Holocene climate change promoted allopatric divergence and explains the current disjunct geographic distribution of the *Ophrys aveyronensis* species complex (Orchidaceae)

Anaïs Gibert[1], Roselyne Buscail[2], Michel Baguette[3,4], Christelle Fraïsse[5], Camille Roux[5], Bertrand Schatz[6] & Joris A. M. Bertrand[1]

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

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

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

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

[5] CNRS, Univ. Lille, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille.

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

Correspondence: anais.gibert@gmail.com; joris.bertrand@univ-perp.fr

Abstract

Species with disjunct geographic distributions provide natural opportunities to investigate recent or incipient allopatric divergence. Although not rare, many of the cases observed result from successful colonization from a historical to a new range through dispersal or human-induced introduction which make the underlying eco-evolutionary processes sometimes difficult to decipher. The *Ophrys aveyronensis* species complex presents a disjunct geographic distribution with two ranges currently separated by 600 km on both sides of the Pyrenees Mountains. To uncover the causes of such intriguing biogeographic pattern, we combined population genomics and Ecological Niche Modelling approaches. Population genomic data show that all the populations studied display similar patterns of genetic diversity and dramatic decrease in effective size. We found significant genetic differentiation between the two ranges of the *O. aveyronensis* species complex. Our results support a very recent divergence (*ca.* 1500 generations ago). Ecological Niche Modelling results further indicate that the disjunct geographic distribution of the *O. aveyronensis* species complex is consistent with a range split of a broad ancestral range, contraction and shifts in opposite directions in response to climate warming during the Holocene. The congruence of both the results of population genomics and ENM approaches demonstrates that continental allopatric divergence is involved in the *Ophrys* adaptive radiation. The *O.*

aveyronensis complex is a promising candidate to study the onset of reproductive isolation immediately following an initial stage of geographic separation.

Keywords Population genomics**,** RAD sequencing, Species Distribution Modelling (SDM), Biogeography, Speciation, Orchids

Introduction

Allopatric speciation (also referred as geographic speciation, or vicariant speciation) occurs when populations of a given species become reproductively isolated from each other because of geographic separation. Aside from being conceptually intuitive, allopatric speciation has been evidenced by theoretical models, laboratory experiments and empirical studies to the point that it is generally regarded as the "default mode" of speciation (see Coyne & Orr, 2004). In allopatric speciation, spatial and reproductive isolation are often assumed to result from geologic-caused topographic changes that form the geographic barrier. In continental systems however, allopatric divergence is prone to experience only transient geographic and reproductive isolation involving one of several phases of secondary contact (e.g. through climate change-induced oscillations during the Quaternary, Hewitt, 2000). Thus, allopatric divergence is a spatio-temporally dynamic process. Aside, not only the fore-mentioned barriers but any kind of fragmentation of species distribution range, including human activities may also initiate and be responsible for evolutionary divergence.

The putative complexity of the speciation process makes difficult to elucidate its ecological and evolutionary causes as current spatial patterns of distribution only provide a snapshot from the speciation continuum (Nosil et al., 2017; Stankowski & Ravinet, 2021). In this regard, species with disjunct geographic distributions are likely to correspond to species whose populations became recently allopatric, *i.e.* with insufficient time to accumulate enough divergence to be considered as distinct species yet. Although such natural example may help to understand the consequences of physical barriers in early stages of evolutionary divergence, disjunct geographic distributions do not necessary result from the fragmentation of an ancestral range. Indeed, dispersal and successful colonization of a new range from the historical one may also lead to the same pattern.

74 For instance, several animal species including terrestrial mammals and mollusks (e.g.
75 Davison et al., 2001; Mascheretti et al., 2003; Grindon & Davison, 2013; Reich et al., 2015
76 and references therein) and plants (e.g. Beatty & Provan, 2013; 2014; Beatty et al., 2015;
77 Fagúndez & Díaz-Tapia, 2023) occur in Northern Iberia and in Ireland (a so called 'Lusitanian'
78 distribution) and display noticeable genetic similarities between these two regions. To
79 explain these disjunct geographic distributions, several authors have considered that all
80 these species may have survived in ice-free pocket (located in southwest Ireland) and in
81 southern European refugia (such as the Iberian Peninsula) during the Last Glacial Maximum
82 (LGM, 21000 years *Before Present*) or may have colonized Ireland thanks to land-bridges,
83 that disappeared ever since. However, molecular data have more recently supported that
84 most of the Irish animal populations have been introduced in the Mesolithic times (more
85 than 8000 years ago) through human activities whereas plants could have been more prone
86 to accomplish post-glacial recolonization of this region. In such examples, the combination of
87 genetic data with sufficient resolution and/or Environmental Niche Modelling (ENM) have
88 been crucial to compare the different scenarios and better understand the biogeography of
89 both species and regions.

90 The species of the genus *Ophrys* have long been known for their unusual pollination
91 syndrome (called 'sexual swindling') that involves insect-like flowers mimicking their female
92 pollinator to lure conspecific males and ensure pollination but are now also increasingly
93 considered as a promising biological model to investigate adaptive radiations. The degree of
94 plant-pollinator specialization (Joffard et al., 2019) required to ensure the success of the
95 *Ophrys* pollination strategy was shown to be the driver of the impressive rate of
96 diversification of the genus within the Mediterranean basin, and to a lesser extent in other
97 parts of Western Europe (see Baguette et al., 2020, for a recent review). If *Ophrys* are

98 notorious for their relatively high number of likely cases of speciation with gene flow, the
99 number of island endemic taxa suggests that geographic isolation induced by insularity is
100 also a key factor to explain *Ophrys* species richness observable nowadays (see Bertrand et
101 al., 2021a).

102 On continents, assessing the role of allopatry in *Ophrys* diversification is challenging,
103 in particular because clear evidences of taxa displaying subranges without being connected
104 by gene flow are lacking. Nevertheless, one species, *Ophrys aveyronensis*, displays an
105 intriguing case of disjunct geographic distribution. This species was initially described as
106 endemic to a very restricted area in the Grands Causses region (Southern France).
107 Observations of morphologically very similar populations were reported later from Northern
108 Spain and described as a distinct species (*Ophrys vitorica*, Kreutz, 2007). *Ophrys aveyronensis*
109 is now rather considered as a species complex consisting of two subspecies (*O. aveyronensis*
110 subsp. *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica*) that share, at least a mutual
111 insect pollinator (*i.e.* the solitary bee *Andrena hattorfiana*, Paulus & Gack, 1999; Paulus,
112 2017; Benito Ayuso, 2019 (but see Gibert et al., 2022 reporting subtle differences in
113 morphometry and coloration)). This taxon is thus present in two rather restricted ranges
114 separated by a gap of about 600 km (and Pyrenees Mountains), in the Grands Causse region
115 (France) and in Northern Spain.

116 The formation of the Pyrenees starting during the Eocene (from 45 Myrs ago) cannot
117 explain by itself the split of the ancestral range of the species as all phylogenetic hypotheses
118 published so far conclude that the *Ophrys* genus emerged no more than 5-6 Myrs ago and
119 that the *Sphegodes* clade, to which the *O. aveyronensis* complex belongs may have
120 diversified over the last million year (Breitkopf et al., 2015, see also Inda et al., 2012). This
121 does not rule out allopatry as factor to explain current disjunct geographic distribution but

122 suggests that geographic separation would have occurred more recently. According to this
123 'vicariance hypothesis', the current disjunct geographic distribution may correspond to two
124 relict pockets of an ancestral range that got fragmented following a scenario according to
125 which populations from Northern and Southern ranges diverge in allopatry. In this case, the
126 degree of genetic differentiation between Northern and Southern populations is expected to
127 be significant and positively correlated with the duration of reproductive isolation induced
128 by geographic separation. Namely, genetic differentiation is expected to be relatively higher
129 between Northern and Southern ranges than between populations within subranges.
130 Alternatively, we also envisage a 'biological introduction hypothesis': a colonization from
131 one historical range to the other as a result of human activities, or not. The notoriously small
132 *Ophrys* seeds are likely to be wind-dispersed over hundreds of kilometers (*i.e.* thus being
133 able to travel across most physical barriers including mountains and vast open water areas)
134 and may have found in the newly colonized area, suitable conditions for their germination,
135 development and reproduction. Although 'orchid dusty seed rains' may allow to generate
136 founder populations with non negligeable genetic diversity, such newly established
137 populations typically consist of a subset of the genetic pool of the source population (*i.e.* a
138 relatively little divergent population with lower effective population size and genetic
139 diversity).

140 In this study, we used a population genomic approach (*i.e.* through a double digest
141 RAD-seq like protocol called 'nGBS') to first investigate patterns of genetic diversity and
142 differentiation among *Ophrys aveyronensis* populations from both part of its disjunct
143 geographic range. We then couple demographic inference analyses and Ecological Niche
144 Modelling approaches to test whether the current disjunct geographic distribution of this
145 species complex is rather consistent with either allopatry or recent introduction. Specifically,

146 we aimed at determining how recent the divergence between Northern and Southern
147 populations of the *O. aveyronensis* complex may be and whether or not this disjunct
148 geographic distribution may have originated with a colonization event from an ancestral
149 range to a new one.
150

151 Materials and Methods

152 Sampling and study sites

153 We used a minimally destructive sampling method to collect a total of 86 plant tissue
154 samples from six localities that span the entire geographic range of the *O. aveyronensis*
155 species complex: three in Southern France (Grands Causses region): 1-Guilhaumard, 2-
156 Lapanouse-de-Cernon and 3-Saint-Affrique, and three in Northern Spain: 4-Valgañón, 5-
157 Larraona and 6-Bercedo, in June 2019, Fig. 1, see also Gibert et al., 2022). At each sampling
158 site, we collected 6 to 10, 0.6 mm diameter leaf punches from 13 to 15 individuals which
159 were stored in 100% ethanol until DNA extraction (see Table 1 and Supplementary Table S1).
160 As *Ophrys aveyronensis* is a nationally protected plant species in France, sampling was done
161 under permit 'Arrêté préfectoral n°2019-s-16' issued by the 'Direction Régionale de
162 l'Environnement de l'Aménagement et du Logement (DREAL)' from the 'Région Occitanie',
163 on 07-May-2019). *Ophrys aveyronensis* is not legally protected in Spain.
164

165 DNA extraction, library preparation and genotyping

166 Genomic DNA extraction as well as library preparation and genotyping were subcontracted
167 to LGC Genomics GmbH (Berlin, Germany). A fractional genome sequencing strategy called
168 normalized Genotyping-by-Sequencing (nGBS) was used to subsample the *Ophrys*
169 *aveyronensis* genome for which we estimated a relatively high size of ~5-6 Gbp based on

170 flow cytometry techniques (Bertrand, unpublished data). The nGBS protocol makes use of a
171 combination of two restriction enzymes (PstI and ApeKI, in our case) to produce a
172 reproducible set of fragments across samples/individuals. It also comprises a normalization
173 step whose aim is to avoid further sequencing of highly repetitive regions. Illumina
174 technology was then used to aim at obtaining a minimum of 1.5 million reads per sample
175 (with 2 x 150 bp read length) for each of the 86 individually barcoded libraries.
176
177 **Population genomic dataset processing**
178 We used different scripts included in *Stack*s v.2.60 (Catchen et al., 2011; 2013) to build loci
179 from Illumina reads, *de novo* (*i.e.* without aligning reads to a reference genome). We first
180 used *process_radtags* to demultiplex and clean reads, then *denovo_map.pl* to build loci
181 within individuals, create a catalog and match all samples against it and finally *populations* to
182 further filter the SNPs obtained at the population level and compute basic population
183 genetic statistics. Before running the pipeline on the complete dataset, we first optimized
184 several key parameters: *-m* (the minimum number of identical raw reads required to form a
185 putative allele), *-M* (the number of mismatches allowed between alleles to form a locus) and
186 *-n* (the number of mismatches allowed between loci during construction of the catalog) by
187 running it on a subset of 12 individuals representative of the whole dataset (*i.e.* geographic
188 origin and coverage). As recommended by several authors, we varied *M* and *n* from 1 to 9
189 (fixing *M* = *n*) while keeping *m* = 3 (see Paris et al., 2017). The combination *-m* 3, *-M* 5 and *-n*
190 5 was found to be the most suitable to maximize the number of SNPs, assembled and
191 polymorphic loci in our case and was used to run *Stacks* on the whole data set (see
192 Supplementary Table S2). Then, we ran the *populations* script included in *Stacks*, from which
193 only SNPs that were genotyped in at least 80% of the individuals per population (*-r* 80) and

194 found in all six sampling sites were kept (-*p* 6), while SNPs exhibiting estimates of observed
195 heterozygosity greater than 70% (--*max-obs-het* 0.7) were filtered out to reduce the risk of
196 including remaining paralogs. We also discarded sites whose minor allele frequency was
197 lower than 5% (--*min-maf* 0.05). The genotype matrix of the remaining 9301 SNPs was then
198 exported in several formats for downstream analyses.

