number of SNPs, assembled and
233 polymorphic loci in our case and was used to run *Stacks* on the whole data set (see
234 Supplementary Table S2). Then, we ran the *populations* script, from which only SNPs that were
235 genotyped in at least 80% of the individuals per population (*-r* 80) and found in all six sampling
236 sites were kept (*-p* 6), while SNPs exhibiting estimates of observed heterozygosity greater than
237 70% (*--max-obs-het* 0.7) were filtered out to reduce the risk of including remaining paralogs.
238 We also discarded sites whose minor allele frequency was lower than 5% (*--min-maf* 0.05).
239 The genotype matrix of the remaining 9301 SNPs was then exported in several formats for
240 downstream analyses.

241 The following population genomic analyses aimed to i) infer patterns of genetic
242 diversity and divergence between subspecies/countries and sampling sites and ii) infer the
243 demographic scenario (and associated parameters) likely to best explain the current
244 biogeographic pattern observed in the *O. aveyronensis*.

245

246 **Genetic diversity, population differentiation and evolutionary divergence**

247 To get an overview of the overall genetic diversity and differentiation among individuals and
248 populations, we first performed a Principal Component Analysis (PCA) based on a matrix of 86
249 individuals (as rows) and 9301 SNPs (as columns) coded as a *genlight* object with the R-

250 package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011). The number of private alleles
251 and the average nucleotide diversity (π) were directly available from the output of the
252 *populations* script from *Stacks*. We used Genodive v.3.04 (Meirmans, 2020) to compute
253 expected and observed heterozygosity (H_E and H_O , respectively) as well as to evaluate the
254 deviation from panmixia by computing G_{IS} . Allelic richness per sampling site (A_R) was
255 computed with the R-package *hierfstat* (Weir & Goudet, 2017). Overall and pairwise genetic
256 differentiation was also assessed based on G -statistics (G_{ST} , G'_{ST} , G''_{ST}) as well as Jost's D also
257 implemented in Genodive. Genodive was also used to carry out a hierarchical Analysis of
258 Molecular Variance (AMOVA) to assess proportions of genetic variance (i) among groups
259 (here, subspecies/countries) (F_{CT}) and (ii) among populations within each subspecies (F_{SC}). The
260 statistical significance of the obtained values was estimated based on 10000 permutations.

261 To further investigate population structure and characterize putative
262 migration/admixture event, we used sNMF (Frichot, Mathieu, Trouillon, Bouchard & François,
263 2014) as implemented in the R-package *LEA* (Frichot & François, 2015) to estimate individual
264 ancestry coefficients based on sparse non-negative matrix factorization algorithms. The
265 number of ancestral populations (genetic clusters) ranged from $K = 1$ to 10, and analyses were
266 run with 10 replicates at each value of K . We followed the cross-entropy criterion to determine
267 the optimal number of clusters and the ancestry proportions matrices (Q -matrices) obtained
268 were plotted with the R-package *pophelper* (Francis, 2017).

269 Finally, we performed Maximum-Likelihood phylogenetic tree reconstruction with IQ-
270 Tree v.2.2.6 (Minh *et al.*, 2020) based on the concatenated SNP matrix with 1000 replicates (-
271 B 1000) of Ultrafast Bootstrap Approximation (UFBoot) to assess nodes support. The
272 substitution model that best fitted the data was selected with *ModelFinder Plus*
273 (Kalyaanamoorthy, Minh, Wong, Von Haeseler & Jermini, 2017) and we used Ascertainment
274 bias correction (-MFP+ASC) as recommended for SNP data.

275

276 **Demographic and evolutionary history reconstruction**

277 We reconstructed the past demographic history of each of the six populations, separately,
278 using STAIRWAYPLOT2 (Liu & Fu, 2015; 2020). This program relies on the Site Frequency
279 Spectrum (SFS) to infer the temporal dynamics of effective population size (N_e) and has the
280 advantage of not requiring whole-genome sequence data or reference genome information.
281 First, we used the *easySFS.py* script (<https://github.com/isacovercast/easySFS>) to compute
282 the folded SFS of each population. To maximize the number of segregating sites, the
283 populations were downsampled to 24, 24, 22, 24, 24 and 22 individuals for 1-Guilhaumard, 2-
284 Lapanouse, 3-St-Affrique, 4-Valgañón, 5-Larraona and 6-Bercedo, respectively. We then ran
285 STAIRWAYPLOT2 that fits a flexible multi-epoch demographic model that estimates N_e at
286 different time periods using the SFS to infer estimated N_e fluctuation coinciding with
287 coalescent events. Here, we assumed a mutation rate of 7×10^{-9} per site per generation for
288 angiosperms autosomal markers (following Krasovec, Chester, Ridout & Filatov, 2018) a mean
289 generation time of 5 years for *Ophrys*, and performed 200 bootstrap replicates to estimate
290 95% confidence intervals.

291 We then used the Approximate Bayesian Computation (ABC) framework implemented
292 in DILS (Fraïsse *et al.*, 2021) to infer the best demo-genomic scenario and associated
293 parameters of the evolutionary history between *O. a. subsp. aveyronensis* and *O. a. subsp.*
294 *vitorica*. We considered two types of datasets i) a pooled dataset grouping the three
295 populations *O. a. subsp. aveyronensis* together and the 3 populations of *O. a. subsp. vitorica*
296 together and ii) paired datasets consisting of all possible *O. a. subsp. aveyronensis* and *O. a.*

297 subsp. *vitorica* population pairs ($3^2 = 9$ pairwise comparisons). The two-population model of
298 DILS compares different scenarios in a hierarchical manner. First, current isolation models, *i.e.*
299 Strict Isolation (SI) versus Ancient Migration (AM) are compared to ongoing migration models,
300 *i.e.* Isolation-with-Migration (IM) versus Secondary Contact (SC). Then, for the best
301 demographic model, DILS compares genomic models to test for linked selection by assuming
302 variation in effective population size (N_e) across sites (*i.e.* hetero versus homo N_e models are
303 compared). For migration models (IM and SC), DILS also tests for selection against migrants by
304 assuming variation in migration rates $N.m$ across sites (*i.e.* hetero versus homo $N.m$ models
305 are compared). Depending on the best selected model, DILS provides estimates and
306 confidence intervals for parameters such as the time of split (T_{split}), the migration rate ($N.m$),
307 current and ancestral effective population sizes ($N_{e_{current}}$ and $N_{e_{past}}$) and whenever relevant,
308 the time of ancient migration (T_{AM}) or the time of secondary contact (T_{SC}). We formatted the
309 input file with a custom Perl script and used the online facility (<https://www.france-bioinformatique.fr/cluster-ifb-core/>) to run computations with the following settings and
310 priors: *population_growth*: constant, *modeBarrier*: bimodal (*i.e.* there is a class of species
311 barrier loci with $N.m = 0$ and a class of loci migrating at the background rate $N.m$),
312 *max_N_tolerated*: 0.1, *LMin* = 30, *nMin* = 10 (20 for the pooled dataset), $\mu = 7.10^{-9}$,
313 *rho_over_theta*: 0.1, *N_min* = 100, *N_max*: 200000, *Tsplit_min*: 100, *Tsplit_max*: 25000,
314 *M_min*: 0.4 and *M_max*: 4. To quantify the fit of the best model to the data we conducted a
315 goodness-of-fit test using 2000 simulations performed under the best model based on the
316 estimated parameter values.
317

318

319 **Past and current Species Distribution Modelling**

320 We used an Ecological Niche Modelling (or Environmental Niche Modelling, ENM or Species
321 Distribution Modelling, SDM) framework similar to the one we followed and described in
322 detail in Salvado et al. (2022) to infer the spatio-temporal dynamics of suitable habitat for *O.*
323 *aveyronensis* since the Last Glacial Maximum (LGM, 21 kyrs BP). We then compared the results
324 with those obtained from population genomic analyses. To model and spatially project the
325 current bioclimatic niche, we first downloaded 19 bioclimatic layers at a resolution of 2.5 arc-
326 minutes ($\sim 5 \text{ km}^2$) available from WorldClim 2 (Fick & Hijmans, 2017). Current climate data
327 correspond to time averaged variables over the period 1970-2000. To spatially project the
328 inferred niche onto past bioclimatic conditions data, we then downloaded paleoclimate data
329 either from WorldClim v1.4 (Hijmans, Cmaeron, Para, Jones & Jarvis, 2005) or from PaleoClim
330 (www.paleoclim.org).

331 Paleoclimate data from WorldClim are available for both the LGM (21 kyrs ago) and
332 the Mid-Holocene (about 6 kyrs ago), downscaled at a resolution of 2.5 arc-minutes for at least
333 three commonly used General Circulation Models (GCMs) proposed by the National Center
334 for Atmospheric Research (NCAR) though the Community Climate System Model (CCSM4), the
335 Japan Agency for Marine-Earth Science and Technology, the Center for Climate System
336 Research of the University of Tokyo, and the National Institute for Environmental Studies
337 (MIROC-ESM) and the Max Plant Institute for Meteorology Earth System Model (MPI-ESM-P).
338 Only CCSM data are available from PaleoClim but over different time periods over the
339 Pleistocene (at a resolution of 2.5 arc-minutes): the LGM (as from CHELSA, Karger, Nobis,
340 Normand, Graham & Zimmermann, 2021), the Heinrich Stadial1 (17.0-14.7 kyrs BP), the
341 Bølling-Allerød (14.7-12.9 kyrs BP) the Younger Dryas Stadial (12.9-11.7 kyrs BP) and the
342 Holocene: early-Holocene, Greenlandian (11.7-8.326 kyrs BP), mid-Holocene, Northgrippian

343 (8.326-4.2 ka BP) and late-Holocene, Meghalayan (4.2-0.3 kyrs BP) (as from Fordham et al.,
344 2017).

