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] Evo-Eco-Paleo, UMR 8198, Centre National de la Recherche Scientifique (CNRS) - Université de Lille, 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 indicate a recent divergence (*i.e.* ~3500 generations ago). Ecological Niche Modelling results further support 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 since the Last Glacial Maximum (LGM, about 21000 years ago). The congruence of both the results of population genomics and ENM approaches indicate that allopatric speciation is involved in the *Ophrys* adaptive

radiation. The *O. aveyronensis* complex thus provides a promising candidate to further investigate putative incipient ecological divergence following an initial phase of geographic isolation.

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

52 **Introduction**

53 Allopatric speciation (also referred as geographic speciation, or vicariant speciation) occurs
54 when populations of a given species become reproductively isolated from each other
55 because of geographic separation. Aside from being conceptually intuitive, allopatric
56 speciation has been evidenced by theoretical models, laboratory experiments and empirical
57 studies to the point that it is generally regarded as the "default mode" of speciation (see
58 Coyne & Orr, 2004). In allopatric speciation, spatial and reproductive isolation are often
59 assumed to result from geologic-caused topographic changes that form the geographic
60 barrier. In continental systems however, allopatric divergence is prone to experience only
61 transient geographic and reproductive isolation involving one of several phases of secondary
62 contact (e.g. through climate change-induced oscillations during the Quaternary, Hewitt,
63 2000). Thus, allopatry is a spatio-temporally dynamic process. Aside, not only the fore-
64 mentioned barriers but any kind of fragmentation of species distribution range, including
65 climate change and/or human activities may also initiate and be responsible for evolutionary
66 divergence.

67 The putative complexity of the speciation process makes difficult to elucidate its
68 ecological and evolutionary causes as current spatial patterns of distribution only provide a
69 snapshot from the speciation continuum. In this regard, species with disjunct geographic
70 distributions are likely to correspond to species whose populations became recently
71 allopatric, *i.e.* with insufficient time to accumulate enough divergence to be considered as
72 distinct species yet. Although such natural example may help to understand the
73 consequences of physical barriers in early stages of evolutionary divergence, disjunct
74 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.

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

The species of the genus *Ophrys* have long been known for their unusual pollination syndrome (called 'sexual swindling') that involves insect-like flowers mimicking their female pollinator to lure conspecific males and ensure pollination but are now also increasingly considered as a promising biological model to investigate adaptive radiations. The degree of plant-pollinator specialization (Joffard et al., 2019) required to ensure the success of the

99 *Ophrys* pollination strategy was shown to be the driver of the impressive rate of
100 diversification of the genus within the Mediterranean basin, and to a lesser extent in other
101 parts of Western Europe (see Baguette et al., 2020, for a recent review). If *Ophrys* are
102 notorious for their relatively high number of likely cases of speciation with gene flow, the
103 number of island endemic taxa prove that allopatric speciation is also an important
104 mechanism to explain *Ophrys* species richness observable nowadays (see Bertrand et al.,
105 2021a).

106 For continental taxa, clear evidences of allopatric speciation in *Ophrys* are missing.
107 Nevertheless, one species, *Ophrys aveyronensis*, displays an intriguing case of disjunct
108 geographic distribution. This species was initially described as endemic to a very restricted
109 area in the Grands Causses region (Southern France). Observations of morphologically very
110 similar populations were reported later from Northern Spain and described as a distinct
111 species (*Ophrys vitorica*, Kreutz, 2007). *Ophrys aveyronensis* is now rather considered as a
112 complex consisting of two subspecies (*O. aveyronensis* subsp. *aveyronensis* and *Ophrys*
113 *aveyronensis* subsp. *vitorica*) that share a mutual insect pollinator (*i.e.* the solitary bee
114 *Andrena hattorfiana*, Paulus & Gack, 1999; Paulus, 2017; Benito Ayuso, 2019 (but see Gibert
115 et al., 2022 reporting subtle differences in morphometry and coloration). This taxon is thus
116 present in two rather restricted ranges separated by a gap of about 600 km (and Pyrenees
117 Mountains), in the Grands Causse region (France) and in Northern Spain.