199 The following population genomic analyses aimed to i) infer patterns of genetic
200 diversity and divergence between subspecies/countries and ii) infer the demographic
201 scenario (and associated parameters) likely to best explain the current biogeographic
202 pattern observed in the *O. aveyronensis* species complex.

203

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

205 To get an overview of the overall genetic diversity and differentiation among individuals and
206 populations, we first performed a Principal Component Analysis (PCA) based on a matrix of
207 86 individuals (as rows) and 9301 SNPs (as columns) coded as a *genlight* object with the R-
208 package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011). The number of private alleles
209 and the average nucleotide diversity (π) were directly available from the output of the
210 *populations* script from *Stacks*. We used Genodive v.3.04 (Meirmans, 2020) to compute
211 expected and observed heterozygosity (H_E and H_O, respectively) as well as to evaluate the
212 deviation from panmixia by computing G_{IS}. Allelic richness (A_R) was computed with the R-
213 package *hierfstat* (Weir & Goudet, 2017). Overall and pairwise genetic differentiation was
214 also assessed based on *G*-statistics (G_{ST}, G'_{ST}, G''_{ST}) as well as Jost's *D* also implemented in
215 Genodive. An Analysis of Molecular Variance (AMOVA) allowed to assess the proportion of
216 genetic variance at different hierarchical levels, especially between sampling sites among

217 subspecies/country (F_{SC}) and between subspecies/country (F_{CT}). The statistical significance of
218 the obtained values was estimated based on 10000 permutations.

219 To further investigate population structure and characterize putative
220 migration/admixture event, we used sNMF (Frichot et al., 2014) as implemented in the R-
221 package *LEA* (Frichot & François, 2015) to estimate individual ancestry coefficients based on
222 sparse non-negative matrix factorization algorithms. The number of ancestral populations
223 (genetic clusters) was varied from K = 1 to 10, and analyses were run with 10 replicates at
224 each value of K. We followed the cross-entropy criterion to determine the optimal number
225 of clusters and the ancestry proportions matrices (Q-matrices) obtained were plotted with
226 the R-package *pophelper* (Francis, 2017).

227 Finally, we performed Maximum-Likelihood phylogenetic tree reconstruction with IQ-
228 Tree v.2.2.6 (Minh *et al*., 2020) based on the concatenated SNP matrix with 1000 replicates
229 (-B 1000) of Ultrafast Bootstrap Approximation (UFBoot) to assess nodes support. The
230 substitution model that best fitted the data was selected with *ModelFinder Plus*
231 (Kalyaanamoorthy *et al*., 2017) and Ascertainment bias correction (-MFP+ASC) as
232 recommended for SNP data.

233

234 **Demographic and evolutionary history reconstruction**

235 We reconstructed the past demographic history of each of the six populations, separately,
236 using STAIRWAYPLOT2 (Liu & Fu, 2015; 2020). This program relies on the Site Frequency
237 Spectrum (SFS) to infer the temporal dynamics of effective population size (N_e) and has the
238 advantage of not requiring whole-genome sequence data or reference genome information.
239 First, we used the *easySFS.py* script (https://github.com/isacovercast/easySFS) to compute
240 the folded SFS of each population. To maximize the number of segregating sites, the

241 population was downsampled to 24, 24, 22, 24, 24 and 22 individuals for 1-Guilhaumard, 2-
242 Lapanouse, 3-St-Affrique, 4-Valgañón, 5-Larraona and 6-Bercedo haploid genomes,
243 respectively. We then ran STAIRWAYPLOT2 that fits a flexible multi-epoch demographic
244 model that estimates N_e at different time periods using the SFS to infer estimated N_e
245 fluctuation coinciding with coalescent events. We assumed a generation time of 5 years for
246 *Ophrys* and a mutation rate of 7×10^{-9} per site per generation for angiosperms autosomal
247 markers (following Krasovec et al., 2018) and performed 200 bootstrap replicates to
248 estimate 95% confidence intervals.

249 We then used the Approximate Bayesian Computation (ABC) framework
250 implemented in DILS (Fraïsse et al., 2021) to infer the best demo-genomic scenario and
251 associated parameters of the evolutionary history between *O. a.* subsp. *aveyronensis* and *O.*
252 *a.* subsp. *vitorica*. We considered two types of datasets i) a pooled dataset grouping the 3
253 populations *O. a.* subsp. *aveyronensis* together and the 3 populations of *O. a.* subsp. *vitorica*
254 together and ii) paired datasets consisting of all possible *O. a.* subsp. *aveyronensis* and *O. a.*
255 subsp. *vitorica* population pairs ($3^2 = 9$ pairwise comparisons). The two-population model of
256 DILS compares different scenarios in a hierarchical manner. First, current isolation models,
257 *i.e.* Strict Isolation (SI) versus Ancient Migration (AM) are compared to ongoing migration
258 models, *i.e.* Isolation Migration (IM) versus Secondary Contact (SC). Then, for the best
259 demographic model, DILS compares genomic models to test for background selection by
260 assuming variation in effective population size (N_e) across sites (*i.e.* hetero versus homo *N*e
261 models are compared). For migration models (IM and SC), DILS also tests for selection
262 against migrants by assuming variation in migration rates *N.m* across sites (*i.e.* hetero versus
263 homo *N.m* models are compared). Depending on the best selected model, DILS provides
264 estimates and confident intervals for parameters such as the time of split (T_{split}), the

265 migration rate (*N.m*), current and ancestral effective population sizes ($Ne_{current}$ and Ne_{past})
266 and whenever relevant, the time of ancient migration (T_{AM}) or the time of secondary contact
267 (T_{SC}). We formatted the input file with a custom Perl script and used the online facility
268 (https://www.france-bioinformatique.fr/cluster-ifb-core/) to run computations with the
269 following settings and priors: *population_growth*: constant, *modeBarrier*: bimodal (*i.e.* there
270 is a class of species barrier loci with *N.m* = 0 and a class of loci migrating at the background
271 rate *N.m*) for model type to test), *max_N_tolerated:* 0.1*, LMin* = 30, *nMin* = 10 (20 for the
272 pooled dataset), *mu* = 7.10^{-9}, *rho_over_theta*: 0.1, *N*_min = 100, *N_max*: 200000, *Tsplit_min*:
273 100, *Tsplit_max*: 25000, *M_min*: 0.4 and *M_max*: 4. To quantify the fit of the best model to
274 the data we conducted a goodness-of-fit test using 2000 simulations performed under the
275 best model based on the estimated parameter values.
276
277 **Past and current Species Distribution Modelling**
278 We used an Ecological Niche Modelling (or Environmental Niche Modelling, ENM or Species
279 Distribution Modelling, SDM) framework similar to the one we followed and described in
280 detail in Salvado et al., (2022) to infer the spatio-temporal dynamics of suitable habitat for
281 the *O. aveyronensis* species complex since the Last Glacial Maximum (LGM, 21 kyrs BP). We
282 then compared the results with those obtained from population genomic analyses. To model
283 and spatially project the current bioclimatic niche, we first downloaded 19 bioclimatic layers
284 at a resolution of 2.5 arc-minutes (~5 km^2) available from WorldClim 2 (Fick & Hijmans,
285 2017). Current climate data correspond to time averaged variables over the period 1970-
286 2000. To spatially project the inferred niche onto past bioclimatic conditions data, we then
287 downloaded paleoclimate data either from WorldClim v1.4 (Hijmans et al., 2005) and from
288 PaleoClim (www.paleoclim.org).

289 Paleoclimate data from WorldClim are available for both the LGM (21 kyrs ago) and
290 the Mid-Holocene (about 6 kyrs ago), downscaled at a resolution of 2.5 arc-minutes for at
291 least three commonly used General Circulation Models (GCMs) proposed by the National
292 Center for Atmospheric Research (NCAR) though the Community Climate System Model
293 (CCSM4), the Japan Agency for Marine-Earth Science and Technology, the Center for Climate
294 System Research of the University of Tokyo, and the National Institute for Environmental
295 Studies (MIROC-ESM) and the Max Plant Institute for Meteorology Earth System Model
296 (MPI-ESM-P). Only CCSM data are available from PaleoClim but over different time periods
297 over the Pleistocene (at a resolution of 2.5 arc-minutes): the LGM (as from CHELSA, Karger et
298 al., 2021), the Heinrich Stadial1 (17.0-14.7 kyrs BP), the Bølling-Allerød (14.7-12.9 kyrs BP)
299 the Younger Dryas Stadial (12.9-11.7 kyrs BP) and the Holocene: early-Holocene,
300 Greenlandian (11.7-8.326 kyrs BP), mid-Holocene, Northgrippian (8.326-4.2 ka BP) and late-
301 Holocene, Meghalayan (4.2-0.3 kyrs BP) (as from Fordham et al., 2017).

302 The *O. aveyronensis* occurrences dataset consists of a set of field observations as well
303 as from records downloaded from the gbif (www.gbif.org) and the iNaturalist
304 (www.inaturalist.org) databases before manual curation. The Ecological Niche Modelling
305 analyses were conducted thanks to the R-package *ENMwizard v.0.3.7* (Heming et al., 2018,
306 see also Bagley et al. 2020) which is a convenient wrapper of several tools. We first spatially
307 filtered occurrences to keep only those that were at least 5 km away from each other using
308 the R-package *spThin* (Aiello-Lammens et al., 2015). The calibration area for the models was
309 created as a buffer of 0.5° around the minimum convex polygon encompassing all
310 occurrences. From the 19 bioclimatic variables, we selected the less correlated ones (Person
311 correlation coefficient < 0.75) thanks to the R-package *caret* (Kuhn, 2019) and kept 7

312 variables: bio2, bio3, bio8, bio9, bio10 and bio17 (see details in Supplementary Appendix S5)
313 for further analyses.