345 The *O. aveyronensis* occurrences dataset consists of a set of field observations as well
346 as from records downloaded from the gbif (www.gbif.org) and the iNaturalist
347 (www.inaturalist.org) databases before manual curation. The Ecological Niche Modelling
348 analyses were conducted with the R-package *ENMwizard v.0.3.7* (Heming, Dambros &
349 Gutiérrez, 2018), which is a convenient wrapper of several tools. We first spatially filtered
350 occurrences to keep only those that were at least 5 km away from each other using the R-
351 package *spThin* (Aiello-Lammens, Boria, Radosavljevic, Vilella & Anderson, 2015). The
352 calibration area for the models was created as a buffer of 0.5° around the minimum convex
353 polygon encompassing all occurrences. From the 19 bioclimatic variables, we selected the less
354 correlated ones (Pearson correlation coefficient < 0.75) using the R-package *caret* (Kuhn,
355 2019) and kept seven variables: bio2, bio3, bio8, bio9, bio10 and bio17 (see details in
356 Supplementary Appendix S5) for further analyses.

357 We used the maximum entropy method (implemented in MaxEnt ver. 3.4.1, Phillips,
358 Anderson & Schapire, 2006; Philips & Dudík, 2008; Phillips et al., 2017) to calibrate models and
359 evaluated models' performance with the package *ENMeval* (Muscarella et al., 2014) as
360 implemented in *ENMwizard*. We evaluated models using a geographic partition scheme of
361 type "block" and optimized two parameters of MaxEnt: the Regularization Multipliers (RM)
362 and the Feature Classes (FCs). RM varied between 0.5 and 4.5, incremented by 0.5 whereas a
363 suite of 15 FCs (L, for Linear, P, for Product, Q, for Quadratic and H for Hinge) or combination
364 of them were evaluated: L, P, Q, H, LP, LQ, LH, PQ, PH, QH, LPQ, LPH, LQH, PQH, LPQH, resulting
365 in a total of 135 models. Model selection was done by computing the corrected Akaike
366 Information Criterion ("LowAIC"). Model accuracy was also evaluated by calculating omission
367 rates calculated when a 10th percentile threshold is applied (x10ptp). The final model was
368 projected on current and paleoclimatic contexts.

369

370 **Results**

371 **Population genomic dataset**

372 We obtained a total of 150 694 350 of read pairs across the 86 individuals (1 382 768 - 7 370
373 366 of raw reads per ind., mean 3 504 519, SD = 1 282011) of which 290 431 210 barcoded
374 reads (*i.e.* 96.36%) were retained, 10 819 014 (3.6%) were removed because RAD cutsite was
375 not found in the sequence and 138 476 reads (< 1%) were removed because of low quality.
376 We genotyped a total of 859 035 *loci* (composed of 135 255 870 sites) including 368 712
377 variant sites (SNPs) with *Stacks*. Average read depth ranged from 24.1X to 81.6X (mean 48.7X,
378 SD = 12.4X) based on the combination of parameters we used. After filtering the data with
379 *populations*, we finally kept 4302 *loci* from which we retained 9301 variant sites (SNPs).

380

381 **Genetic diversity, population differentiation and divergence**

382 The PCA biplot shows that individuals are arranged by subspecies/country of origin along PC1
383 (which explains 8.50% of the total genetic variance) and by sampling site along PC2 (which
384 explains 3.88% of the total genetic variance) even though the sampling sites of 1-Guilhaumard
385 and 2-Lapanouse-de-Cernon slightly overlap (Fig. 2a). Based on PC2, intra-population genetic
386 variation seems minimal for 1-Guilhaumard and 2-Lapanouse-de-Cernon and maximal for 3-
387 Saint-Affrique and 5-Larraona. Basic population genetics statistics are indicated in Table 1. All
388 populations displayed weak but significant deviation from panmixia with a mean $G_{IS} = 0.060$
389 ($p < 0.01$, 95% CI: 0.056-0.064) and population G_{IS} values ranging from 0.006 in 3-Saint-

390 Affrique to 0.081 in 4-Valgañón (all $p < 0.01$). Average nucleotide diversity (π) varies from
391 0.269 in 3-Saint-Affrique to 0.290 in 4-Valgañón and 5-Larraona, and allelic richness (A_R) from
392 16696 (in 3-Saint-Affrique) to 17297 (in 5-Larraona). Based on the filtering procedure we
393 followed, the number of private alleles was relatively low but consistently higher for *O. a.*
394 subsp. *vitorica* (*i.e.* from 9 to 23) compared to *O. a.* subsp. *aveyronensis* (*i.e.* from 0 to 2).

395 Overall G -statistics show that the whole set of populations is genetically structured:
396 $G_{ST} = 0.091$ ($p < 0.01$, 95% CI: 0.089-0.093); $G'_{ST} = 0.134$ ($p < 0.01$, 95% CI: 0.105-0.011); $G''_{ST} =$
397 0.150 ($p < 0.01$, 95% CI: 0.146-0.153) and Jost's $D = 0.047$ ($p < 0.01$, 95% CI: 0.046-0.048). The
398 AMOVA confirms that significant proportions of the total genetic variance are found among
399 populations within subspecies/country (6.8%, $F_{SC} = 0.072$, 95% CI: 0.070-0.074) and between
400 subspecies/country (6.1%, $F_{CT} = 0.061$, $F_{CT} = 0.058$ -0.064). Pairwise differentiation between
401 sampling sites varies from $F_{ST} = 0.049$ to 0.082 within subspecies/country and reaches $F_{ST} =$
402 0.111 to 0.147 between subspecies/country of origin (all $p < 0.01$; see Supplementary
403 Appendix S3 for details).

404 The clustering analysis carried out with sNMF suggests an optimal number of genetic
405 clusters of $K=5$ based on the cross-entropy criterion (Fig. 2c). We plotted the Q -matrices (that
406 contain individual admixture coefficients) corresponding to the best replicate runs for $K=2$ (to
407 investigate whether the program could distinguish the two subspecies) and $K=5$
408 (Supplementary Appendix S3). For $K=2$, the individuals of *O. a.* subsp. *aveyronensis* and *O. a.*
409 subsp. *vitorica* were unambiguously assigned to two distinct clusters. For the most likely
410 combination of $K=5$, we observe one cluster per sampling site except for the geographically
411 close localities of 1-Guilhaumard and 2-Lapanouse-de-Cernon that were grouped together in
412 a unique cluster. In the locality of 3-Saint-Affrique 4 out of 13 individuals (19-Oav-039, 19-Oav-
413 040, 19-Oav-049 and 19-Oav-051) display >50% of ancestry proportion that are associated to
414 the cluster associated with 1-Guilhaumard and 2-Lapanouse-de-Cernon.

415 The phylogenetic tree inferred from the SNP matrix corroborates these results. *O. a.*
416 subsp. *aveyronensis* and *O. a.* subsp. *vitorica* appeared as two well supported reciprocally
417 monophyletic groups (Fig. 2b). Sub-clades corresponding to sampling sites also confirm a
418 phylogenetic sub-clustering consistent with geography in each country for each subspecies
419 (with the exception of a group consisting of a mixture of individuals of all sites in France).

420

421 **Inference of past demographic history**

422 Reconstruction of historical dynamics of effective population size using STAIRWAYPLOT2
423 shows congruent trends consisting of progressive >95% decline in estimated N_e for all 6
424 populations, since the LGM, or at least, over the last 25000 years (see Fig. 3). Current effective
425 population sizes were estimated to vary between 302 (2-Lapanouse-de-Cernon) and 642
426 individuals (3-St-Affrique) with this method. Despite the current geographical discontinuity
427 between *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* populations, we attempted to
428 statistically compare demographic scenarios with and without ongoing migration using DILS.
429 The Posterior Probabilities associated with the models of ongoing migration (*i.e.* IM and SC)
430 relative to models without (*i.e.* SI and AM) was 0.59 based on the pooled dataset and ranged
431 from 0.55 to 0.66 based on the population pairs analyses. These probabilities are low and lead
432 to ambiguous model choices. Therefore, we cannot draw strong conclusions from model
433 comparisons with DILS (see Table 2, Fig. 4 and Appendix S4). Ambiguous inferences can
434 happen when the separation time between the two subspecies (here, *O. a.* subsp.
435 *aveyronensis* and *O. a.* subsp. *vitorica*) is too recent compared to their current population sizes
436 (Burban, Tenaillon & Glémin, 2024). The median of the parameters estimated based on

437 posterior distributions made under the most probable model (*i.e.* Isolation-with-Migration,
438 IM) indicate very low migration rates ($N.m = 0.60$ for the pooled dataset, and it ranges from
439 0.56 to 0.92 based on population pairs). With the pooled dataset, we inferred a subdivision of
440 the ancestral population of about 75 000 individuals based on two pools (90 000 to 103 200
441 based on population pairs) into two daughter populations about 1500 generations ago (650
442 to 3880 based on population pairs), which convert into 4500 years assuming a generation time
443 of 3 years, 7500 years assuming a generation time of 5 years and 10500 years assuming a
444 generation time of 7 years. These estimates of splitting times were found to be even more
445 recent based on the population pairs (*i.e.* namely, sampling sites) data set. This suggests that
446 most coalescence events are taking place in the ancestral population and not in the respective
447 daughter populations. In other words, the amount of shared polymorphism is still relatively
448 high because of insufficient time for drift to differentially fix/eliminate alleles within the two
449 subranges (*i.e.* Incomplete Lineage Sorting). Similar to what was inferred with STAIWAYPLOT2,
450 our analyses also point to a decrease of effective sizes compared to the ancestral population
451 (*i.e.* current N_e have values corresponding to about 10% of N_{anc}). The goodness-of-fit test
452 indicates that the best estimated scenario (IM) reproduces the observed data set with high
453 fidelity (see Supplementary Appendix S4).