118 The formation of the Pyrenees starting during the Eocene (from 45 Myrs ago) cannot
119 explain by itself the split of the ancestral range of the species as all phylogenetic hypotheses
120 published so far conclude that the *Ophrys* genus emerged no more than 5-6 Myrs ago
121 (Breitkopf et al., 2015, see also Inda et al., 2012). This does not rule out 'natural' allopatry as
122 factor to explain current disjunct geographic distribution but suggests that it would have

occurred more recently. According to this "vicariance hypothesis", the current disjunct geographic distribution may correspond to two relict pockets of an ancestral range that got fragmented following a scenario according to which populations from Northern and Southern ranges diverged in allopatry. In this case, the splitting time between northern and southern populations is expected to be relatively old and current levels of genetic diversity are not necessary expected to vary among populations. Alternatively, we also envisage recent successful colonization from one historical range to the other. This event could have occurred 'naturally'. The notoriously small *Ophrys* seeds are likely to be wind-dispersed over hundreds of kilometers and may have found in the newly colonized area suitable conditions for their germination and development. In this case, it may be difficult to provide a dating of splitting time, but the levels of genetic diversity are expected to be lower in the newly colonized area, as the recently founded *Ophrys* population is only a subset of the genetic pool of the origin population. The colonization of one part of the range from the other may also have occurred 'non-naturally' (namely, because of anthropogenic activities). In all cases, the environmental conditions are expected to be highly similar in the two regions.

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

Materials and Methods

Sampling and study sites

147 We used a minimally destructive sampling method to collect a total of 86 plant tissue
148 samples from six localities that span the entire geographic range of the *O. aveyronensis*
149 species complex: three in Southern France (Grands Causses region): 1-Guilhaumard, 2-
150 Lapanouse-de-Cernon and 3-Saint-Affrique, and three in Northern Spain: 4-Valgañón, 5-
151 Larraona and 6-Bercedo, in June 2019, Fig. 1, see also Gibert et al., 2022). At each sampling
152 site, we collected 6 to 10, 0.6 mm diameter leaf punches from 13 to 15 individuals which
153 were stored in 100% ethanol until DNA extraction (see Table 1 and Supplementary Table S1).
154 As *Ophrys aveyronensis* is a nationally protected plant species in France, sampling was done
155 under permit 'Arrêté préfectoral n°2019-s-16' issued by the 'Direction Régionale de
156 l'Environnement de l'Aménagement et du Logement (DREAL)' from the 'Région Occitanie',
157 on 07-May-2019). *Ophrys aveyronensis* is not legally protected in Spain.
158
159 **DNA extraction, library preparation and genotyping**
160 Genomic DNA extraction as well as library preparation and genotyping were subcontracted
161 to LGC Genomics GmbH (Berlin, Germany). A fractional genome sequencing strategy called
162 normalized Genotyping-by-Sequencing (nGBS) was used to subsample the *Ophrys*
163 *aveyronensis* genome for which we estimated a relatively high size of ~5-6 Gbp based on
164 flow cytometry techniques (Bertrand, unpublished data). The nGBS protocol makes use of a
165 combination of two restriction enzymes (PstI and ApeKI, in our case) to produce a
166 reproducible set of fragments across samples/individuals. It also comprises a normalization
167 step whose aim is to avoid further sequencing of highly repetitive regions. Illumina
168 technology was then used to aim at obtaining a minimum of 1.5 million reads per sample
169 (with 2 x 150 bp read length) for each of the 86 individually barcoded libraries.
170