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

326

327 ## Results

328 **Population genomic dataset**

329 We obtained a total of 150 694 350 of read pairs across the 86 individuals (1 382 768 - 7 370
330 366 of raw reads per ind., mean 3 504 519, SD = 1 282011.52) of which 290 431 210
331 barcoded reads (*i.e.* 96.36%) were retained, 10 819 014 (3.6%) were removed because RAD
332 cutsite was not found in the sequence and 138 476 reads (< 1%) were removed because of
333 low quality. We genotyped a total of 859 035 *loci* (composed of 135 255 870 sites) including
334 368 712 variant sites (SNPs) with *Stacks*. Average read depth ranged from 24.1X to 81.6X
335 (mean 48.7X, SD = 12.4X) based on the combination of parameters we used. After filtering

336 the data with *populations*, we finally kept 4302 *loci* from which we retained 9301 variant
337 sites (SNPs).
338
339 **Genetic diversity, population differentiation and divergence**
340 The PCA biplot shows that individuals are arranged by subspecies/country of origin along
341 PC1 (which explains 8.50% of the total genetic variance) and by sampling site along PC2
342 (which explains 3.88% of the total genetic variance) even though the sampling sites of 1-
343 Guilhaumard and 2-Lapanouse-de-Cernon slightly overlap (Fig. 2A). Based on PC2, intra-
344 population genetic variation seems minimal for 1-Guilhaumard and 2-Lapanouse-de-Cernon
345 and maximal for 3-Saint-Affrique and 5-Larraona. Basic population genetics statistics are
346 indicated in Table 1. All populations displayed weak but significant deviation from panmixia
347 with a mean G_{IS} = 0.060 (p < 0.01, 95% CI: 0.056-0.064) and population G_{IS} values ranging
348 from 0.006 in 3-Saint-Affrique to 0.081 in 4-Valgañón (all p < 0.01). Average nucleotide
349 diversity (π) varies from 0.269 in 3-Saint-Affrique to 0.290 in 4-Valgañón and 5-Larraona, and
350 allelic richness (A_R) from 16696.70 (in 3-Saint-Affrique) to 17297.27 (in 5-Larraona). Based on
351 the filtering procedure we followed, the number of private alleles was relatively low but
352 consistently higher for *O. a.* subsp. *vitorica* (*i.e.* from 9 to 23) compared to *O. a.* subsp.
353 *aveyronensis* (*i.e.* from 0 to 2).
354 Overall *G*-statistics show that the whole set of populations is genetically structured:
355 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}
356 = 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).
357 The AMOVA confirms that although a major part of the genetic variance is found within
358 individual (81.8%, F_{IT} = 0.182, 95% CI: 0.177-0.186), significant proportions of the total
359 genetic variance are found among populations within subspecies/country (6.8%, F_{SC} = 0.072,

360 95% CI: 0.070-0.074) and between subspecies/country (6.1%, F_{CT} = 0.061, F_{CT} = 0.058-0.064).
361 Pairwise differentiation between sampling sites varies from F_{ST} = 0.049 to 0.082 within
362 subspecies/country and reaches F_{ST} = 0.111 to 0.147 between subspecies/country of origin
363 (all $p < 0.01$; see Supplementary Appendix S3 for details).

364 The clustering analysis carried out with sNMF suggests an optimal number of genetic
365 clusters of *K*=5 based on the cross-entropy criterion (Fig. 2B). We plotted the *Q*-matrices
366 (that contain individual admixture coefficients) corresponding to the best replicate runs for
367 *K*=2 (to investigate whether the program could distinguish the two subspecies) and *K*=5 (Fig.
368 2C). For *K*=2, the individuals of *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* were
369 unambiguously assigned to two distinct clusters. For the most likely combination of *K*=5, we
370 observe one cluster per sampling site except for the geographically close localities of 1-
371 Guilhaumard and 2-Lapanouse-de-Cernon that were grouped together in a unique cluster. In
372 the locality of 3-Saint-Affrique 4 out of 13 individuals (19-Oav-039, 19-Oav-040, 19-Oav-049
373 and 19-Oav-051) display >50% of ancestry proportion that are associated to the cluster
374 associated with 1-Guilhaumard and 2-Lapanouse-de-Cernon.

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

380

381 **Inference of past demographic history**

382 Reconstruction of historical dynamics of effective population size using STAIRWAYPLOT2
383 shows congruent trends consisting of progressive >95% decline in estimated N_e for all 6

384 populations, since the LGM, or at least, over the last 25000 years (see Fig. 3). Current
385 effective population sizes were estimated to vary between 302 (2-Lapanouse-de-Cernon)
386 and 642 individuals (3-St-Affrique) with this method. Despite the current geographical
387 discontinuity between *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* populations, we
388 attempted to statistically compare demographic scenarios with and without ongoing
389 migration using DILS. The Posterior Probabilities associated with the models of ongoing
390 migration (*i.e.* IM and SC) relative to models without (*i.e.* SI and AM) was 0.59 based on the
391 pooled dataset and ranged from 0.55 to 0.66 based on the population pairs analyses. These
392 probabilities are low and lead to ambiguous model choices. Therefore, we cannot draw
393 strong conclusions from model comparisons with DILS (see Table 2 and Appendix S4). This
394 can happen when the real history is an intermediate model, or because the separation time
395 between *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* is recent compared to the
396 current population sizes. The median of the parameters estimated based on posterior
397 distributions made under the most probable model (*i.e. Isolation with continuous Migration*,
398 IM) indicate very low migration rates ($N.m$ = 0.60 for the pooled dataset, and it ranges from
399 0.56 to 0.92 based on population pairs). With the pooled dataset, we inferred a subdivision
400 of the ancestral population of about 75 000 individuals based on two pools (90 000 to 103
401 200 based on population pairs) into two daughter populations about 1500 generations ago
402 (650 to 3880 based on population pairs), which convert into 4500 years assuming a
403 generation time of 3 years, 7500 years assuming a generation time of 5 years and 10500
404 years assuming a generation time of 7 years, an globally less based on population pairs
405 estimates. This subdivision is indeed considered recent as it is about five times smaller than
406 current population sizes. This suggests that most coalescence events are taking place in the
407 ancestral population and not in the respective daughter populations. In other words, the

408 amount of shared polymorphism is still relatively high because of insufficient time for drift to
409 differentially fix/eliminate alleles within the two subranges (*i.e.* Incomplete Lineage Sorting).
410 Similar to what was inferred with STAIWAYPLOT2, our analyses also point to a decrease of
411 effective sizes compared to the ancestral population (*i.e.* current N_e have values
412 corresponding to about 10% of $N_{anc.}$. The goodness-of-fit test indicates that the best
413 estimated scenario reproduces the observed data set with high fidelity (see Supplementary
414 Appendix S4).
415
416 **Current and past Species Distribution Modelling**
417 ENMeval analyses identified RM = 1 and a mixture of Product (P), Quadratic (Q) and Hinge
418 (H) feature classes as the best-performing parameters for calibrating the final ENM. These
419 parameters yielded a single "best" candidate model with the lowest AICc score (925.02) and
420 an AICc weight of ~0.23. This model had mean omission rates of 0.165 (for the 10th
421 percentile) and a mean test AUC of 0.834 (see variable contributions following this model in
422 Table 3 and Supplementary Appendix S5, for additional details). The spatial projection of the
423 current bioclimatic inferred for the *O. aveyronensis* species complex corresponds to its
424 actual geographic distribution but show that some additional areas may be climatically
425 suitable for this orchid (Fig. 5). Spatial projections considering paleoclimatic conditions show
426 that the extent of suitable habitat for the *O. aveyronensis* species complex may have
427 reached its apex during the Younger Dryas cold period (12.9-11.7 kyrs BP) before starting to
428 decrease across the Holocene. The LGM conditions (21 kyrs BP) were not suitable for the
429 establishment of viable populations in the studied area but important suitable habitats may
430 have existed during the Heinrich Stadial 1 event (17-14.7 kyrs BP) and the Bøllering-Allerød
431 warming (14.7 – 12.9 kyrs BP). Altogether, these results are consistent with the possible

existence of a unique ancestral area, centered on South-Western France that could have experienced a split and a combination of reduction and shift in two distinct directions during the Holocene. The results available from WorldClim data do not contradict these conclusions but show significant differences across GCMs and even for a given GCM but downscaled based on different interpolation algorithms: namely, WorldClim instead of CHELSA (see Supplementary Appendix S5).

Discussion

Some *Ophrys* orchid lineages exhibit among the highest rate of speciation reported so far (Breitkopf et al., 2015) and are being increasingly mentioned as biological model of interest to investigate ecological speciation and adaptive radiations (e.g. Nosil et al., 2017; Baguette et al., 2020; Nürk et al., 2020). The selection pressures exerted by pollinators and the high degree of specificity necessary to ensure the success of the sexual swindling strategy is likely to explain the overwhelming importance of ecological and sometimes sympatric speciation in *Ophrys* (see Baguette et al., 2020). This does not preclude for a significant role of allopatry in *Ophrys* diversification, even in non-insular systems (see Bertrand et al., 2021a). Our study, is to our knowledge, the first to thoroughly characterize a scenario of very recent allopatric divergence in a continental *Ophrys* species. The *Ophrys aveyronensis* species complex therefore provides a promising opportunity to investigate the relative contribution of biogeographic and ecological factors in shaping early stages of evolutionary divergence in rapidly diversifying groups of organisms.

Patterns of genetic diversity and differentiation based on population genomics

455 Although Northern and Southern individuals of *Ophrys aveyronensis* only present subtle
456 differences based on phenotypic data (Gibert et al., 2022), the French and Iberian
457 counterparts of this species complex can be unambiguously teased apart based on
458 population genomics data. Both the PCA, the *F*-statistics (and/or their analogues), the
459 clustering analysis and the phylogenetic tree confirm that the two taxa form distinct genetic
460 entities, themselves exhibiting a lower genetic substructure between sampling sites within
461 subranges. Altogether, these results show that each sampled population of *O. aveyronensis*
462 is more genetically differentiated from populations from the other region than from
463 populations of its own region. This pattern *a priori* rules out a recent colonization and
464 expansion from one range to the other from a unique source population. Patterns of genetic
465 diversity corroborate this conclusion. We found that all the populations sampled displayed
466 similar levels of allelic richness, heterozygosity, departure from panmixia and number of
467 private alleles. There is therefore no evidence for any of these six populations to appear as a
468 subsample of the genetic diversity of a putative source population elsewhere. Finally, the
469 inference of the temporal dynamics of effective population sizes suggests that each sampled
470 population may be considered as kind of replicate of each other. All the populations
471 displayed consistent decreases in effective population size over the last tens of thousands of
472 years which suggests that they shared the same demographic history. In summary,
473 population genomics results are fully consistent with a scenario of range split, and
474 subsequent contractions of two subranges experiencing restricted gene flow.

475 Delforge (2016) hypothesized that the Iberian populations may result from a hybrid
476 swarm (without mentioning the putative parental species) undergoing spatial and
477 demographic expansion. None of our results support the hypothesis according to which
478 Northern and Southern populations of *Ophrys aveyronensis* may result from parallel

479 convergent evolution from distinct regional hybrid swarms. However, the demographic
480 scenarios we tested remain relatively simple and the Bayesian Posterior Probabilities we
481 obtained with DILS show that the approach implemented in this program could not strongly
482 support Migration over a Isolation models. Events that are known to deceives demographic
483 inferences such as 'ghost introgression' from an unsampled lineage have a non-negligible
484 probability in *Ophrys*. Investigating whether the *Ophrys aveyronensis* complex has or not a
485 hybrid origin and at least whether the two taxa have been introgressed, and to which extent,
486 by other *Ophrys* taxa will thus require the genomic analyses of additional species in a future
487 study, including data with higher marker density (transcriptomes or whole-genome
488 resequencing data).
489

490 **On the causes of the disjunct geographic distribution of the *O. aveyronensis* complex**

491 The Ecological Niche Modelling results show that the LGM conditions were not suitable for
492 the establishment of viable populations of the *O. aveyronensis* species complex. This
493 suggests that either this species complex diversified from an ancestral *Ophrys* lineage in the
494 study area after this period, or that the populations of the complex occupied a relatively
495 remote glacial refugia during this period (the ENM results suggest that the Italian peninsula
496 could have provided such refugia, see Fig 5). Over the last 21000 years, the extent of the
497 habitat suitable for *O. aveyronensis* was rather large until the Holocene (about 12000 years
498 ago) (Fig. 5) where its area started to dramatically decrease while keeping an area of
499 relatively high suitability centered on the Southwest of France. The split and, to a lesser
500 extent, the shift Northward and Southward may have occurred during the Late Holocene
501 (starting about 4000 years ago) (Fig. 5). This timing is on the same order of magnitude than
502 the time of split we inferred based on population genomic data and is consistent with a

503 scenario of allopatric divergence (without being able to fully exclude transient gene flow
504 through climate-induced oscillations all along the Holocene) from an ancestral population
505 with relatively high effective size.