454

455 **Current and past Species Distribution Modelling**

456 ENMeval analyses identified RM = 1 and a mixture of Product (P), Quadratic (Q) and Hinge (H)
457 feature classes as the best-performing parameters for calibrating the final ENM. These
458 parameters yielded a single “best” candidate model with the lowest AICc score (925.02) and
459 an AICc weight of ~0.23. This model had mean omission rates of 0.165 (for the 10th percentile)
460 and a mean test AUC of 0.834 (see variable contributions following this model in Table 3 and
461 Supplementary Appendix S5, for additional details). The spatial projection of the current
462 bioclimatic niches inferred for *O. aveyronensis* corresponds to its actual geographic
463 distribution but show that some additional areas may be climatically suitable for this orchid
464 (Fig. 5). Spatial projections considering paleoclimatic conditions show that the extent of
465 suitable habitat for *O. aveyronensis* may have reached its apex during the Younger Dryas cold
466 period (12.9-11.7 kyrs BP) before starting to decrease across the Holocene. The LGM
467 conditions (21 kyrs BP) were not suitable for the establishment of viable populations in the
468 studied area but important suitable habitats may have existed during the Heinrich Stadial 1
469 event (17-14.7 kyrs BP) and the Bøllering-Allerød warming (14.7 – 12.9 kyrs BP). Altogether,
470 these results are consistent with the possible existence of a unique ancestral area, centered
471 on South-Western France that could have experienced a split and a combination of reduction
472 and shift towards cooler mid-elevation areas, in the North-East (for *O. a.* subsp. *aveyronensis*)
473 and in the South-West (for *O. a.* subsp. *vtorica*) during the Holocene. The results available
474 from WorldClim data do not contradict these conclusions (Appendix S5). However, the latter
475 results show significant differences across GCMs and even for a given GCM but downscaled
476 based on different interpolation algorithms: namely, WorldClim instead of CHELSA.

477

478 **Discussion**

479 Some *Ophrys* orchid lineages (e.g. *O. sphegodes sensu lato*) boast some of the highest
480 diversification rates reported worldwide for orchids and angiosperms in general (Breitkopf et
481 al., 2015; see also Thompson, Davis, Dodd, Wills & Priest, 2023) and are being increasingly
482 mentioned as biological models of interest for studying ecological speciation and adaptive
483 radiations (e.g. Nosil et al., 2017; Baguette et al., 2020; Nürk et al., 2020). Ecological and

484 sometimes sympatric speciation in *Ophrys* is confidently explained by the selective pressures
485 exerted by pollinators and the high degree of specificity required to ensure the success of the
486 sexual swindling strategy (see Baguette et al., 2020). Nevertheless, allopatric-vicariant
487 divergence undoubtedly also plays an important role in the diversification of *Ophrys*. In this
488 study, we used population genomics and ENM to characterise one of the most recent cases of
489 allopatric-vicariant divergence for a continental plant. *Ophrys aveyronensis* therefore provides
490 a promising opportunity to investigate the relative contribution of biogeographic and
491 ecological factors in shaping early stages of evolutionary divergence in rapidly diversifying
492 groups of organisms.

493

494 **Genetic diversity and differentiation in *O. aveyronensis***

495 Although populations of *Ophrys aveyronensis* from southern France (subsp. *aveyronensis*) and
496 northern Spain (subsp. *vitorica*) differ only in subtle ways on the basis of phenotypic data
497 (Gibert et al., 2022; 2024), our results allowed to unambiguously tease them apart based on
498 population genomics. Both the PCA, the *F*-statistics (and/or their analogues), the clustering
499 analysis and the phylogenetic tree confirm that the two subspecies form distinct genetic
500 entities, which themselves show a lower genetic substructure between sampling sites within
501 subranges. The reciprocal monophyly of the two disjunct subspecies (with no outlier
502 individuals deviating from the global pattern) *a priori* rules out a recent dispersal scenario. We
503 found that all the populations sampled had similar levels of allelic richness, heterozygosity,
504 departure from panmixia and number of private alleles. There is therefore no evidence that
505 any of these six populations represent a subset of the genetic diversity of a putative source
506 population elsewhere. Finally, the inference of the temporal dynamics of effective population
507 size suggests that each of the sampled populations can be considered a kind of replicate of
508 each other. All the populations displayed consistent decline in effective population size over
509 the last tens of thousands of years, suggesting that they shared a similar demographic history.
510 In summary, these results are fully consistent with a scenario of range splitting, and
511 subsequent contraction of two subranges with restricted inter-regional gene flow.

512 Delforge (2016) hypothesised that the Iberian populations could be the result of a
513 hybrid swarm (perhaps involving *Ophrys castellana* as one putative parental species),
514 undergoing spatial and demographic expansion. According to this scenario, the two
515 morphologically similar population groups may have evolved convergent phenotypes as a
516 result of their adaptation to the same insect pollinator species on both sides of the Pyrenees.
517 Such evidence of phenotypic convergence between *Ophrys* species originating from different
518 phylogenetic lineages has been already documented (Stöckl et al., 2005; Gögler et al., 2009;
519 Sramkó, Gulyas & Molnár, 2010). However, without ruling out the possibility that both
520 subspecies have experienced gene flow with other sympatric *Ophrys* taxa, the striking
521 similarity we found in patterns of genetic diversity, as well as in the temporal trends we
522 inferred for the effective sizes of all populations sampled in this study (regardless their
523 geographic origin) do not support this hypothesis, but rather suggest that *O. a.* subsp.
524 *aveyronensis* and *O. a.* subsp. *vitorica* originated from a single and direct common ancestor.

525 The demographic scenarios we tested remain relatively simple and the Bayesian
526 posterior probabilities we obtained with DILS show that the approach implemented in this
527 program, while favouring ongoing migration, could not strongly support such a model over
528 isolation ones. A theoretical explanation for this otherwise puzzling result is that the time of
529 split is too recent to observe sufficient allele sorting (and signal) from a diverse ancestral
530 population with relatively high effective population size to be able to detect a cessation of

531 gene flow between the two species (see Smith & Hahn, 2024, Burban et al. 2024). In addition,
532 events that are known to deceive demographic inferences such as ‘ghost introgression’ from
533 unsampled lineages (Tricou, Tannier & De Vienne, 2022) may have a non-negligible probability
534 to occur in *Ophrys*. Investigating whether or not *Ophrys aveyronensis* has a hybrid origin, and
535 at least whether and to what extent the two taxa have been introgressed by other *Ophrys* taxa
536 will require the genomic analyses of additional species in a future study. Nevertheless, we are
537 confident that the time of split we inferred here is among the rare cases documenting a so
538 recent speciation event in a plant continental system (see de Queiroz et al., 2022), given that
539 ghost introgression may have led to an overestimation of the time of split (Pang & Zang 2023).

540

541 **Causes of the disjunct distribution of *O. aveyronensis***

542 The results of the Ecological Niche Modelling show that the conditions of the LGM were not
543 suitable for the establishment of viable populations of *O. aveyronensis*. This suggests that
544 either this species diversified from an ancestral *Ophrys* lineage in the study area after this
545 period, or that its populations occupied a relatively remote glacial refugium during this period
546 (the ENM results suggest that the Italian peninsula could have provided such a refugium, see
547 Fig. 5). During the last 21000 years, the extent of suitable habitat for *O. aveyronensis* was
548 rather large until the Holocene (about 12000 years ago) (Fig. 5), when its area began to
549 decrease dramatically while maintaining an area of relatively high suitability centred on
550 southwestern France. The split and, to a lesser extent, the northward and southward shifts
551 may have occurred as late as during the Late Holocene (from 4000 years ago) (Fig. 5).
552 Interestingly, this timing is the same order of magnitude as the time of the split we inferred
553 from the population genomic data, and is consistent with a scenario of allopatric divergence
554 (without being able to completely exclude transient gene flow through climate-induced
555 oscillations throughout the Holocene) from an ancestral population with relatively large
556 effective size.

557 The Pyrenees mountains have the peculiarity of acting as a biogeographic boundary
558 for many terrestrial organisms, while being at the crossroads between temperate and
559 Mediterranean biomes (see Pironon, Gómez, Font & García, 2022 and references therein).
560 Although orchid propagules have the capacity to disperse efficiently over long distances,
561 several taxa occur on either the southern or the northern slopes of the Pyrenees, but have
562 ranges that do not extend latitudinally on either side, or only weakly so (*i.e.* the so-called
563 northern or southern peripheral species in Pironon et al., 2022). The Pyrenees may therefore
564 alter the dispersal of at least some orchid species and populations (e.g. *Dactylorhiza insularis*,
565 *Gymnadenia gabasiana*, *Neotinea conica*, *Ophrys castellana*, *O. riojana* and *O. subinsectifera*,
566 *Orchis langei* and *O. simia*) due to various biotic or abiotic factors. This mostly granitic and
567 relatively high mountain channel forms a vast area of unsuitable habitat for *Ophrys* orchids,
568 which generally require calcareous soils and lowland or midland environments. Although data
569 on their geographical distribution remain incomplete, we may also speculate that the absence
570 of *Ophrys*-specific pollinator species and/or of fungi on which the orchids depend for their
571 germination and their development may also have limited the possibility of establishment of
572 intermediate orchid populations between the two current subranges. For *O. aveyronensis*, our
573 results are consistent with a general reduction in gene flow between both sides of the
574 Pyrenees through the formation and the contraction of the two subranges during the
575 Holocene until the present cessation of gene exchange, which is too recent to be supported
576 on the basis of the model-based demographic inferences but that is corroborated by our
577 descriptive genetic analyses.

578

579 **Ecological speciation in addition to vicariance?**

580 Selection pressures exerted by insect pollinators are likely to be the main driver of ecological
581 and presumably sympatric speciation in *Ophrys* orchids (see Baguette et al., 2020). Thus,
582 populations of a given *Ophrys* species that are pollinated by distinct insect pollinator species
583 can be considered to be undergoing a speciation process. So far, field observations support
584 that *O. aveyronensis* from the Grands Causses and their Iberian counterparts are pollinated by
585 the same pollinator species: the solitary bee *Andrena hattorfiana* (see Benito Ayuso, 2019).
586 However, in a recent study on the same set of individuals, we showed that *O. aveyronensis*
587 from the Grands Causses and from the Iberian Peninsula show subtle differences in a number
588 of morphological and colour traits that may be involved in pollination (Gibert et al., 2022;
589 2024). These subtle differences may arise as a result of local adaptation to a “secondary”
590 pollinator species (see Joffard et al., 2019; Schatz, Genoud, Claessens & Kleynen, 2020;
591 Baguette et al., 2021). Ongoing work is now addressing whether this variation is indeed
592 actually adaptive and involved in evolutionary divergence i) by assessing which phenotypic
593 traits may be under divergent selection among the *O. aveyronensis* populations across their
594 range and which ones may be evolving ‘non-ecologically’ and ii) by investigating putative
595 genomic regions associated with phenotypic trait values, genetic structure or ecological
596 factors, something that remains largely underexplored in *Ophrys* so far (but see Gibert et al.,
597 2024).