171 **Population genomic dataset processing**

172 We used different scripts included in *Stacks* v.2.60 (Catchen et al., 2011; 2013) to build loci
173 from Illumina reads, *de novo* (*i.e.* without aligning reads to a reference genome). We first
174 used *process_radtags* to demultiplex and clean reads, then *denovo_map.pl* to build loci
175 within individuals, create a catalog and match all samples against it and finally *populations* to
176 further filter the SNPs obtained at the population level and compute basic population
177 genetic statistics. Before running the pipeline on the complete dataset, we first optimized
178 several key parameters: *-m* (the minimum number of identical raw reads required to form a
179 putative allele), *-M* (the number of mismatches allowed between alleles to form a locus) and
180 *-n* (the number of mismatches allowed between loci during construction of the catalog) by
181 running it on a subset of 12 individuals representative of the whole data set (*i.e.* geographic
182 origin and coverage). As recommended by several authors, we varied *M* and *n* from 1 to 9
183 (fixing *M* = *n*) while keeping *m* = 3 (see Paris et al., 2017). The combination *-m* 3, *-M* 5 and *-n*
184 5 was found to be the most suitable to maximize the number of SNPs, assembled and
185 polymorphic loci in our case and was used to run *Stacks* on the whole data set (see
186 Supplementary Table S2). Then, we ran the *populations* script included in *Stacks*, from which
187 only loci that were genotyped in at least 80% of the individuals per population (*-r* 80) and
188 found in all six sampling sites were kept (*-p* 6), while loci exhibiting estimates of observed
189 heterozygosity greater than 70% (*--max-obs-het* 0.7) were filtered out to reduce the risk of
190 including remaining paralogs. We also discarded sites whose minor allele frequency was
191 lower than 5% (*--min-maf* 0.05). The genotype matrix of the remaining 9301 SNPs was then
192 exported in several formats for downstream analyses.

193

194 **Genetic diversity and population differentiation**

195 To get an overview of the overall genetic diversity and differentiation among individuals and
196 populations, we first performed a Principal Component Analysis (PCA) based on a matrix of
197 86 individuals (as rows) and 9301 SNPs (as columns) coded as a *genlight* object with the R-
198 package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011). The number of private alleles
199 and the average nucleotide diversity (π) were directly available from the output of the
200 *populations* script from *Stacks*. We used Genodive v.3.04 (Meirmans, 2020) to compute
201 expected and observed heterozygosity (H_E and H_O, respectively) as well as to evaluate the
202 deviation from panmixia by computing G_{IS}. Allelic richness (A_R) was computed with the R-
203 package *hierfstat* (Weir & Goudet, 2017). Overall and pairwise genetic differentiation was
204 also assessed based on *G*-statistics (G_{ST}, G'_{ST}, G''_{ST}) as well as Jost's *D* also implemented in
205 Genodive. An Analysis of Molecular Variance (AMOVA) allowed to assess the proportion of
206 genetic variance at different hierarchical levels, especially between sampling sites among
207 subspecies/country (F_{SC}) and between subspecies/country (F_{CT}). The statistical significance of
208 the obtained values was estimated based on 10000 permutations.

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

217

218 **Demographic and evolutionary history reconstruction**

219 We reconstructed the past demographic history of each of the six populations, separately,
220 using STAIRWAYPLOT2 (Liu & Fu, 2015; 2020). This program relies on the Site Frequency
221 Spectrum (SFS) to infer the temporal dynamics of effective population size (N_e) and has the
222 advantage of not requiring whole-genome sequence data or reference genome information.
223 First, we used the *easySFS.py* script (https://github.com/isacovercast/easySFS) to compute
224 the folded SFS of each population. To maximize the number of segregating sites, the
225 population was downsampled to 24, 24, 22, 24, 24 and 22 individuals for 1-Guilhaumard, 2-
226 Lapanouse, 3-St-Affrique, 4-Valgañón, 5-Larraona and 6-Bercedo haploid genomes,
227 respectively. We then ran STAIRWAYPLOT2 that fits a flexible multi-epoch demographic
228 model that estimates N_e at different time periods using the SFS to infer estimated N_e
229 fluctuation coinciding with coalescent events. We assumed a generation time of 5 years for
230 *Ophrys* and a mutation rate of 7×10^{-9} per site per generation for angiosperms autosomal
231 markers (following Krasovec et al., 2018) and performed 200 bootstrap replicates to
232 estimate 95% confidence intervals.