506 The Pyrenees mountains have the particularity to act as a biogeographic boundary
507 for many terrestrial organisms while being at the crossroads between temperate and
508 Mediterranean biomes (see Pironon et al., 2022 and references therein). Although orchid
509 propagules have the capacity to efficiently spread across long distances, several taxa occur
510 on either the Southern or the Northern slopes of the Pyrenees but have ranges that do not
511 extend latitudinally on both sides, or only weakly so (*i.e.* the so-called Northern or Southern
512 peripheral species in Pironon et al., 2022). The Pyrenees may thus alter the dispersal of at
513 least some orchid populations (e.g. *Dactylorhiza insularis*, *Gymnadenia gabasiana*, *Neotinea*
514 *conica*, *Ophrys castellana*, *O. riojana* and *O. subinsectifera*, *Orchis langei* and *O. simia*)
515 because of various biotic or abiotic factors. This mostly granitic and relatively high mountain
516 channel forms a vast area of unsuitable habitat for *Ophrys* orchids that generally require
517 calcareous soils and lowland or midland environments. Although data on their geographic
518 distribution remain incomplete, we may also speculate that the absence of *Ophrys* specific
519 pollinator species and/or of fungi to which orchid depend for their germination and their
520 development may also have limited the possibility of establishment for intermediate orchid
521 populations between the two current subranges. For *O. aveyronensis*, our results are
522 consistent with an overall reduction of gene flow between both sides of the Pyrenees
523 through the formation and the contraction of the two subranges during the Holocene until
524 the contemporary arrest of gene exchange that seems to be too recent to be determined
525 based on current data.

526

527 **Evidence for ecological speciation in addition to vicariance?**

528 Selection pressures exerted by insect pollinators are likely to be the main driver of ecological
529 and putatively sympatric speciation in *Ophrys* orchids (see Baguette et al., 2020). In this way,
530 populations of a given *Ophrys* species being pollinated by distinct insect pollinator species
531 may be considered as engaged in a speciation process. So far, field observations support that
532 *O. aveyronensis* from the Grands Causses and their Iberian counterparts are pollinated by
533 the same pollinator species: the solitary bee *Andrena hattorfiana* (see Benito Ayuso, 2019).
534 In a recent study on this system however, we showed that *O. aveyronensis* from the Grands
535 Causses and from the Iberian Peninsula present subtle differences at a set of morphological
536 and coloration traits that may be involved in pollination (Gibert et al., 2022). These subtle
537 differences could be emerging as a result of local adaptation to a "secondary" pollinator
538 species. Although additional work is required to address whether this variation may be
539 actually adaptive and involved in evolutionary divergence, these results call for i) an
540 assessment of which phenotypic traits may be under divergent selection across the *O.*
541 *aveyronensis* species complex geographic distribution and which ones may evolve through
542 drift only and ii) an investigation of putative genomic regions associated with phenotypic
543 trait values, genetic structure or ecological features, something that remains largely
544 underexplored in *Ophrys* (but see Sedeek et al., 2014, for a noticeable exception).

545 Carrying out cross-pollination experiments between individuals of the two subranges
546 may also help to highlight the insurgence of post-zygotic reproductive isolation within the
547 *Ophrys aveyronensis* species complex. However, such work may be impeded by the difficulty
548 to grow *Ophrys* under controlled conditions and because of the protection status of *Ophrys*
549 *aveyronensis* in France. Current knowledge suggest that such reproductive barriers may be
550 weak or absent at the *Ophrys* genus scale (Soliva and Widmer, 2003; Scopece et al., 2007,

551 Gervasi et al., 2017). To our best knowledge, only two noticeable exceptions to this pattern
552 involve the phylogenetically distant *Ophrys incubacea* and *Ophrys iricolor* (Cortis et al., 2009)
553 or species with different levels of ploidy: the diploid *Ophrys exaltata* and the tetraploid
554 *Ophrys lupercalis* (Vereecken et al., 2010).
555

556 **Taxonomic and conservation implications**

557 Considering that populations of the *O. aveyronensis* species complex are pollinated by the
558 same insect pollinator species across their range and have similar environmental niches, we
559 have currently no evidence to even consider the existence of different ecotypes in this
560 species complex. However, the disjunct geographic distribution in two subranges, separated
561 by ~600 km and the Pyrenees Mountains suggest that populations from the Grands Causses
562 and from the Iberian Peninsula may be considered as distinct evolutionary (and
563 conservation) units. Our population genomic analyses indicate that both sets of populations
564 are genetically differentiated and do not show evidence of ongoing gene flow. Accordingly,
565 the proposition to consider them as two subspecies: *Ophrys aveyronensis* subsp.
566 *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica*, (Paulus, 2017) seems appropriate.

567 Recent studies elaborated from some empirical evidence that delineating the so-
568 called 'micro-species' in *Ophrys* is practically cumbersome (Bateman et al., 2021; Bateman
569 and Rudall, 2023), especially in the section *Sphegodes* to which the *O. aveyronensis* species
570 complex belongs. Some genomic data such as whole plastomes that are useful to delineate
571 the main *Ophrys lineages*, present sequence similarity degree of more than 99.5% between
572 'microspecies' and therefore do not present enough variation to allow to discriminate them
573 (Bertrand et al., 2019; Bertrand et al., 2021b; Bateman et al., 2021). Here, however, our
574 results demonstrate that an approach consisting of sampling populations (*i.e.* a sufficient

575 number of individuals at a given location) and appropriate genetic markers (e.g. SNPs
576 genotyped on populations based on RAD-seq protocols) give signals that are consistent with
577 taxonomy. We therefore claim that such approaches of population genomics within an
578 integrative taxonomy framework (relying on flower morphology, pollinator identity,
579 ecological features…) is by far the most fruitful to better understand the process of
580 evolutionary divergence and the most relevant to improve at the same time our knowledge
581 on taxonomy, systematics, ecology and evolution of *Ophrys* orchids.
582
583 ***Ophrys aveyronensis* as a promising biological model to study speciation**
584 Allopatric divergence is a very common route toward speciation (Hewitt et al., 2000), but
585 most of empirical cases document it retrospectively, between nascent species that exhibit at
586 the same time obvious geographic separation and substantial levels of both phenotypic and
587 genetic differentiation (Dool et al., 2021). By contrast studies reporting cases of incipient
588 allopatric divergence are rare, especially in continental systems. The allopatric divergence in
589 the *Ophrys aveyronensis* species complex we report here is extremely recent compared to
590 other species complex harboring disjunct geographic distribution in terrestrial (e.g.
591 González-Serna et al., 2018; Tomasello et al., 2020; Wang et al., 2022; Cao et al., 2022;
592 Balmori-de la Puente et al., 2022) or marine systems (Schmid et al., 2023; Marcionetti et al.,
593 2023). It is likely to have arisen because of climate warming that dissected an ancestral
594 geographic distribution and subsequently constrained the spatial dynamics of its subranges
595 and does not *a priori* imply a human-induced biological introduction. This system therefore
596 may serve as a natural laboratory to assess early stages of reproductive isolation insurgence,
597 immediately after geographic separation.
598

599 **Author contributions**

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

601 AG and JB performed the analyses with contribution from CF and CR and wrote the first

602 version of the manuscript. All authors contributed substantially to the revisions.

603

604 **Data availability statement**

605 Sequencing data have been submitted to the European Nucleotide Archive (ENA;

606 https:/www.ebi.ac.uk/ena/browser/home) under Study with primary accession

607 n°PRJEB61037 (and secondary accession number ERP146119) and samples accession

608 n°ERS14864169 (SAMEA112857509) to n°ERS14864254 (SAMEA112857594).

609

610 ## Acknowledgements

611 This work was primarily supported by an inter-LabEX TULIP-CEMEB initiative (n°197310) to

612 Joris Bertrand and Bertrand Schatz. This research was also funded by an Agence Nationale

613 pour la Recherche Jeune Chercheur Jeune Chercheuse (ANR JCJC) grant to Joris Bertrand,

614 grant number ANR-21-CE02-0022-01, and is set within the framework of the "Laboratoires

615 d'Excellences (LABEX)" TULIP [ANR-10- LABX-41]. Lastly, this research was also funded by the

616 Observatoire de REcherche Méditerranéen de l'Environnement (SO Ocove & SO PolliMed,

617 OSU OREME) to Bertrand Schatz. We thank Jean-Luc Roux and Daniel Vizcaïno for the

618 information provided about sampling sites in Spain. We thank Christian Moliné for his

619 contribution to fieldwork, Moaine El Baidouri and Pascaline Salvado for their support with

620 bioinformatic analyses.

621

622 **Conflict of interest**

The authors declare no conflict of interest.

References

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., & Anderson, R.P. (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38, 541-545. doi: doi.org/10.1111/ecog.01132

Bagley, J.C., Heming, N.M., Gutiérrez, E. E., Devisetty, U.K., Mock, K. E., Eckert, A.J., & Strauss, S.H. (2020) Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (*Populus tremuloides*) *Ecology and Evolution*, 10, 4609-4629. doi: 10.1002/ece3.6214

Baguette, M., Bertrand, J.A.M, Stevens, V.M., & Schatz, B. (2020) Why are there so many bee-orchid species? Adaptive radiation by intra-specific competition for mnesic pollinators. *Biol Rev.*, 95, 1630-1663. doi: 10.1111/brv.12633

Balmori-de la Puente, A., Ventura, J., Miñarro, M., Somoano, A., Hey, J. & Castesana, J. (2022) Divergence time estimation using ddRAD data and an isolation-with-migration model applied to water vole populations of *Arvicola*, *Scientific Reports*, 12, 4065. doi: 10.1038/s41598-022-07877-y

Bateman, R.M., Rudall, P. J., Murphy, A.R.M., Cowan, R.S., Devey, D.S. & Peréz-Escobar, O.A. (2021) Whole plastomes are not enough: phylogenomic and morphometric exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect. *Sphegodes*. *Journal of Experimental Botany*, 72, 654-681. doi: 10.1093/jxb/eraa467

Bateman, R.M. & Rudall, P.J. (2023) Morphological continua make poor species: genus-wide morphometric survey of the European Bee Orchids (*Ophrys* L.). *Biology*, 12, 136. doi: 10.3390/ biology12010136

647 Beatty, G.E., & Provan, J. (2013) Post-glacial dispersal, rather than *in situ* glacial survival, best
648 explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica*
649 (Ericacea). *Journal of Biogeography*, 40, 335-344. doi: 10.1111/j.1365-
650 2699.2012.02789.x
651 Beatty, G.E., & Provan, J. (2014) Phylogeographical analysis of two cold-tolerant plants with
652 disjunct Lusitanian distributions does not support *in situ* survival during the last
653 glaciation. *Journal of Biogeography*, 41, 2185-2193. doi: 10.1111/jbi.12371
654 Beatty, G.E., Lennon, J.J., O'Sullivan, C.J., & Provan, J. (2015) The not-so-Irish spurge:
655 *Euphorbia hyberna* (Euphorbiaceae) and the Littletonian plant 'steeplechase'.
656 *Biological Journal of the Linnean Society*, 114, 249-259. doi: 10.1111/bij.12435
657 Benito Ayuso, J. (2019) Estudios sobre polinización en el généro *Ophrys* (Orchidaceae), I,
658 *Flora Montiberica*, 74, 32-37.
659 Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2019) Characterization of the
660 complete plastome of *Ophrys aveyronensis*, a Euro-Mediterranan orchid with an
661 intriguing geographic distribution. *Mitochondrial DNA Part B*, 4, 3256-3257. doi:
662 10.1080/23802359.2019.1670748
663 Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2021) Whole plastid genome-based
664 phylogenomics supports an inner placement of the *O. insectifera* group rather than a
665 basal position in the rapidly diversifying *Ophrys* genus (Orchidaceae). *Biology Letters*,
666 168, 452-457. doi: 10.1080/23818107.2021.1893216
667 Bertrand, J. A. M., Baguette, M., Joffard, N., & Schatz, B. (2021) Challenges inherent in the
668 systematics and taxonomy of genera that have recently experienced explosive
669 radiation: the case of orchids of the genus *Ophrys*. In Grancolas, P. & Maurel, M.-C.
670 *Systematics and the Exploration of Life* (pp 113-134). Hoboken, New Jersey: Wiley.