598 Cross-pollination experiments between individuals of the two subranges could also
599 help to highlight the insurgence of post-zygotic reproductive isolation in *Ophrys aveyronensis*.
600 However, such work may be hampered by the difficulty of growing *Ophrys* under controlled
601 conditions and by the protection status of *Ophrys aveyronensis* in France. Current knowledge
602 suggests that such reproductive barriers may be weak or absent at the scale of the *Ophrys*
603 genus (Soliva and Widmer, 2003; Scopece, Mesacchio, Widmer & Cozzolino, 2007, Gervasi et
604 al., 2017). Notable exceptions to this pattern generally involve phylogenetically distant taxa
605 such *O. incubacea* and *O. iricolor* (Cortis et al., 2009) or species with different ploidy levels,
606 such as the diploid *O. exaltata* and the tetraploid *O. lupercalis* (Vereecken, Cozzolino &
607 Schiestl, 2010), but recent studies also suggest that postmating barriers may exist between
608 *Ophrys* taxa from a same group: *O. annae* and *O. chestermanii* (Lussu, Se Agostini, Marignani,
609 Cogoni & Cortis, 2018) and remain to be tested in *O. aveyronensis*.

610

611 **Taxonomic and conservation implications**

612 Given that populations of *O. aveyronensis* share a same insect pollinator species (*Andrena*
613 *hattorfiana*) throughout their range and have similar environmental niches, we currently have
614 no evidence to even consider the existence of distinct ecotypes in *Ophrys aveyronensis*.
615 However, the disjunct geographic distribution in two subranges, separated by ~600 km and
616 the Pyrenees Mountains suggests that populations from the Grands Causses and from the
617 Iberian Peninsula may be considered as distinct evolutionary (and conservation) units. These
618 two sets of populations display differentiation at several phenotypic traits (Gibert et al., 2022;
619 Gibert et al., 2024). Here, our population genomic analyses suggest that both sets of populations
620 are genetically differentiated and are unlikely to experience ongoing gene flow. All these
621 considerations are consistent with the generally admitted definition of the subspecies and we
622 propose to use the taxonomic status proposed by Paulus (2017) to describe the France and
623 Iberian populations with the names: *Ophrys aveyronensis* subsp. *aveyronensis* and *Ophrys*
624 *aveyronensis* subsp. *vitorica*, respectively.

625 Recent studies have elaborated on some empirical evidence that the delineation of the
626 so-called ‘micro-species’ in *Ophrys* is practically cumbersome (Bateman et al., 2021; Bateman
627 and Rudall, 2023), especially in the section *Sphegodes* to which *O. aveyronensis* belongs. Some
628 genomic data, such as whole plastomes, which are useful to delineate the main *Ophrys*
629 *lineages*, show a sequence similarity degree of more than 99.5% between ‘microspecies’ and
630 therefore do not show enough variation to allow to distinguish them (Bertrand, Gibert, Llauro
631 & Panaud, 2019; Bertrand Gibert, Llauro & Panaud, 2021; Bateman et al., 2021). Here,
632 however, our results demonstrate that an approach consisting of sampling populations (*i.e.* a
633 sufficient number of individuals at a given location) and appropriate genetic markers (e.g.
634 SNPs genotyped on populations based on RAD-seq protocols) provides signals that are
635 consistent with taxonomy. We therefore argue that such approaches of population genomics
636 within an integrative taxonomy framework (relying on flower morphology, pollinator identity,
637 ecological features...) is by far the most fruitful for better understanding the process of
638 evolutionary divergence and, at the same time, for improving our knowledge of the taxonomy,
639 systematics, ecology and evolution of *Ophrys* orchids (see also Salvado et al., 2024).

640

641 ***Ophrys aveyronensis* as a promising biological model to study speciation**

642 Vicariance is a very common route toward speciation (Hewitt et al., 2000), but most of
643 empirical cases document it retrospectively, between nascent species that exhibit at the same
644 time obvious geographic separation and substantial levels of both phenotypic and genetic
645 differentiation (Dool, Picker & Eberhard, 2021). In this context, studies reporting cases of
646 incipient allopatric divergence are rare in plants, especially in continental systems (see de
647 Queiroz et al., 2022). The allopatric divergence in the *Ophrys aveyronensis* we report here is
648 indeed extremely recent compared to other species with disjunct geographic distributions in
649 terrestrial systems (e.g. González-Serna, Cordero & Orego, 2018; Tomasello, Karbstein, Hodač,
650 Paetzold & Hörandl, 2020; Wang et al., 2022; Cao et al., 2022; Balmori-de la Puente et al.,
651 2022). It is likely to have arisen as a result of climate warming, not as a consequence of
652 Quaternary climate fluctuations but as late as following the Last Glacial Maximum, which
653 disrupted an ancestral geographic distribution and subsequently constrained the spatial
654 dynamics of its subranges and does not *a priori* imply a human-induced biological
655 introduction. This system may therefore serve as a natural laboratory to study parallel
656 adaptation at early stages of reproductive isolation insurgence, following climate-induced
657 range fragmentation.

658

659 **Acknowledgements**

660 This work was primarily supported by an inter-LabEX TULIP-CEMEB initiative (n°197310) to J.
661 Bertrand and B. Schatz. This research was also funded by an Agence Nationale pour la
662 Recherche Jeune Chercheur Jeune Chercheuse (ANR JCJC) grant to J. Bertrand, grant number
663 ANR-21-CE02-0022-01, and is set within the framework of the “Laboratoires d’Excellences
664 (LABEX)” TULIP [ANR-10- LABX-41]. Lastly, this research was also funded by the Observatoire
665 de REcherche Méditerranéen de l’Environnement (SO Ocove & SO PolliMed, OSU OREME) to
666 B. Schatz. We thank J.-L. Roux and D. Vizcaïno for the information provided about sampling
667 sites in Spain. We thank C. Moliné for his contribution to fieldwork, M. El Baidouri and P.
668 Salvado for their support with bioinformatic analyses. P. Comes, M. James and other
669 anonymous reviewers also provided constructive feedback on earlier versions of the
670 manuscript.

671

672 **Conflict of interest**

673 The authors declare no conflict of interest.

674

675 **Data availability statement**

676 Sequencing data have been submitted to the European Nucleotide Archive (ENA;
677 <https://www.ebi.ac.uk/ena/browser/home>) under Study with primary accession
678 n°PRJEB61037 (and secondary accession number ERP146119) and samples accession
679 n°ERS14864169 (SAMEA112857509) to n°ERS14864254 (SAMEA112857594).

680

681 **Author contributions**

682 JB, BS, AG and MB designed the study. AG, JB, BS and RB gathered the data.

683 AG and JB performed the analyses with contribution from CF and CR and wrote the first
684 version of the manuscript. All authors contributed substantially to the revisions.

685

686 **References**

- 687 Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., & Anderson, R.P. (2015) spThin:
688 an R package for spatial thinning of species occurrence records for use in ecological
689 niche models. *Ecography*, 38, 541-545.
- 690 Baguette, M., Bertrand, J.A.M, Stevens, V.M., & Schatz, B. (2020) Why are there so many bee-
691 orchid species? Adaptive radiation by intra-specific competition for mnesic pollinators.
692 *Biol Rev.*, 95, 1630-1663.
- 693 Balmori-de la Puente, A., Ventura, J., Miñarro, M., Somoano, A., Hey, J., & Castesana, J. (2022)
694 Divergence time estimation using ddRAD data and an isolation-with-migration model
695 applied to water vole populations of *Arvicola*, *Scientific Reports*, 12, 4065.
- 696 Bateman, R.M., Rudall, P. J., Murphy, A.R.M., Cowan, R.S., Devey, D.S., & Pérez-Escobar, O.A.
697 (2021) Whole plastomes are not enough: phylogenomic and morphometric
698 exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect.
699 *Sphegodes*. *Journal of Experimental Botany*, 72, 654-681.
- 700 Bateman, R.M., & Rudall, P.J. (2023) Morphological continua make poor species: genus-wide
701 morphometric survey of the European Bee Orchids (*Ophrys* L.). *Biology*, 12, 136.
- 702 Benito Ayuso, J. (2019) Estudios sobre polinización en el género *Ophrys* (Orchidaceae), I, *Flora*
703 *Montiberica*, 74, 32-37.
- 704 Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2019) Characterization of the complete
705 plastome of *Ophrys aveyronensis*, a Euro-Mediterranean orchid with an intriguing
706 geographic distribution. *Mitochondrial DNA Part B*, 4, 3256-3257.
- 707 Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2021) Whole plastid genome-based
708 phylogenomics supports an inner placement of the *O. insectifera* group rather than a
709 basal position in the rapidly diversifying *Ophrys* genus (Orchidaceae). *Biology Letters*,
710 168, 452-457
- 711 Bertrand, J. A. M., Baguette, M., Joffard, N., & Schatz, B. (2021) Challenges inherent in the
712 systematics and taxonomy of genera that have recently experienced explosive
713 radiation: the case of orchids of the genus *Ophrys*. In Grancolas, P. & Maurel, M.-C.
714 *Systematics and the Exploration of Life* (pp 113-134). Hoboken, New Jersey: Wiley.
- 715 Breitkopf, H., Onstein, R.E., Cafasso, D., Schlüter, P.M., & Cozzolino, S. (2015) Multiple shifts
716 to different pollinators fuelled rapid diversification in sexually deceptive *Ophrys*
717 orchids. *New Phytologist*, 207, 377-389.