233 We then used the Approximate Bayesian Computation (ABC) framework
234 implemented in DILS (Fraïsse et al., 2021) to infer the best demo-genomic scenario and
235 associated parameters for each pair of *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica*
236 populations. The two-population model of DILS compares different scenarios in a
237 hierarchical manner. First, current isolation models, *i.e.* Strict Isolation (SI) versus Ancient
238 Migration (AM) are compared to ongoing migration models, *i.e.* Isolation Migration (IM)
239 versus Secondary Contact (SC). Then, for the best demographic model, DILS compares
240 genomic models to test for background selection by assuming variation in effective
241 population size (N_e) across sites (*i.e.* hetero versus homo *N*e models are compared). For
242 migration models (IM and SC), DILS also tests for selection against migrants by assuming

243 variation in migration rates *N.m* across sites (*i.e.* hetero versus homo *N.m* models are
244 compared). Depending on the best selected model, DILS provides estimates and confident
245 intervals for parameters such as the time of split (T_{split}), the migration rate (*N.m*), current and
246 ancestral effective population sizes ($Ne_{current}$ and Ne_{past}) and whenever relevant, the time of
247 ancient migration (T_{AM}) or the time of secondary contact (T_{SC}). We formatted the input file
248 with a custom Perl script and used the online facility (http://dils.univ-lyon1.fr/) to run
249 computations with the following settings and priors: *population_growth*: variable,
250 *modeBarrier*: bimodal (*i.e.* there is a class of species barrier loci with *N.m*=0 and a class of
251 loci migrating at the background rate *N.m*) for model type to test), *max_N_tolerated:* 0.25*,*
252 *LMin*: 10, *nMin*: 8, *mu*= 7.10^{-9}, *rho_over_theta*: 0.2, *N*_min = 0, *N_max*: 200000, *Tsplit_min*:
253 0, *Tsplit_max*: 100000, *M_min*: 1 and *M_max*: 40. As the first step of DILS comparing the
254 models 'with' *versus* 'without' current migration did not result in strong statistical support
255 for either model, we estimated the parameters of the strict isolation model based on our
256 knowledge of current biogeography. This parameter estimation was carried out using the
257 same prior distribution as described above. Finally, to quantify the fit of the estimated model
258 to the data we conducted a goodness-of-fit test using 2000 simulations performed under the
259 SI model based on the estimated parameter values.
260

261 **Past and current Species Distribution Modelling**

262 We used an Ecological Niche Modelling (or Environmental Niche Modelling, ENM or Species
263 Distribution Modelling, SDM) framework similar to the one we followed and described in
264 detail in Salvado et al., (2022) to infer the spatio-temporal dynamics of suitable habitat for
265 the *O. aveyronensis* species complex since the Last Glacial Maximum (LGM, 21 kyrs BP). We
266 then compared the results with those obtained from population genomic analyses. To model

267 and spatially project the current bioclimatic niche, we first downloaded 19 bioclimatic layers
268 at a resolution of 2.5 arc-minutes (~5 km^2) available from WorldClim 2 (Fick & Hijmans,
269 2017). Current climate data correspond to time averaged variables over the period 1970-
270 2000. To spatially project the inferred niche onto past bioclimatic conditions data, we then
271 downloaded paleoclimate data either from WorldClim v1.4 (Hijmans et al., 2005) and from
272 PaleoClim (www.paleoclim.org).