671 Breitkopf, H., Onstein, R.E., Cafasso, D., Schlüter, P.M., & Cozzolino, S. (2015) Multiple shifts
672 to different pollinators fuelled rapid diversification in sexually deceptive *Ophrys*
673 orchids. *New Phytologist*, 207, 377-389. doi: 10.1111/nph.13219
674 Cao, Y., Almeida-Silva, F., Zhang, W.-P., Ding, Y.-M., Bai, D., Bai, W.-N., Zhang, B.W., Van de
675 Peer, Y.- & Zhang, D.Y. (2023) Genomic Insights into Adaptation to Karst Limestone
676 and Incipient Speciation in East Asian *Platycarya* spp. (Juglandaceae). *Molecular*
677 *Biology and Evolution*, 40, msad121. doi: 10.1093/molbev/msad121
678 Catchen, J., Amores, A., Cresko, W., & Postlewait, J. (2011) Stacks: building and genotyping
679 loci de novo from short-read sequences. *G3: genes, Genomes, Genetics*, 1, 171-182.
680 doi: 10.1534/g3.111.000240
681 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2013) Stacks: an
682 analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140. doi:
683 /10.1111/mec.12354
684 Cortis, P., Vereecken, N.J., Schiestl, F.P., Barone Lumaga, M.R., Scrugli, A. & Cozzolino, S.
685 (2009) Pollinator convergence and the nature of species' boundaries in sympatric
686 Sardinian *Ophrys* (Orchidaceae). *Annals of Botany*, 104, 497-506. doi:
687 /10.1093/aob/mcn219
688 Coyne, J.A., & Orr, H.A. (2004). *Speciation*. Sunderland, Mass: Sinauer Associates.
689 Claessens, J., & Kleynen, J. (2016) *Orchidées d'Europe : fleurs et pollinisation*. Biotope, Mèze,
690 France: Biotope.
691 Davison, A., Birks, J.D.S., Brookes, R.C., Messenger, J.E., & Griffiths, H.I. (2001) Mitochondrial
692 phylogeography and population history of pine martens *Martes martes* compared
693 with polecats *Mustela putorius*. *Molecular Ecology*, 10, 2479-2488. doi:
694 10.1046/j.1365-294x.2001.01381.x.

Delforge, P. (2016) *Orchidées d'Europe, d'Afrique du Nord et du Proche Orient*. Lausanne, Switzerland: Delachaux & Niestlé.

Dool, S.E., Picker, M.D. & Eberhard, M.J.B. (2021) Limited dispersal and local adaptation promote allopatric speciation in a biodiversity hotspot. *Molecular Ecology*, 31, 279-295. doi: 10.1111/mec.16219

Fagúndez, J. & Díaz-Tapia, P. (2023) Comparative phylogeography of a restricted and a widespread heather: genetic evidence of multiple independent introductions of *Erica mackayana* into Ireland from northern Spain. *Botanical Journal of the Linnean Society*. doi: 10.1093/botlinnean/boac071

Fick, S.E., & Hijmans, R.J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302-4315. doi: 10.1002/joc.5086

Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T. M. L., Otto-Bliesner, B. L., Chan, K. C., & Brook, B.W. (2017) PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40, 1348-1358. Doi: doi.org/10.1111/ecog.03031

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire E., Simon, A., Galtier, N., Duret, L., Bierne, N., Vekemans, X., & Roux C. (2021) DILS : Demographic inferences with linked selection using ABC. *Molecular Ecology Resources*, 2629-2644. doi: doi.org/10.1111/1755-0998.13323

Francis, R.M. (2017) POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17, 27-32. doi: 10.1111/1755-0998.12509

718 Frichot, E., Mathieu, F., Trouillon T., Bouchard, G., & François O. (2014) Fast and efficient
719 estimation of individual ancestry coefficients. *Genetics*, 196, 973-983. doi:
720 10.1534/genetics.113.160572
721 Frichot, E. & François O. (2015) LEA: An R package for landscape and ecological association
722 studies. *Methods in Ecology and Evolution*, 6, 925-929. doi: 10.1111/2041-
723 210X.12382
724 Gervasi, D.L., Selosse, M.-A., Sauve, M., Francke, W., Vereecken, N.J., Cozzolino, S. & Schiestl,
725 F.P. (2017) Floral scent and species divergence in a pair of sexually deceptive orchids.
726 *Ecology and Evolution*, 7, 6023-6034. doi: 10.1002/ece3.3147
727 Gibert, A., Louty, F., Buscail, R., Baguette, M., Schatz, B. &, Bertrand, J.A.M. (2022) Extracting
728 quantitative information from images taken in the wild: a case study of two vicariants
729 of the *Ophrys aveyronensis* species complex. *Diversity*, **14**, 400. doi:
730 10.3390/d14050400
731 Grindon, A. J., & Davison, A. (2013) Irish *Cepea nemoralis* land snails have a cryptic Franco-
732 Iberian origin that is most easily explained by the movements of Mesolithic humans.
733 *PLoS ONE*, 8, e65792. doi: doi.org/10.1371/journal.pone.0065792
734 González-Serna, M., Cordero, P.J. & Ortego, J. (2018) Using high-throughput sequencing to
735 investigate the factors structuring genomic variation of a Mediterranean grasshopper
736 of great conservation concern. *Scientific Reports*, 8, 13436. doi: 10.1038/s41598-018-
737 31775-x
738 Heming, N. M., Dambros, C., & Gutiérrez, E. E. (2018). ENMwizard: AIC model averaging and
739 other advanced techniques in Ecological Niche Modeling made easy. R package
740 version 0.1.7. Retrieved from https:// github.com/HemingNM/ENMwizard

741 Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913. doi:
742 10.1038/35016000/
743 Hijmans, R. J., Cameron, S.E., Parra, J. -L., Jones, P. J., & Jarvis A. (2005) Very high resolution
744 interpolated climate surfaces for global land areas. *International Journal of*
745 *Climatology*, 25, 1965-1978. doi: 10.1002/joc.1276
746 Inda, L.A., Pimentel, M., & Chase, M.W. (2012) Phylogenetics of tribe Orchideae
747 (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding
748 timing of diversification and evolution of pollination syndromes. *Annals of Botany.*,
749 110, 71-90. doi: 10.1093/aob/mcs083
750 Joffard, N., Massol, F., Grenié, M., Montgellard, C. & Schatz, B. (2019) Effect of pollination
751 strategy, phylogeny and distribution niches of Euro-Mediterranean orchids. *Journal of*
752 *Ecology*, 107, 478-490. doi: 10.1111/1365-2745.13013
753 Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers.
754 *Bioinformatics*, 24, 1403-1405. doi: 10.1093/bioinformatics/btn129
755 Jombart, T., & Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide
756 SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr521
757 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.F.K., von Haeseler, A. & Jermiin, L.S. (2017)
758 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
759 *Methods*, 14, 587-589. doi: 10.1038/nmeth.4285
760 Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021):
761 CHELSA-TraCE21k v1. 0. Downscaled transient temperature and precipitation data
762 since the last glacial maximum. *Climate of the Past Discussions*, 1-27. doi: 10.5194/cp-
763 2021-30

764 Krasovec, M., Chester, M., Ridout, K., & Filatov, D.A. (2018) The mutation rate and the age of
765 the sex chromosomes in *Silene latifolia*. *Current Biology*, 28, 1832-1838. doi:
766 10.1016/j.cub.2018.04.069
767 Kreutz, C.A. (2007) Beitrag zur Taxonomie und Nomenklatur europäischer, mediterraner,
768 nordafrikanischer und vorderasiatischer Orchideen. *Ber. Arbeitskrs. Heim. Orchid*. 24,
769 77-141.
770 Kuhn, M. (2019) Caret: classification and regression training. https://CRAN. R-
771 project.org/package=caret.
772 Liu, X., & Fu, Y.-X. (2015) Exploring population changes using SNP frequency spectra. *Nature*
773 *Genetics*, 47, 555-559. doi: 10.1038/ng.3254.
774 Liu, X., & Fu, Y.-X. (2020) Stairway Plot 2: demographic inference with folded SNP frequency
775 spectra. *Genome Biology*, 21, 280. doi: 10.1186/s13059-020-02196-9
776 Marcionetti, A., Bertrand, J. A.M, Cortesi, F., Donati, G.F.A., Heim, S., Huyghe, F., Kochzius,
777 M., Pelissier, L. & Salamin N. (2023) Recurrent gene flow events shaped the
778 diversification of the clownfish skunk complex. *bioRxiv*. doi:
779 10.1101/2023.10.24.562491
780 Mascheretti, S., Rogatcheva M. B., Gündüz, I., Fredga, K., & Searle, J.B. (2003) How did
781 pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial
782 cytochrome *b* sequences. *Proceedings of the Royal Society B*, 270, 1593-1599. doi:
783 doi.org/10.1098/rspb.2003.2406
784 Meirmans, P. G. (2020) GENODIVE version 3.0: Easy-to-use software for the analysis of
785 genetic data of diploids and polyploids. *Molecular Ecology Resources*, 20, 1126-1131.
786 doi: 10.1111/1755-0998.13145

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, A., von Haeseler, A. & Lanfear, R. (2020) IQ-TREE 2: New methods and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530-1534. Doi: 10.1093/molbev/msaa015

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., & Anderson, R.P. (2014) ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198-1205. doi: 10.1111/2041-210X.12261

Nosil, P., Feder, J.L., Flaxman, S.M. & Gompert Z. (2017) Tipping points in the dynamics of speciation. *Nature Ecology and Evolution*, 1, 0001. doi: 10.1038/s41559-016-0001

Nürk, M.N., Linder, H.P., Onstein, R.E., Larcombe, M.J., Hughes, C.H., Piñeiro Fernández, L., Schlüter, P.M., Valente, L.M., Beierkuhnlein, C., Cutts, V., Donoghue, M.J., Edwards, E.J., Field, R., Flantua, S.G.A., Higgins, S.II, Liede-Schumann, S. & Pirie, M.D. (2020) Diversification in evolutionary arenas – assessment and synthesis. *Ecology and Evolution*, 10, 6163-6182. doi:/10.1002/ece3.6313

Paris, J., Stevens, J.R., & Catchen J.M. (2017) Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution*, 8, 1360-1373. doi: 10.1111/2041-210X.12775

Paulus, H.F. & Gack, C. (1999) Bestäubungsbiologische Untersuchungen an der Gattung *Ophrys* in der Provence (SO Frankreich), Ligurien und Toscana (NW Italien) (Orchidaceae und Insecta, Apoidea). *Journal Europäischer Orchideen*. 31, 347-422.

Paulus, H.F. (2017) Bestäubungs-biologie *Ophrys* in Nordspanien. *Journal Europäischer Orchideen*. 49, 427-471.