- 718 Burban, E., Tenailon, M.I., & Glémin, S. (2024) RIDGE, a tool tailored to detect gene flow
719 barriers across species pairs. *Molecular Ecology Resources*, 24, e13944.
- 720 Cao, Y., Almeida-Silva, F., Zhang, W.-P., Ding, Y.-M., Bai, D., Bai, W.-N., Zhang, B.W., Van de
721 Peer, Y., & Zhang, D.Y. (2023) Genomic Insights into Adaptation to Karst Limestone
722 and Incipient Speciation in East Asian *Platycarya* spp. (Juglandaceae). *Molecular*
723 *Biology and Evolution*, 40, msad121.
- 724 Catchen, J., Amores, A., Cresko, W., & Postlewait, J. (2011) Stacks: building and genotyping
725 loci de novo from short-read sequences. *G3: genes, Genomes, Genetics*, 1, 171-182.
- 726 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2013) Stacks: an
727 analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140.
- 728 Cortis, P., Vereecken, N.J., Schiestl, F.P., Barone Lumaga, M.R., Scrugli, A., & Cozzolino, S.
729 (2009) Pollinator convergence and the nature of species' boundaries in sympatric
730 Sardinian *Ophrys* (Orchidaceae). *Annals of Botany*, 104, 497-506.
- 731 Coyne, J.A., & Orr, H.A. (2004). *Speciation*. Sunderland, Mass: Sinauer Associates.
- 732 Claessens, J., & Kleynen, J. (2016) *Orchidées d'Europe : fleurs et pollinisation*. Biotope, Mèze,
733 France.
- 734 De Queiroz, L.J., Doenz, C.J., Altermatt, F., Alther, R., Borko, S, Brodersen, J., Gossner, M.M.,
735 Graham, C., Matthews, B., McFadden, I.R., Pélissier, L., Schmitt, T., Selz, O.M., Villalba,
736 S., Rüber, L., Zimmermann, N.E., & Seehausen, O. (2022) Climate, immigration and
737 speciation shape terrestrial and aquatic biodiversity in the European Alps. *Proceedings*
738 *of the Royal Society B*, 289: 20221020.
- 739 Delforge, P. (2016) *Orchidées d'Europe, d'Afrique du Nord et du Proche Orient*. Lausanne,
740 Switzerland: Delachaux & Niestlé.
- 741 Dool, S.E., Picker, M.D., & Eberhard, M.J.B. (2021) Limited dispersal and local adaptation
742 promote allopatric speciation in a biodiversity hotspot. *Molecular Ecology*, 31, 279-
743 295.
- 744 Fick, S.E., & Hijmans, R.J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces for
745 global land areas. *International Journal of Climatology*, 37, 4302-4315.
- 746 Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T. M. L., Otto-Bliesner, B. L., Chan, K. C., &
747 Brook, B.W. (2017) PaleoView: a tool for generating continuous climate projections
748 spanning the last 21 000 years at regional and global scales. *Ecography*, 40, 1348-1358.
- 749 Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguer, J., Loire E., Simon, A.,
750 Galtier, N., Duret, L., Bierne, N., Vekemans, X., & Roux C. (2021) DILS: Demographic
751 inferences with linked selection using ABC. *Molecular Ecology Resources*, 2629-2644.
- 752 Francis, R.M. (2017) POPHELPER: an R package and web app to analyse and visualize
753 population structure. *Molecular Ecology Resources*, 17, 27-32.
- 754 Frichot, E., Mathieu, F., Trouillon T., Bouchard, G., & François O. (2014) Fast and efficient
755 estimation of individual ancestry coefficients. *Genetics*, 196, 973-983.
- 756 Frichot, E., & François O. (2015) LEA: An R package for landscape and ecological association
757 studies. *Methods in Ecology and Evolution*, 6, 925-929
- 758 Gervasi, D.L., Selosse, M.-A., Sauve, M., Francke, W., Vereecken, N.J., Cozzolino, S., & Schiestl,
759 F.P. (2017) Floral scent and species divergence in a pair of sexually deceptive orchids.
760 *Ecology and Evolution*, 7, 6023-6034.
- 761 Gibert, A., Louty, F., Buscail, R., Baguette, M., Schatz, B., & Bertrand, J.A.M. (2022) Extracting
762 quantitative information from images taken in the wild: a case study of two vicariants
763 of the *Ophrys aveyronensis* species complex. *Diversity*, 14, 400.

764 Gibert, A., Schatz, B., Buscail, R. Nguyen, D., Baguette, M., Barthes, N., & Bertrand, J.A.M.
765 (2024) Floral phenotypic divergence and genomic insights in an *Ophrys* orchid:
766 unraveling early speciation processes. *bioRxiv*.

767 Gögler, J., Stökl, S., Sramkova, A., Twele, R., Francke, W., Cozzolino, S., Cortis, P., Srugli, A., &
768 Ayasse, M. (2009) Ménage à trois – two endemic species of deceptive orchids and one
769 pollinator species. *Evolution*, 63, 222-2234.

770 González-Serna, M., Cordero, P.J., & Ortego, J. (2018) Using high-throughput sequencing to
771 investigate the factors structuring genomic variation of a Mediterranean grasshopper
772 of great conservation concern. *Scientific Reports*, 8, 13436.

773 Heming, N. M., Dambros, C., & Gutiérrez, E. E. (2018). ENMwizard: AIC model averaging and
774 other advanced techniques in Ecological Niche Modeling made easy. R package version
775 0.1.7. Retrieved from <https://github.com/HemingNM/ENMwizard>

776 Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

777 Hijmans, R. J., Cameron, S.E., Parra, J. -L., Jones, P. J., & Jarvis A. (2005) Very high resolution
778 interpolated climate surfaces for global land areas. *International Journal of*
779 *Climatology*, 25, 1965-1978.

780 Inda, L.A., Pimentel, M., & Chase, M.W. (2012) Phylogenetics of tribe Orchideae (Orchidaceae:
781 Orchidoideae) based on combined DNA matrices: inferences regarding timing of
782 diversification and evolution of pollination syndromes. *Annals of Botany.*, 110, 71-90.

783 James, M.E., Allsopp, R.N., Groh, J.S., Kaur, A., Wilkinson, M.J., & Ortiz-Barrientos, D. (2023)
784 Uncovering the genetic architecture of parallel evolution. *Molecular Ecology*, 32, 5575-
785 5589.

786 Joffard, N., Massol, F., Grenié, M., Montgellard, C., & Schatz, B. (2019) Effect of pollination
787 strategy, phylogeny and distribution niches of Euro-Mediterranean orchids. *Journal of*
788 *Ecology*, 107, 478-490.

789 Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers.
790 *Bioinformatics*, 24, 1403-1405.

791 Jombart, T., & Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP
792 data. *Bioinformatics*.

793 Kadereit, J.W., & Abbott, R.J. (2023) Plant speciation in the Quaternary. *Plants Ecology &*
794 *Diversity*, 14, 105)-142.

795 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.F.K., von Haeseler, A., & Jermin, L.S. (2017)
796 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
797 *Methods*, 14, 587-589.

798 Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021): CHELSA-
799 TraCE21k v1. 0. Downscaled transient temperature and precipitation data since the
800 last glacial maximum. *Climate of the Past Discussions*, 19, 439-456.

801 Krasovec, M., Chester, M., Ridout, K., & Filatov, D.A. (2018) The mutation rate and the age of
802 the sex chromosomes in *Silene latifolia*. *Current Biology*, 28, 1832-1838.

803 Kreutz, C.A. (2007) Beitrag zur Taxonomie und Nomenklatur europäischer, mediterraner,
804 nordafrikanischer und vorderasiatischer Orchideen. *Ber. Arbeitskrs. Heim. Orchid.* 24,
805 77-141.

806 Kuhn, M. (2019) caret: classification and regression training. [https://CRAN.R-](https://CRAN.R-project.org/package=caret)
807 [project.org/package=caret](https://CRAN.R-project.org/package=caret).

808 Liu, X., & Fu, Y.-X. (2015) Exploring population changes using SNP frequency spectra. *Nature*
809 *Genetics*, 47, 555-559

- 810 Liu, X., & Fu, Y.-X. (2020) Stairway Plot 2: demographic inference with folded SNP frequency
811 spectra. *Genome Biology*, 21, 280.
- 812 Lussu, M., De Agostini, A., Marignani M., Cogoni, A., & Cotis P. (2018) *Ophrys annae* and
813 *Ophrys chestermanii*: an impossible love between two orchids sister species. *Nordic*
814 *Journal of Botany*, e01798.
- 815 Lussu, M., De Agostini, A., Cogoni, A., Marignani M., & Cotis P. (2019) Does size really matter?
816 A comparative study on floral traits in orchids with two different pollination strategies.
817 *Plant Biology*, 21, 961-966.
- 818 Mayr E. (1942) *Systematics and the Origin of Species*. Columbia University Press, New York,
819 USA.
- 820 Meirmans, P. G. (2020) GENODIVE version 3.0: Easy-to-use software for the analysis of genetic
821 data of diploids and polyploids. *Molecular Ecology Resources*, 20, 1126-1131.
- 822 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, A., von Haeseler, A., &
823 Lanfear, R. (2020) IQ-TREE 2: New methods and efficient methods for phylogenetic
824 inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530-1534.
- 825 Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., &
826 Anderson, R.P. (2014) ENMeval: an R package for conducting spatially independent
827 evaluations and estimating optimal model complexity for Maxent ecological niche
828 models. *Methods in Ecology and Evolution*, 5, 1198-1205.
- 829 Nosil, P., Feder, J.L., Flaxman, S.M., & Gompert Z. (2017) Tipping points in the dynamics of
830 speciation. *Nature Ecology and Evolution*, 1, 0001.
- 831 Nürk, M.N., Linder, H.P., Onstein, R.E., Larcombe, M.J., Hughes, C.H., Piñeiro Fernández, L.,
832 Schlüter, P.M., Valente, L.M., Beierkuhnlein, C., Cutts, V., Donoghue, M.J., Edwards,
833 E.J., Field, R., Flantua, S.G.A., Higgins, S.I., Liede-Schumann, S., & Pirie, M.D. (2020)
834 Diversification in evolutionary arenas – assessment and synthesis. *Ecology and*
835 *Evolution*, 10, 6163-6182.
- 836 Pang, X.-X., & Zhang, D.-Y. (2023) Impact of ghost introgression on coalescent-based species
837 tree inference and estimation of divergence time. *Systematic Biology*, 72, 35-49.
- 838 Paris, J., Stevens, J.R., & Catchen J.M. (2017) Lost in parameter space: a road map for STACKS.
839 *Methods in Ecology and Evolution*, 8, 1360-1373.
- 840 Paulus, H.F., & Gack, C. (1999) Bestäubungsbiologische Untersuchungen an der Gattung
841 *Ophrys* in der Provence (SO Frankreich), Ligurien und Toscana (NW Italien)
842 (Orchidaceae und Insecta, Apoidea). *Journal Europäischer Orchideen*. 31, 347-422.
- 843 Paulus, H.F. (2017) Bestäubungs-biologie *Ophrys* in Nordspanien. *Journal Europäischer*
844 *Orchideen*. 49, 427-471.
- 845 Pironon, S., Gómez, D., Font, X., & García, M.B. (2022) Living at the limit in the Pyrenees:
846 Peripheral and endemic plants are rare but underrepresented in protection lists.
847 *Diversity and distribution*, 28, 930-943.
- 848 Phillips, S.J., Anderson, R.P., & Schapire, R.E. (2006) Maximum entropy modelling of species
849 geographic distributions. *Ecological Modelling*, 190, 231-259.
- 850 Phillips, S.J., & Dudík, M. (2008). Modelling of species distributions with Maxent: New
851 extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.
- 852 Philipps, S.J., Anderson, R.P., Dudík, M., Shapire, R.E., & Blair, M.E. (2017) Opening the black
853 box: an open-source release of Maxent. *Ecography*, 40, 887-893.
- 854 Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig,
855 B., & Westram, A.M. (2017) Interpreting the genomic landscape of speciation: a

856 roadmap for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450-
857 1477.