811 Pironon, S., Gómez, D., Font, X. & García, M.B. (2022) Living at the limit in the Pyrenees:
812 Peripheral and endemic plants are rare but underrepresented in protection lists.
813 *Diversity and distribution*, 28, 930-943. doi: 10.1111/ddi.13487
814 Phillips, S.J., Anderson, R.P., & Schapire, R.E. (2006) Maximum entropy modelling of species
815 geographic distributions. *Ecological Modelling*, 190, 231-259. doi:
816 10.1016/j.ecolmodel.2005.03.026
817 Phillips, S.J., & Dudík, M. (2008). Modelling of species distributions with Maxent: New
818 extensions and a comprehensive evaluation. *Ecography*, 31, 161–175. doi:
819 doi.org/10.1111/j.0906-7590.2008.5203.x
820 Philipps, S.J., Anderson, R.P., Dudík, M., Shapire, R.E., & Blair, M.E. (2017) Opening the black
821 box: an open-source release of Maxent. *Ecography*, 40, 887-893. doi:
822 doi.org/10.1111/ecog.03049
823 Reich, I., Gormally, M., Allcock, A.L., McDonnel, R., Castillejo, J., Iglesias, J., Quinteiro J. &
824 Smith, C.J. (2015) Genetic study reveals close link between Irish and Northern Spanish
825 specimens of the protected Lusitanian slug *Geomalacus maculus*. *Biological Journal of*
826 *the Linnean Society*, 116, 156-168. doi: 10.1111/bij.12568
827 Rochette, N.C., & Catchen J.M. (2017) Deriving genotypes from RAD-seq short-read data
828 using Stacks. *Nature Protocols*, 12, 2640-2659. doi:10.1038/nprot.2017.123
829 Salvado, P., Aymerich Boixader, P., Vila Bonfill, A., Martin, M., Quélennec, C., Lewin, J.-M.,
830 Hinoux, V., & Bertrand, J.A.M. (2022) Little hope for the polyploid endemic Pyrenean
831 Larkspur (*Delphinium montanum*): evidences form population genomics and
832 Ecological Niche Modeling. *Ecology and Evolution*, 12, e8711. doi: 10.1002/ece3.8711
833 Schmid, S., Bachmann Salvy, M., Garcia Jimenez, A., Bertrand, J.A.M., Cortesi F., Heim, S.,
834 Huyghe, F., Kochzius, M., Pelissier, L. & Salamin, N. Gene flow throughout the

835 evolutionary history of a polymorphic and generalist clownfish. *bioRxiv*. Doi:
836 10.1101/2023/10.19.563036
837 Scopece, G., Musacchio, A., Widmer, A. & Cozzolino, S. (2007) Patterns of reproductive
838 isolation in Mediterranean deceptive orchids. *Evolution*, 61, 2623-2642. doi:
839 doi.org/10.1111/j.1558-5646.2007.00231.x
840 Sedeek, K.E.M., Scopece, G., Staedler, Y.M., Schönenberger, J., Cozzolino, S., Schiestl, F. P., &
841 Schlüter, P.M. (2014) Genic rather than genome-wide differences between sexually
842 deceptive *Ophrys* orchids with different pollinators. *Molecular Ecology*, 23, 6192-
843 6205. doi: 10.1111/mec.12992
844 Soliva, M. & Widmer, A. (2003) Gene flow across species boundaries in sympatric sexually
845 deceptive *Ophrys* (Orchidaceae) species. *Evolution*, 57, 2252-2261. doi:
846 10.1111/j.0014-3820.2003.tb00237.x.
847 Stankowski, S., Westram, A.M., Zagrodzka, Z.B., Eyres, I., Broquet, T., Johannesson, K., &
848 Butlin, R.K. (2020). The evolution of strong reproductive isolation between sympatric
849 intertidal snails. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
850 375, 20190545. doi:10.1098/rstb.2019.0545
851 Stankowski, S. & Ravinet, M. (2021) Defining the speciation continuum. *Evolution*, 75, 1256-
852 1273. doi:10.1111/evo.14215
853 Tomasello, S., Karbstein, K., Hodač, L., Paetzold, C. & Hörandl, E. (2020) Phylogenomics
854 unravels Quaternary vicariance and allopatric speciation patterns in temperate-
855 montane plant species: A case study on the *Ranunculus auricomus* species complex.
856 *Molecular Ecology*, 29, 2031, 2049. doi: 10.1111/mec.15458

857 Vereecken, N.J., Cozzolino, S. & Schiestl, F.P. (2010) Hybrid floral scent novelty drives
858 pollinator shift in sexually deceptive orchids. *BMC Evolutionary Biology*, 10, 103. doi:
859 10.1186/1471-2148-10-103
860 Wang, R.-H., Yang, Z.-P., Zhang, Z.-C., Comes, H.P., Qi, Z.-C., Li, P. & Fu, C.-X. (2022) Plio-
861 Pleistocene climatic change drives allopatric speciation and population divergence
862 within the *Scrophularia incisa* complex (Scrophulariaceae) of desert and steppe
863 subshrubs in Northwest China. *Frontiers in Plant Science*, 13, 985372. doi:
864 10.3389/fpls.2022.985372.
865 Weir, B.S. & Goudet, J. (2017) A unified characterization of population structure and
866 relatedness. *Genetics*, 206, 2085-2103. doi:
867 https://doi.org/10.1534/genetics.116.198424
868

869 **Table 1.** Sampling site, sample size (*n*), geographic coordinates and elevation, population
870 summary statistics: number of private alleles, expected and observed heterozygosity (H_E and
871 H_O), inbreeding coefficient (G_{IS}) and nucleotide diversity (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

	86									

872

873 **Table 2.** Summary of the main demographic parameters inferred with DILS i) for pooled
874 dataset (*Ophrys aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica*)
875 and ii) for each *Ophrys aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp.
876 *vitorica* population pairs. N_e 1, N_e 2, N_e anc. correspond to effective population sizes
877 estimated for population 1, population 2 and the ancestral population, t_{split} to the time of
878 split between population 1 and population 2 in number of generations, respectively and
879 M_{12}/M_{21} the migration rates (as N_e.m) from population 1 to population 2 and from
880 population 2 to population 1. Parameters have been estimated based on a model using
881 neural network.

Population 1	Population 2	*Ne 1*	*Ne 2*	*Ne anc.*	*tsplit*	M_{12}/M_{21}
O. a. *aveyronensis* pool (France)	O. a. *vitorica* 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)

882

883 **Table 3.** List of bioclimatic variables available from WorldClim. From these 19 variables (bio1
884 to bio19) the ones used for ENM analyses (Person correlation coefficient < 0.75) are
885 indicated 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	

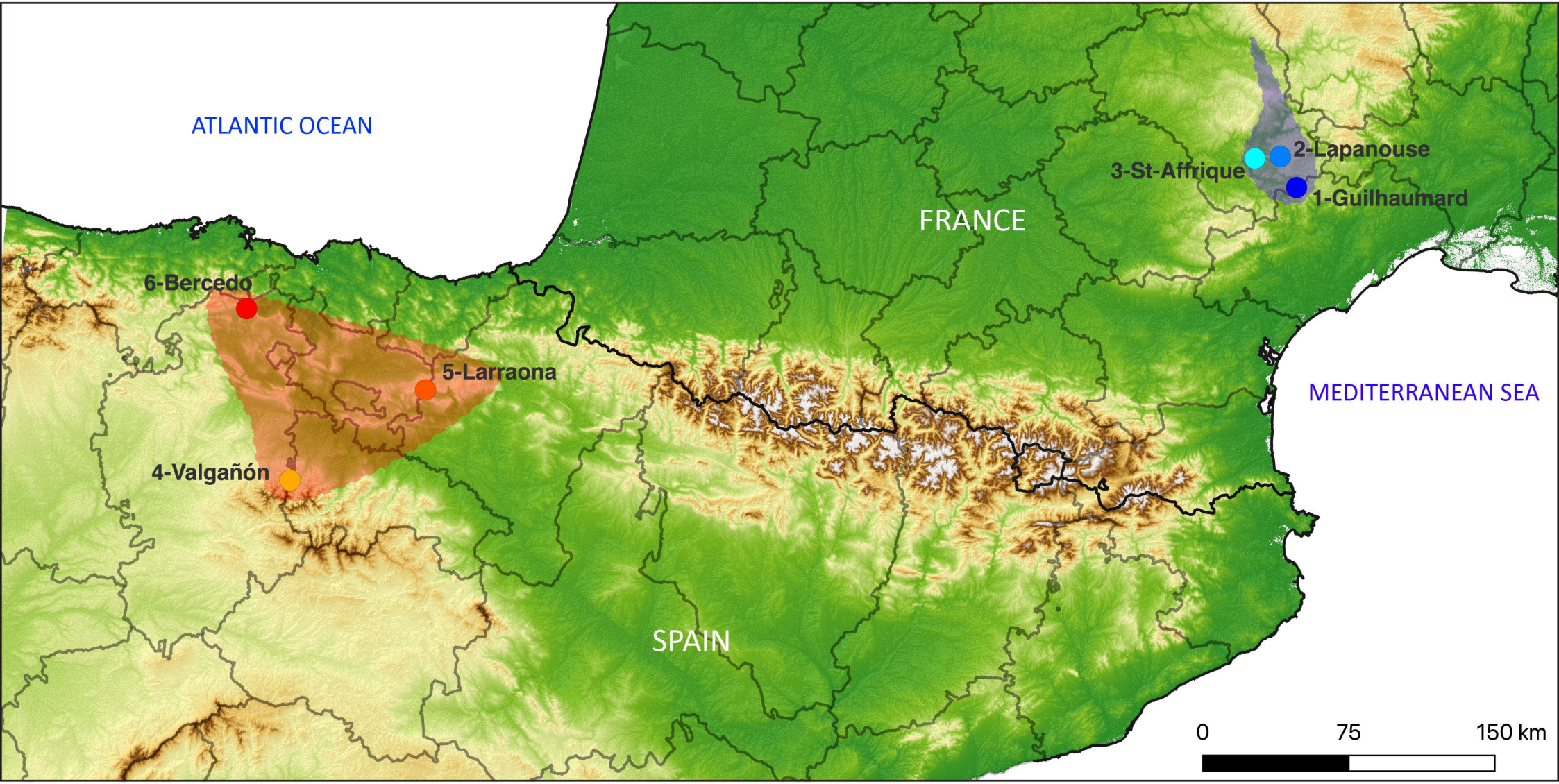
886

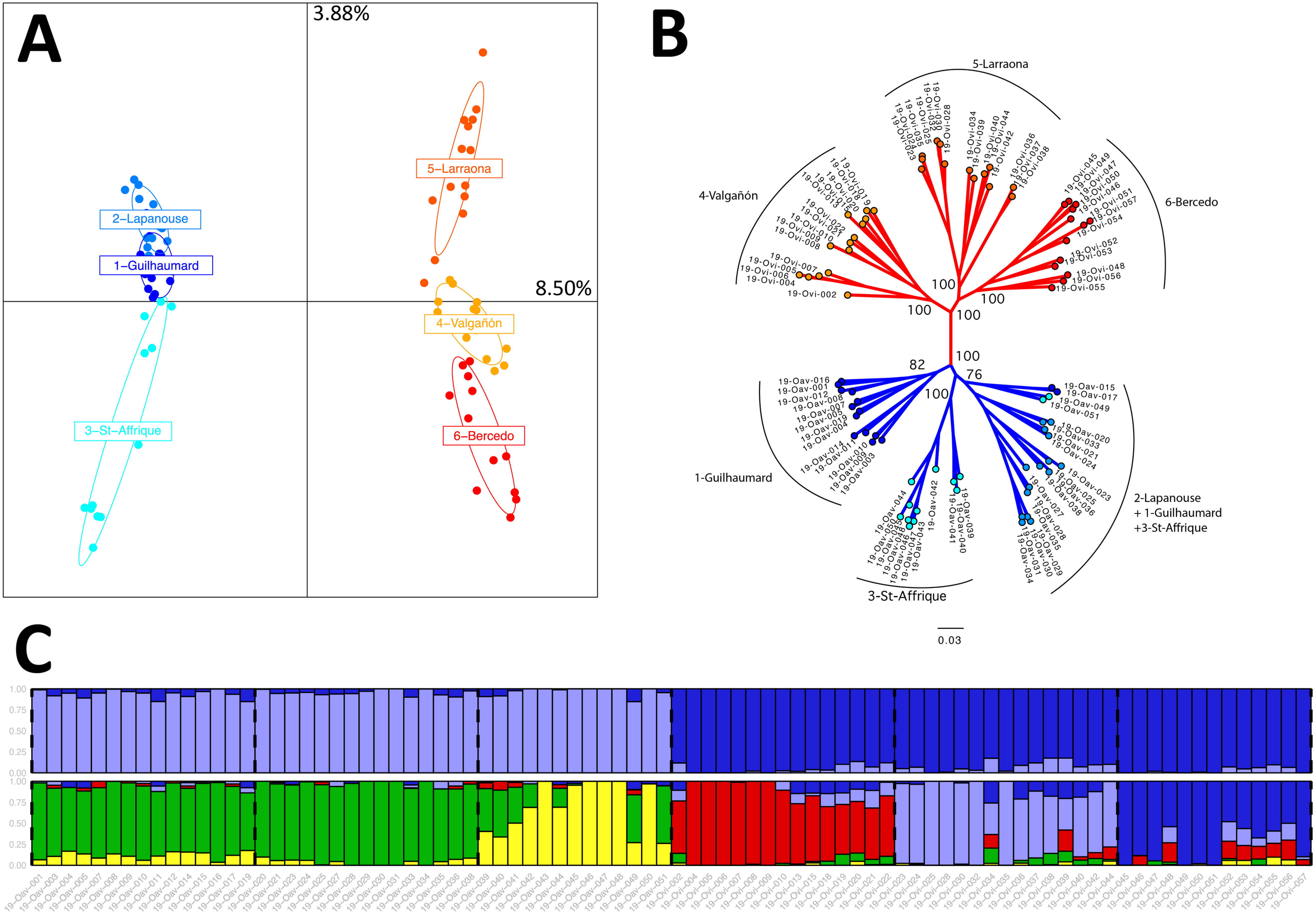
887 **Figure 1** Map displaying the 6 sites sampled: 3 for *Ophrys aveyronensis* subsp. *aveyronensis*:
888 1-Guilhaumard, 2-Lapanouse-de-Cernon, 3-St-Affrique (in blue) and 3 for *O. a.* subsp.
889 *vitorica*: 4- Valgañón, 5-Larraona, 6-Bercedo (in red). The approximate known geographic
890 distribution is highlighted for both taxa with the same color code.

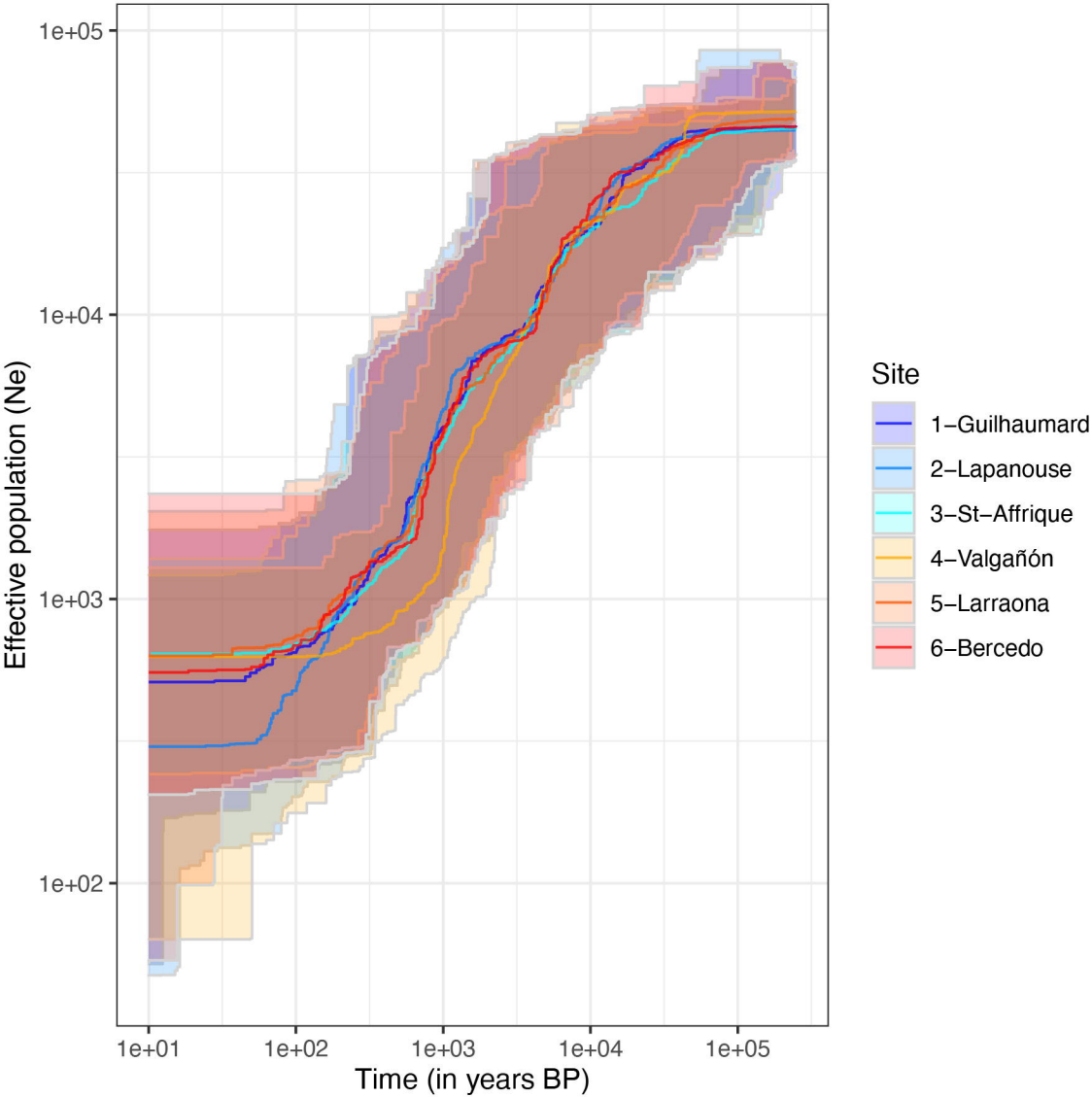
891

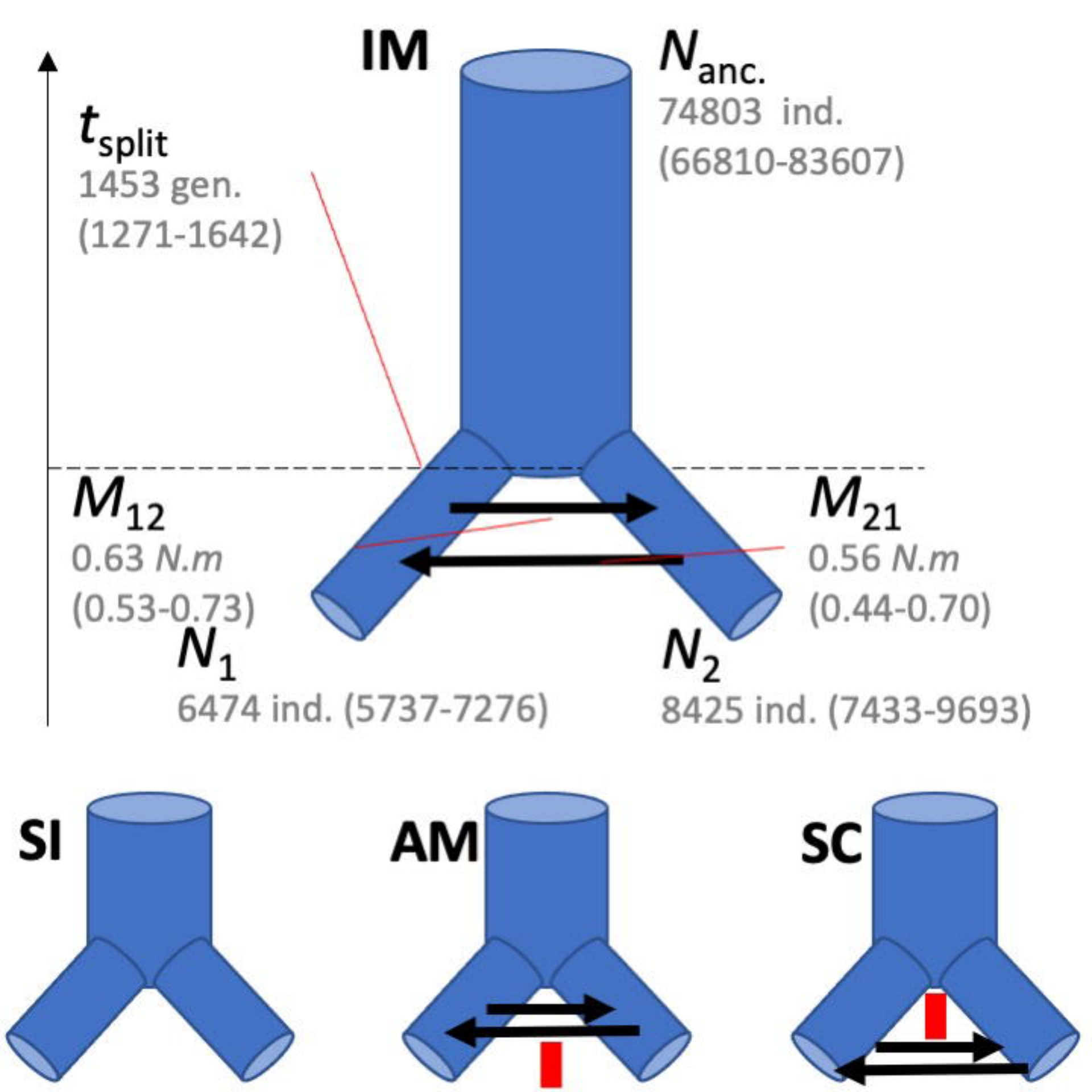
892 **Figure 2** A) Principal Component Analysis (PCA) displaying the two first axes (PC1 and PC2)
893 representing 8.50% and 3.88% of the total genetic variance. PCA was computed based on 86
894 individuals genotyped at 9301 SNP and colors depict sampling localities. B) Maximum
895 Likelihood phylogenetic tree (rooted at midpoint) obtained from the concatenation of the
896 SNPs (node support for main nodes are given as UltraFast Boostrap Approximation based on
897 1000 replicates). C) Barplots of ancestry coefficients obtained from sNMF for 86 individuals
898 for K = 2 and K = 5.
899
900 **Figure 3** Demographic history of the 6 studied populations of *O. a.* subsp. *aveyronensis*
901 (shades of blue) and *O. a.* subsp. *vitorica* (shades of red) depicting the mean effective
902 population size N_e (and associated 95 confidence intervals) over time as inferred from
903 STAIRWAYPLOTv2, assuming a mutation rate of 7 x10^{-9} and a generation time of 5 years.
904
905 **Figure 4** Schematic overview of the best demographic scenario (Isolation with Migration, IM)
906 for pooled dataset (*i.e. O. aveyronensis* subsp. *aveyronensis*/ *O. a.* subsp. *vitorica*). The
907 median values(and 2.5-97.5% boundaries of Highest Posterior Distribution probability) are
908 shown for the parameters inferred with DILS. N_a: ancestral effective population size, N_1:
909 (current) effective population size for the *O. a.* subsp. *aveyronensis* population (given in
910 number of individuals), N_2: (current) effective population size for the *O. a.* subsp. *vitorica*,
911 M_{12}: migration rate from *O. a.* subsp. *aveyronensis* to *O. a.* subsp. *vitorica* and M_{21}: migration
912 rate from *O. a.* subsp. *vitorica* to *O. a.* subsp. *aveyronensis*. Alternative demographic
913 scenarios tested are also illustrated: Strict Isolation (SI), Ancestral Migration (AM, involving a
914 cessation of gene flow after an initial stage of migration after the split) and Secondary
915 Contact (SC, involving a secondary contact after an initial stage of isolation after the split).