858 Salvado, P., Aymerich Boixader, P., Vila Bonfill, A., Martin, M., Quélenec, C., Lewin, J.-M.,
859 Hinoux, V., & Bertrand, J.A.M. (2022) Little hope for the polyploid endemic Pyrenean
860 Larkspur (*Delphinium montanum*): evidences form population genomics and Ecological
861 Niche Modeling. *Ecology and Evolution*, 12, e8711.

862 Salvado, P., Gibert, A., Schatz, B., Vandabeele, L., Buscail, R., Vilasís, D., Feldmann P., &
863 Bertrand, J.A.M. (2024) Contrasting patterns of differentiation among three taxa of the
864 rapidly diversifying orchid genus *Ophrys* sect. *Insectifera* (Orchidaceae) where their
865 ranges overlap. *bioRxiv*. doi: 10.1101/2024.04.23.590674

866 Schatz, B., Genoud, D., Claessens, J., & Kleyen, J. (2020) Orchid-pollinator network in Euro-
867 Mediterranean region: what we know, what we think we know, and what remains to
868 be done. *Acta Oecologica*, 107, 103605.

869 Scopece, G., Musacchio, A., Widmer, A., & Cozzolino, S. (2007) Patterns of reproductive
870 isolation in Mediterranean deceptive orchids. *Evolution*, 61, 2623-2642.

871 Sedeek, K.E.M., Scopece, G., Staedler, Y.M., Schönenberger, J., Cozzolino, S., Schiestl, F. P., &
872 Schlüter, P.M. (2014) Genic rather than genome-wide differences between sexually
873 deceptive *Ophrys* orchids with different pollinators. *Molecular Ecology*, 23, 6192-6205.

874 Smith, M.L., & Hahn, M. (2024) Selection leads to false inferences of introgression using
875 popular methods. *Genetics*, 7, iyae089.

876 Sobel, J.M., Chen, G.F., Watt, L.R., & Schemske, D.W. (2010) The biology of speciation.
877 *Evolution*, 64, 295-315.

878 Sobel, J.M. (2014) Ecogeographic isolation and speciation in the genus *Mimulus*. *American*
879 *Naturalist*, 184, 565-579

880 Soliva, M., & Widmer, A. (2003) Gene flow across species boundaries in sympatric sexually
881 deceptive *Ophrys* (Orchidaceae) species. *Evolution*, 57, 2252-2261.

882 Sramkó, G., Gulyas, G., & Molnár, A. (2011) Convergent evolution in *Ophrys kotschyi*
883 (Orchidaceae) revisited: a study using nrITS and cpIGS sequences. *Annales Botanici*
884 *Fennici*, 48, 97-106.

885 Stankowski, S., & Ravinet, M. (2021) Defining the speciation continuum. *Evolution*, 75, 1256-
886 1273.

887 Stökl, J., Paulus, H., Dafni, A., Schulz, C., Francke, W., & Ayasse, M. (2005) Pollinator attracting
888 odour signals in sexually deceptive orchids of the *Ophrys fusca* group. *Plant*
889 *Systematics and Evolution*, 254, 105-120.

890 Thompson, J.B., Davis, K.E., Dodd, H.E., Wills, M.A., & Priest, N.K. (2023) Speciation across the
891 Earth driven by global cooling in terrestrial orchids. *Proceedings of the National*
892 *Academy of Sciences*, 120, e210240820.

893 Tomasello, S., Karbstein, K., Hodač, L., Paetzold, C., & Hörandl, E. (2020) Phylogenomics
894 unravels Quaternary vicariance and allopatric speciation patterns in temperate-
895 montane plant species: A case study on the *Ranunculus auricomus* species complex.
896 *Molecular Ecology*, 29, 2031, 2049.

897 Tricou, T., Tannier, E., & De Vienne D.M. (2022) Ghost lineages highly influence the
898 interpretation of introgression tests. *Systematics Biology*, 71, 1147-1158.

899 Turco, A., Albano, A., Medagli, P., Wagensommer, R.P., & D'Emerico, S. (2023) Comparative
900 cytogenetic of the 36-chromosomes genera of Orchidinae subtribe (Orchidaceae) in
901 the Mediterranean region: A Summary and New Data. *Plants*, 12, 2798.

902 Vereecken, N.J., Cozzolino, S., & Schiestl, F.P. (2010) Hybrid floral scent novelty drives
903 pollinator shift in sexually deceptive orchids. *BMC Evolutionary Biology*, 10, 103.
904 Wang, R.-H., Yang, Z.-P., Zhang, Z.-C., Comes, H.P., Qi, Z.-C., Li, P., & Fu, C.-X. (2022) Plio-
905 Pleistocene climatic change drives allopatric speciation and population divergence
906 within the *Scrophularia incisa* complex (Scrophulariaceae) of desert and steppe
907 subshrubs in Northwest China. *Frontiers in Plant Science*, 13, 985372.
908 Weir, B.S., & Goudet, J. (2017) A unified characterization of population structure and
909 relatedness. *Genetics*, 206, 2085-2103.

910 **Table 1.** Sampling site, sample size (n), geographic coordinates and elevation, population
911 summary statistics: allelic richness per sampling site (A_R), number of private alleles, expected
912 and observed heterozygosity (H_E and H_O), inbreeding coefficient (G_{IS}) and nucleotide diversity
913 (π)

Sampling site	n	Latitude (°)	Longitude (°)	Elevation (m)	Allelic richness (A_R)	Private alleles	H_O	H_E	G_{IS}	π
1-Guilhaumard	15	43.85	3.19	726	17043.95	0	0.262	0.282	0.071	0.282
2-Lapanouse-de-Cernon	15	43.99	3.09	651	16833.53	2	0.260	0.276	0.059	0.276
3-Saint-Affrique	13	43.98	2.93	585	16669.70	1	0.268	0.269	0.006	0.269
4-Valgañón	15	42.32	-3.08	1023	17167.45	24	0.268	0.291	0.081	0.290
5-Larraona	15	42.78	-2.28	812	17297.27	9	0.269	0.291	0.076	0.290
6-Bercedo	13	43.09	-3.44	774	17102.52	13	0.264	0.283	0.066	0.281
Total	86									

914 **Table 2.** Summary of the main demographic parameters inferred with DILS i) for pooled
915 dataset (*Ophrys aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica*)
916 and ii) for each *Ophrys aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp.
917 *vitorica* population pairs. $N_e 1$, $N_e 2$, N_e anc. correspond to effective population sizes estimated
918 for population 1, population 2 and the ancestral population, t_{split} to the time of split between
919 population 1 and population 2 in number of generations, respectively and M_{12}/M_{21} the
920 migration rates (as $N_e.m$) from population 1 to population 2 and from population 2 to
921 population 1. Parameters have been estimated based on a model using neural network.
922
923

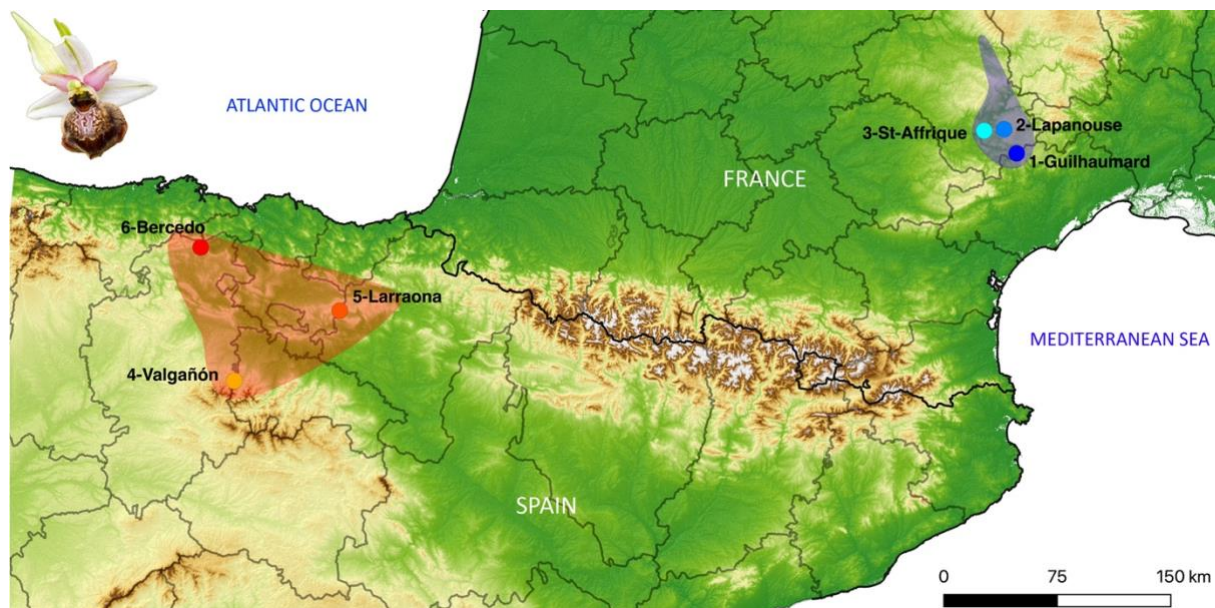
Population 1	Population 2	$N_e 1$	$N_e 2$	N_e anc.	t_{split}	M_{12}/M_{21}
O. a. <i>aveyronensis</i> pool (France)	O. a. <i>vitorica</i> pool (Spain)	6474 (5737-7276)	8425 (7433-9693)	74803 (66810-83607)	1453 (1271-1642)	0.63/0.56 (0.53-0.73/0.44-0.70)
1-Guilhaumard	4-Valgañón	4934 (4290-5662)	4105 (3622-4589)	98474 (92517-105003)	1226 (1079-1383)	0.90/0.74 (0.73-1.07/0.60-0.88)
1-Guilhaumard	5-Larraona	2403 (2089-2751)	2968 (2569-3429)	103206 (97489-109542)	651 (573-724)	0.89/0.64 (0.72-1.06/0.52-0.76)
1-Guilhaumard	6-Bercedo	9824 (8553-11446)	10477 (7866-13078)	97808 (93302-103547)	3029 (2672-3410)	0.73 /0.44 (0.35-0.53/(0.60-0.86)
2-Lapanouse	4-Valgañón	2729 (2506-3035)	2640 (2311-2953)	99913 (94450-105541)	803 (728-872)	0.46/0.67 (0.38-0.55/0.54-0.79)
2-Lapanouse	5-Larraona	4061 (3556-4554)	6049 (5291-6972)	93345 (89784-97211)	1416 (1255-1564)	0.42/1.01 (0.34-0.50/0.71-1.28)
2-Lapanouse	6-Bercedo	5812 (4988-6822)	7719 (6431-9262)	97785 (90460-105961)	2069 (1731-2410)	0.84/0.74 (0.67-1.00/0.58-0.93)
3-St-Affrique	4-Valgañón	9995 (8557-11455)	13510 (11837-15468)	94633 (89865-100052)	3879 (3236-4549)	0.89/0.81 (0.71-1.05/0.62-0.99)
3-St-Affrique	5-Larraona	7270 (6386-8230)	10997 (9603-12667)	90012 (86305-93995)	2891 (2494-3300)	0.82/0.79 (0.66-0.97/0.56-0.98)
3-St-Affrique	6-Bercedo	9562 (8393-10964)	10997 (9632-12459)	99195 (92748-104706)	3682 (3184-4204)	0.95 /0.89 (0.77-1.14-0.69-1.08)