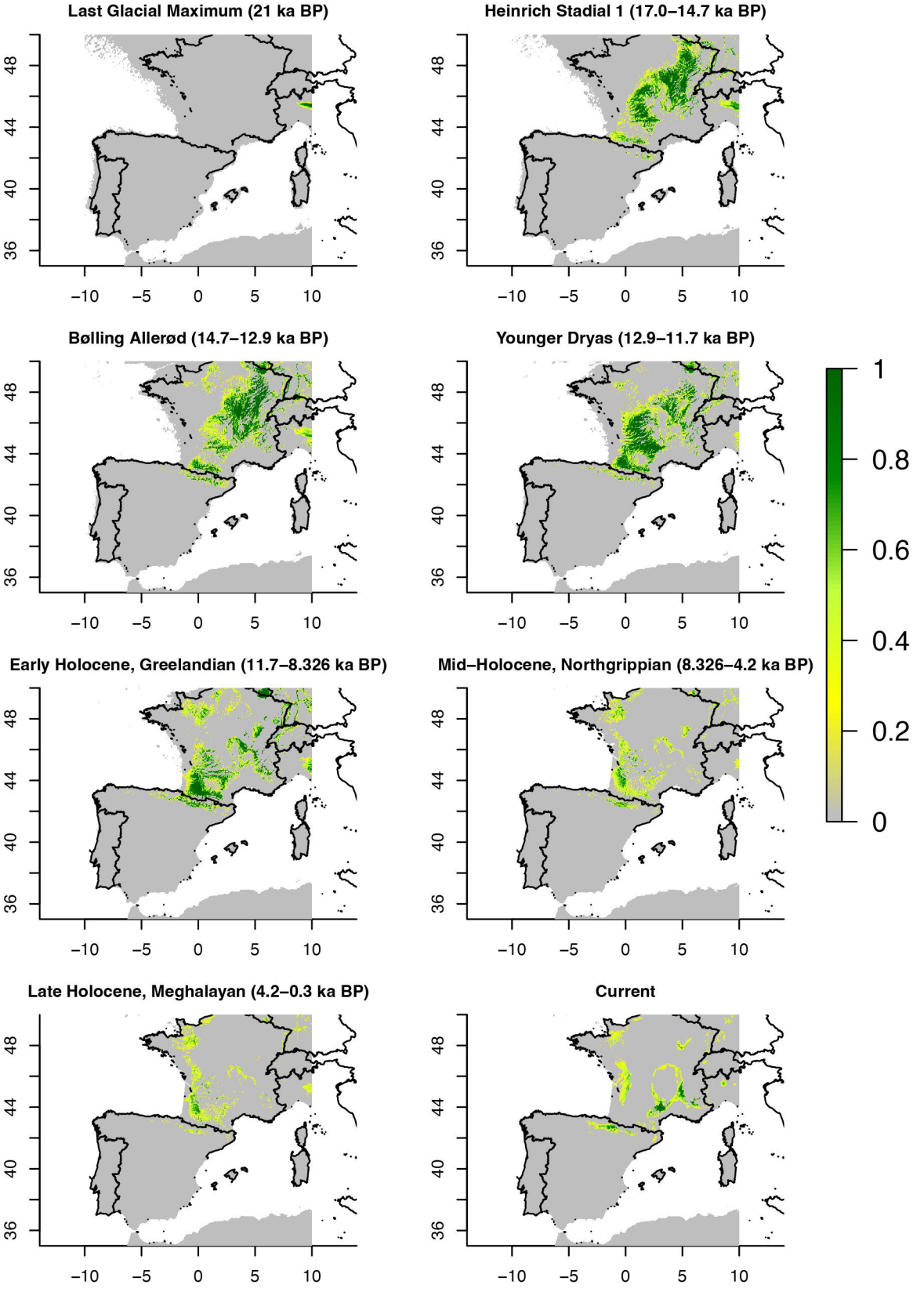
Figure 5 Spatial projections of the bioclimatic niches inferred for the *Ophrys aveyronensis* species complex depicting climate suitability at different time periods from the Last Glacial Maximum (LGM, about 21 000 kyrs ago) and with projections for the Heinrich Stadial1 (17.0-14.7 kyrs BP), the Bølling-Allerød (14.7-12.9 kyrs BP) the Younger Dryas Stadial (12.9-11.7 kyrs BP) and the Holocene: early-Holocene, Greenlandian (11.7-8.326 kyrs BP), mid-Holocene, Northgrippian (8.326-4.2 kyrs BP) and late-Holocene, Meghalayan (4.2-0.3 kyrs BP) and currently.

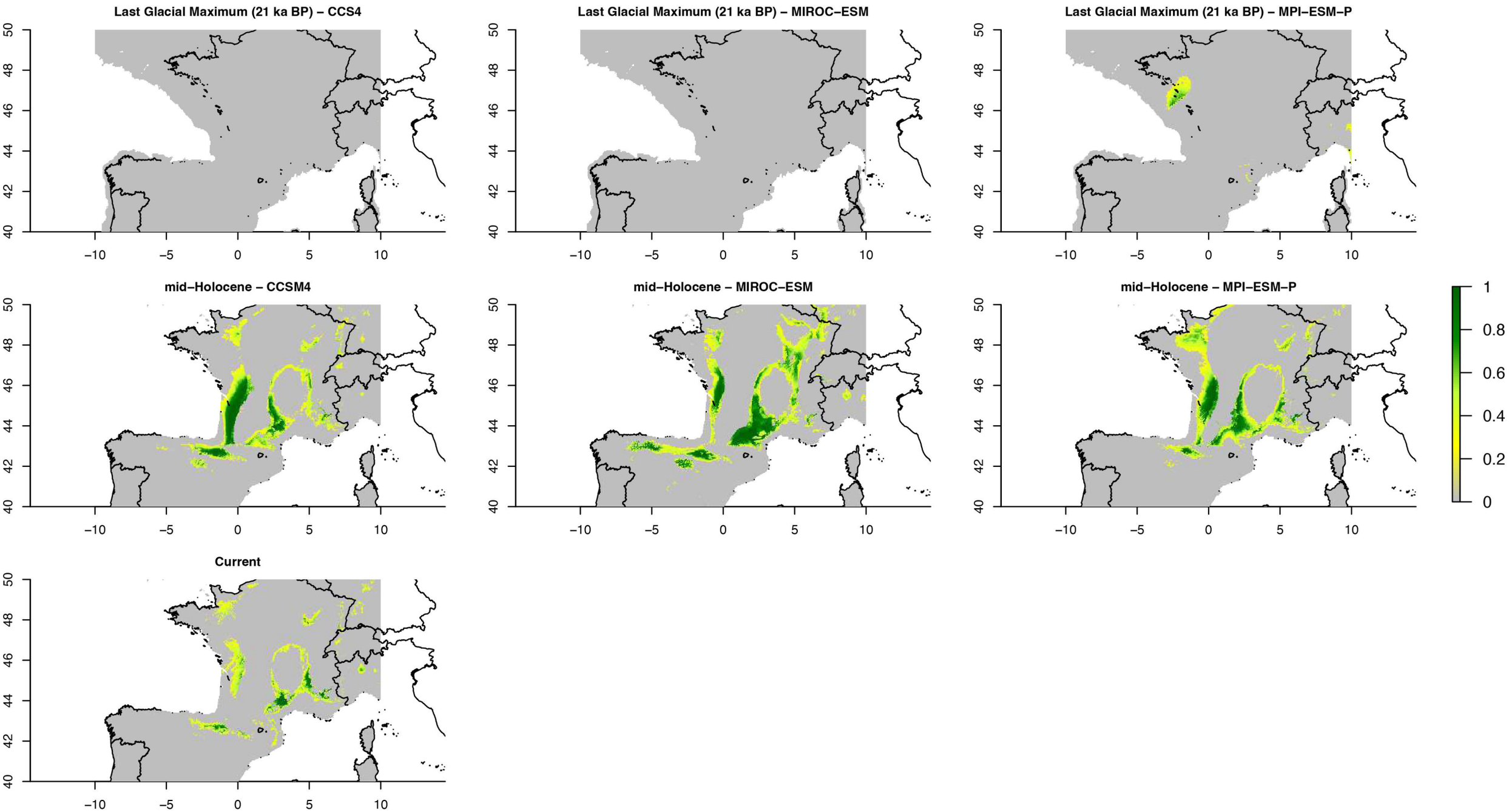

ATLANTIC OCEAN
FRANCE
SPAIN
MEDITERRANEAN SEA
1-Guilhaumard
2-Lapanouse
3-St-Affrique
4-Valgañón
5-Larraona
6-Bercedo
0
75
150 km


A
3.88%
8.50%
2–Lapanouse
1–Guilhaumard
3–St–Affrique
5–Larraona
4–Valgañón
6–Bercedo
B
5-Larraona
4-Valgañón
6-Bercedo
1-Guilhaumard
2-Lapanouse
+ 1-Guilhaumard
+3-St-Affrique
3-St-Affrique
0.03
C
O. a. subsp. aveyronensis
O. a. subsp. vitorica
1–Guilhaumard
2–Lapanouse
3–St–Affrique
4–Valgañón
5–Larraona
6–Bercedo


1e+05
1e+04
1e+03
1e+02
Effective population (Ne)
1e+01
1e+02
1e+03
1e+04
1e+05
Time (in years BP)
Site
1−Guilhaumard
2−Lapanouse
3−St−Affrique
4−Valgañón
5−Larraona
6−Bercedo


IM
$N_{anc.}$
74803 ind.
(66810-83607)
t_{split}
1453 gen.
(1271-1642)
M_{12}
0.63 *N.m*
(0.53-0.73)
M_{21}
0.56 *N.m*
(0.44-0.70)
N_1
6474 ind. (5737-7276)
N_2
8425 ind. (7433-9693)
SI
AM
SC


Last Glacial Maximum (21 ka BP)
Heinrich Stadial 1 (17.0–14.7 ka BP)
Bølling Allerød (14.7–12.9 ka BP)
Younger Dryas (12.9–11.7 ka BP)
Early Holocene, Greelandian (11.7–8.326 ka BP)
Mid–Holocene, Northgrippian (8.326–4.2 ka BP)
Late Holocene, Meghalayan (4.2–0.3 ka BP)
Current
48
44
40
36
−10
−5
0
5
10
1
0.8
0.6
0.4
0.2
0


Last Glacial Maximum (21 ka BP) – CCS4
Last Glacial Maximum (21 ka BP) – MIROC–ESM
Last Glacial Maximum (21 ka BP) – MPI–ESM–P
mid–Holocene – CCSM4
mid–Holocene – MIROC–ESM
mid–Holocene – MPI–ESM–P
Current
40
42
44
46
48
50
–10
–5
0
5
10
0
0.2
0.4
0.6
0.8
1