924

925 **Table 3.** List of bioclimatic variables available from WorldClim. From these 19 variables (bio1
 926 to bio19) the ones used for ENM analyses (Person correlation coefficient < 0.75) are indicated
 927 in bold with their relative contribution in the selected ENM model.

Variable name	Denomination	Relative contribution (in %)
bio1	Annual Mean Temperature	
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	8.14
bio3	Isothermality (bio2/bio7) (×100)	5.32
bio4	Temperature Seasonality (standard deviation ×100)	
bio5	Max Temperature of Warmest Month	
bio6	Min Temperature of Coldest Month	
bio7	Temperature Annual Range (bio5-bio6)	
bio8	Mean Temperature of Wettest Quarter	3.21
bio9	Mean Temperature of Driest Quarter	3.01
bio10	Mean Temperature of Warmest Quarter	11.31
bio11	Mean Temperature of Coldest Quarter	
bio12	Annual Precipitation	
bio13	Precipitation of Wettest Month	
bio14	Precipitation of Driest Month	
bio15	Precipitation Seasonality (Coefficient of Variation)	41.65
bio16	Precipitation of Wettest Quarter	
bio17	Precipitation of Driest Quarter	23.34
bio18	Precipitation of Warmest Quarter	
bio19	Precipitation of Coldest Quarter	

928



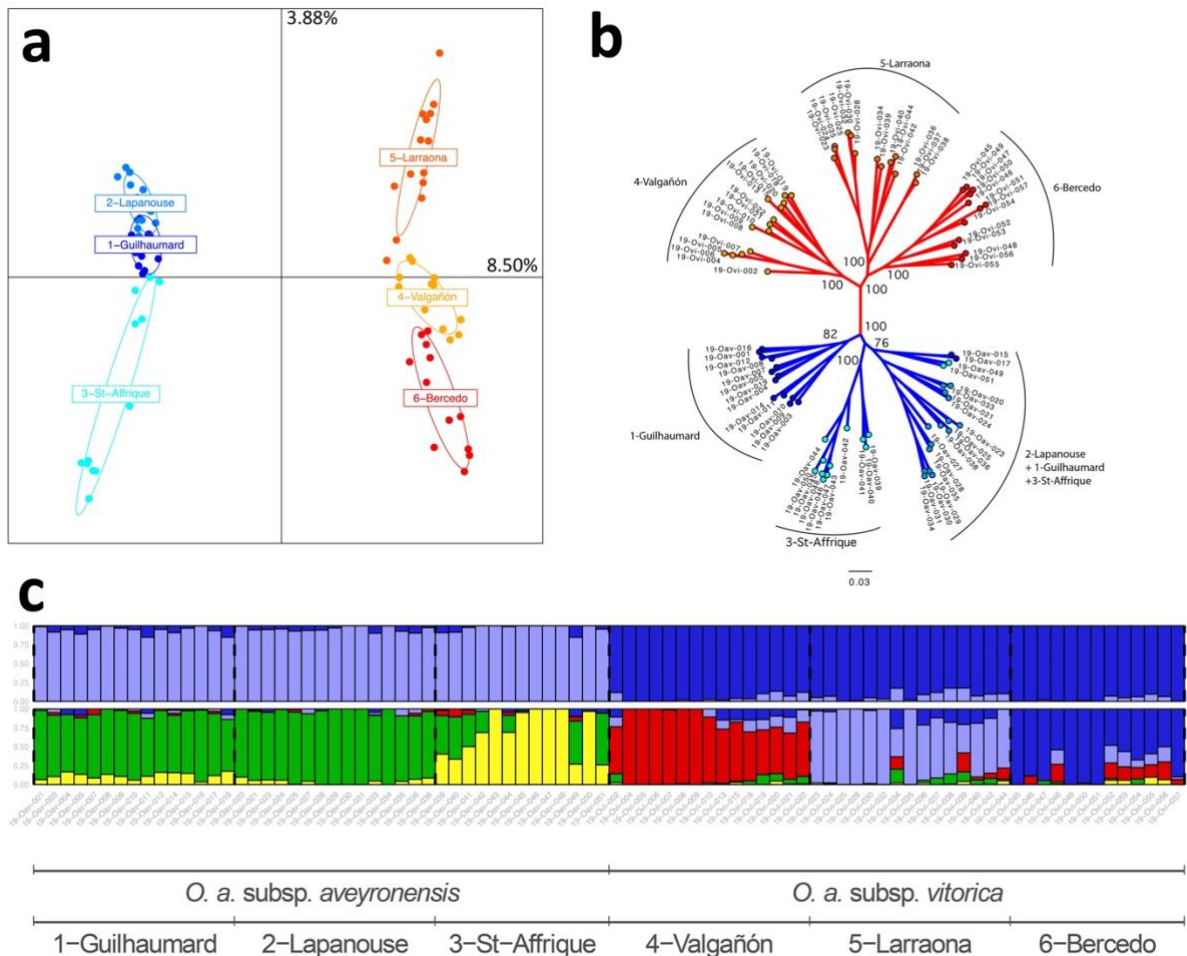
929

930

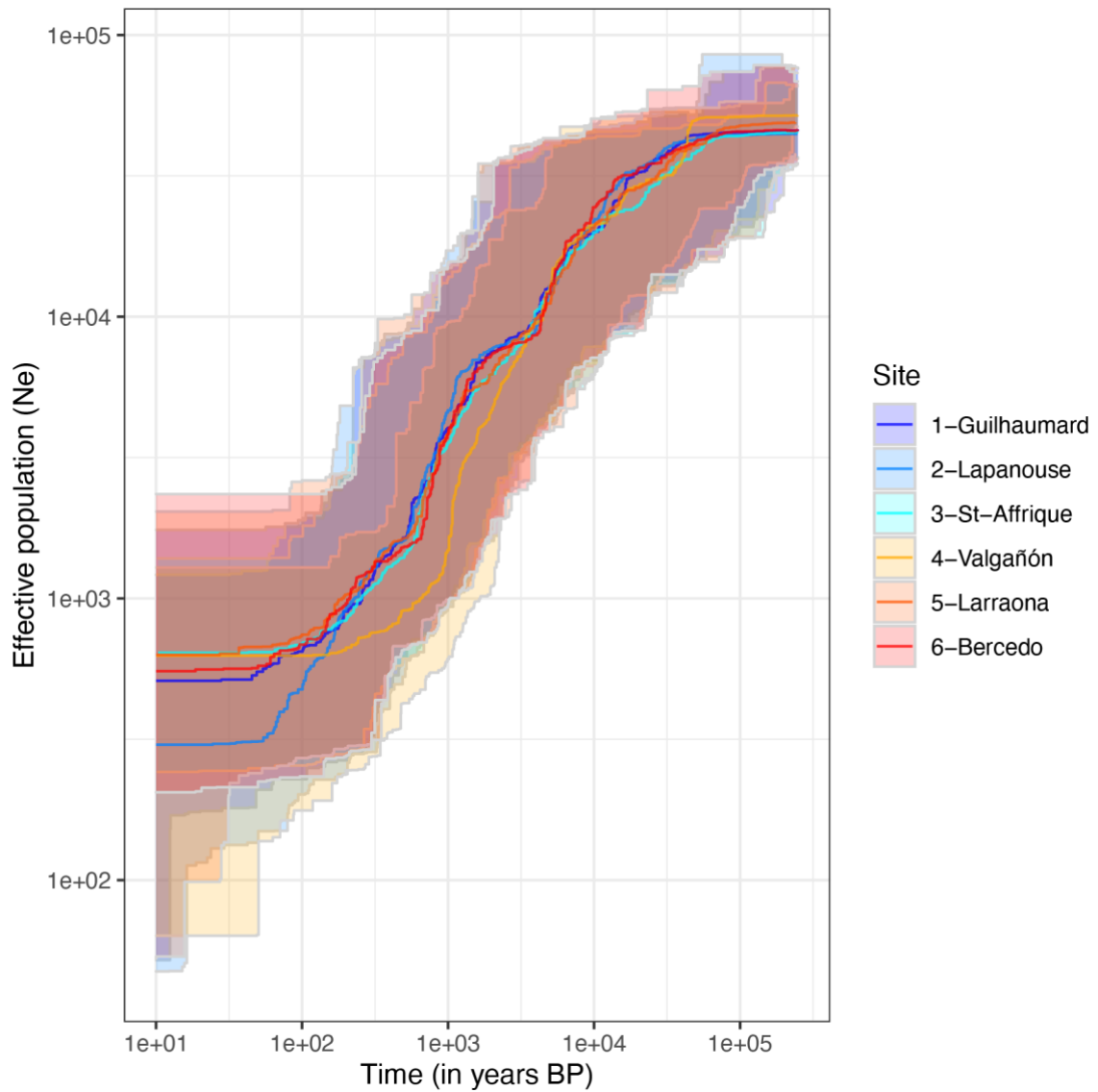
931

Figure 1 Map displaying the 6 sites sampled: 3 for *Ophrys aveyronensis* subsp. *aveyronensis*: 1-Guilhaumard, 2-Lapanouse-de-Cernon, 3-St-Affrique (in blue) and 3 for *O. a.* subsp. *vitorica*:

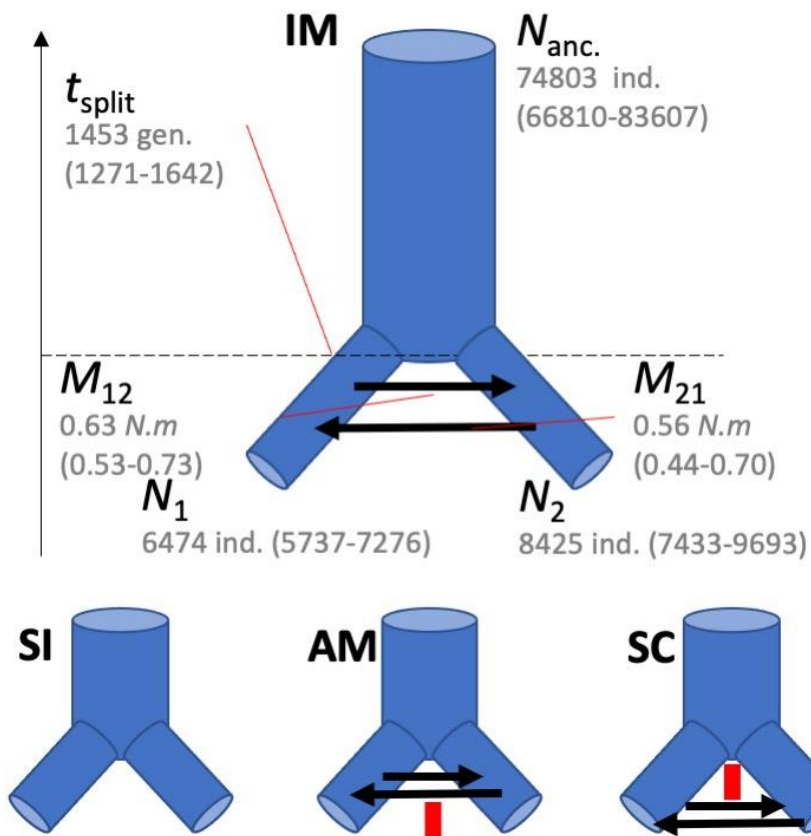
932 4- Valgañón, 5-Larraona, 6-Bercedo (in red). The approximate known geographic distribution
 933 is highlighted for both taxa with the same color code. The photograph (on the top left corner)
 934 illustrates the insect-mimicking flower of a typical *O. aveyronensis* with a marbled labellum.
 935



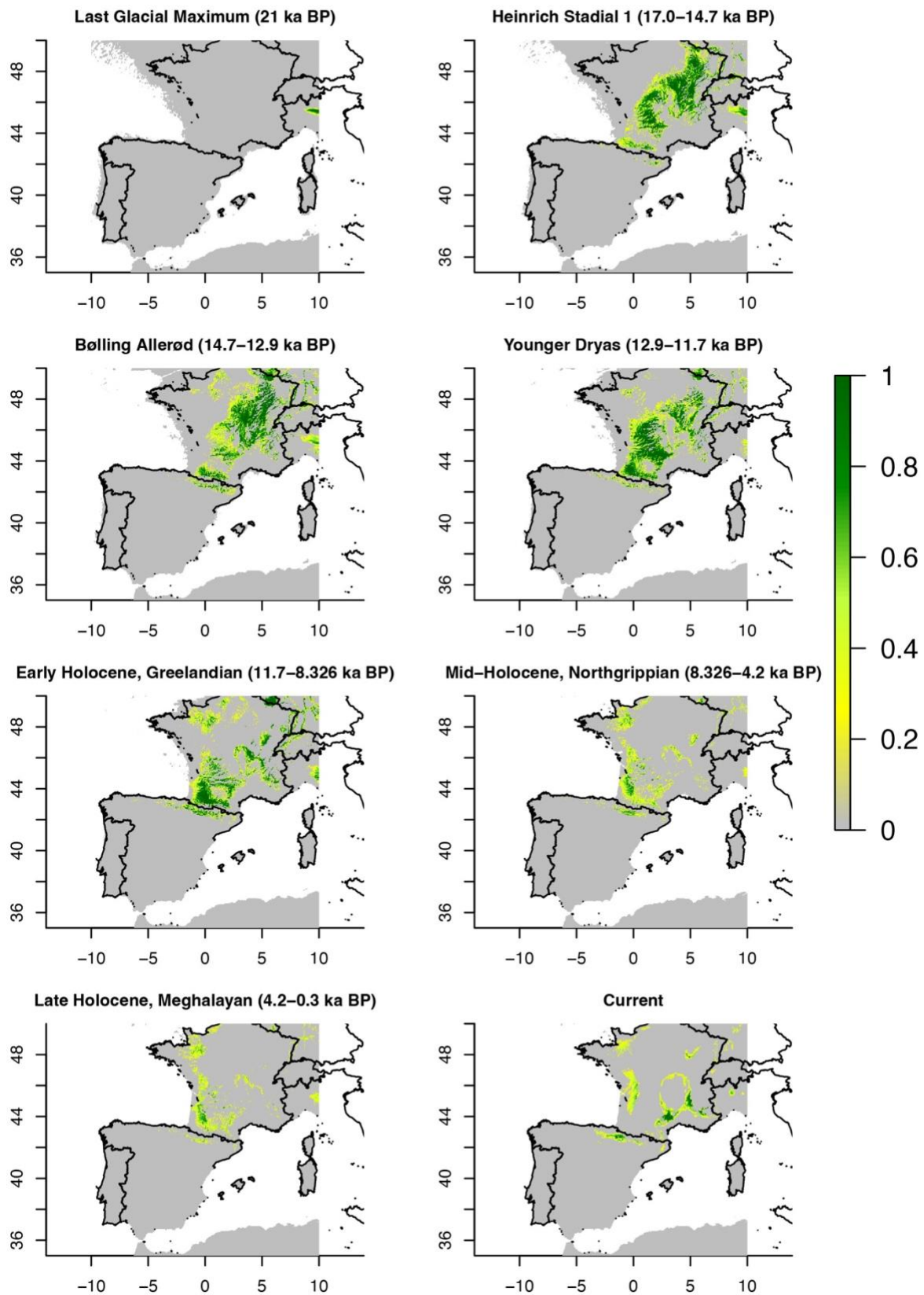
936 **Figure 2** (a) Principal Component Analysis (PCA) displaying the two first axes (PC1 and PC2)
 937 representing 8.50% and 3.88% of the total genetic variance. PCA was computed based on 86
 938 individuals genotyped at 9301 SNPs and colors depict sampling localities. (b) Maximum
 939 Likelihood phylogenetic tree (rooted at midpoint) obtained from the concatenation of the
 940 SNPs (node support for main nodes are given as UltraFast Bootstrap Approximation based on
 941 1000 replicates). (c) Barplots of ancestry coefficients obtained from sNMF for 86 individuals
 942 for $K = 2$ and $K = 5$.
 943
 944



945
 946 **Figure 3** Demographic history of the 6 studied populations of *O. a. subsp. aveyronensis* (shades
 947 of blue) and *O. a. subsp. vitorica* (shades of red) depicting the mean effective population size
 948 N_e (and associated 95 confidence intervals) over time as inferred from STAIRWAYPLOT2,
 949 assuming a mutation rate of 7×10^{-9} and a generation time of 5 years.



950
 951 **Figure 4** Schematic overview of the best demographic scenario (Isolation with Migration, IM)
 952 for pooled dataset (*i.e.* *O. aveyronensis* subsp. *aveyronensis*/*O. a.* subsp. *vitorica*). The median
 953 values (and 2.5-97.5% boundaries of Highest Posterior Distribution probability) are shown for
 954 the parameters inferred with DILS. T_{split} : time of split (given in number of generations), N_a :
 955 ancestral effective population size, N_1 : (current) effective population size for the *O. a.* subsp.
 956 *aveyronensis* population (given in number of individuals), N_2 : (current) effective population
 957 size for the *O. a.* subsp. *vitorica*, M_{12} : migration rate from *O. a.* subsp. *aveyronensis* to *O. a.*
 958 subsp. *vitorica* and M_{21} : migration rate from *O. a.* subsp. *vitorica* to *O. a.* subsp. *aveyronensis*.
 959 Alternative demographic scenarios tested are also illustrated: Strict Isolation (SI), Ancestral
 960 Migration (AM, involving a cessation of gene flow after an initial stage of migration after the
 961 split) and Secondary Contact (SC, involving a secondary contact after an initial stage of
 962 isolation after the split).



963

964

965

Figure 5 Spatial projections of the bioclimatic niches inferred for the *Ophrys aveyronensis* depicting habitat suitability at different time periods from the Last Glacial Maximum (LGM,

966 about 21 kyrs ago) and with projections for the Heinrich Stadial1 (17.0-14.7 kyrs BP), the
967 Bølling-Allerød (14.7-12.9 kyrs BP) the Younger Dryas Stadial (12.9-11.7 kyrs BP) and the
968 Holocene: early-Holocene, Greenlandian (11.7-8.326 kyrs BP), mid-Holocene, Northgrippian
969 (8.326-4.2 kyrs BP) and late-Holocene, Meghalayan (4.2-0.3 kyrs BP) and currently.