273 Paleoclimate data from WorldClim are available for both the LGM (21000 kyrs ago)
274 and the Mid-Holocene (about 6000 years ago), downscaled at a resolution of 2.5 arc-minutes
275 for at least three General Circulation Models (GCMs) proposed by the National Center for
276 Atmospheric Research (NCAR) though the Community Climate System Model (CCSM4), the
277 Japan Agency for Marine-Earth Science and Technology, the Center for Climate System
278 Research of the University of Tokyo, and the National Institute for Environmental Studies
279 (MIROC-ESM) and the Max Plant Institute for Meteorology Earth System Model (MPI-ESM-
280 P). Only CCSM data are available from PaleoClim but over different time periods over the
281 Pleistocene (at a resolution of 2.5 arc-minutes): the LGM (as from CHELSA, Karger et al.,
282 2021), the Heinrich Stadial1 (17.0-14.7 kyrs BP), the Bølling-Allerød (14.7-12.9 kyrs BP) the
283 Younger Dryas Stadial (12.9-11.7 kyrs BP) and the Holocene: early-Holocene, Greenlandian
284 (11.7-8.326 kyrs BP), mid-Holocene, Northgrippian (8.326-4.2 ka BP) and late-Holocene,
285 Meghalayan (4.2-0.3 kyrs BP) (as from Fordham et al., 2017).

286 The *O. aveyronensis* occurrences dataset consists of a set of field observations as well
287 as from records downloaded from the gbif (www.gbif.org) and the iNaturalist
288 (www.inaturalist.org) databases before manual curation. The Ecological Niche Modelling
289 analyses were conducted thanks to the R-package *ENMwizard v.0.3.7* (Heming et al., 2018,
290 see also Bagley et al. 2020) which is a convenient wrapper of several tools. We first spatially

filtered occurrences to keep only those that were at least 5 km away from each other using the R-package *spThin* (Aiello-Lammens et al., 2015). The calibration area for the models was created as a buffer of 0.5° around the minimum convex polygon encompassing all occurrences. From the 19 bioclimatic variables, we selected the less correlated ones (Person correlation coefficient < 0.75) thanks to the R-package *caret* (Kuhn, 2019) and kept 7 variables: bio2 (the mean diurnal range which is the mean of monthly maximum temperature minus the mean of monthly minimum temperature), bio3 (the isothermality, *i.e.* bio2 divided by the temperature annual range), bio8 (the mean temperature of the wettest quarter), bio9 (the mean temperature of the driest quarter), bio10 (the mean temperature of the warmest quarter), bio15 (the precipitation seasonality, *i.e.* its coefficient of variation) and bio17 (the precipitation amount of the driest quarter) for further analyses.

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

315 ## Results

316 **Population genomic dataset**

317 We obtained a total of 150 694 350 of read pairs across the 86 individuals (1 382 768 - 7 370
318 366 of raw reads per ind., mean 3 504 519, SD = 1 282011.52) of which 290 431 210
319 barcoded reads (*i.e.* 96.36%) were retained, 10 819 014 (3.6%) were removed because RAD
320 cutsite was not found in the sequence and 138 476 reads (< 1%) were removed because of
321 low quality. We genotyped a total of 859 035 *loci* (composed of 135 255 870 sites) including
322 368 712 variant sites (SNPs) with *Stacks*. Average read depth ranged from 24.1X to 81.6X
323 (mean 48.7X, SD = 12.4X) based on the combination of parameters we used. After filtering
324 the data with *populations*, we finally kept 4302 *loci* from which we retained 9 301 variant
325 sites (SNPs).

326

327 **Genetic diversity and population differentiation**

328 The PCA biplot shows that individuals are arranged by subspecies/country of origin along
329 PC1 (which explains 8.50% of the total genetic variance) and by sampling site along PC2
330 (which explains 3.88% of the total genetic variance) even though the sampling sites of 1-
331 Guilhaumard and 2-Lapanouse-de-Cernon slightly overlap (Fig. 2A). Based on PC2, intra-
332 population genetic variation seems minimal for 1-Guilhaumard and 2-Lapanouse-de-Cernon
333 and maximal for 3-Saint-Affrique and 5-Larraona. Basic population genetics statistics are
334 indicated in Table 1. All populations displayed weak but significant deviation from panmixia
335 with a mean G_{IS} = 0.060 (p < 0.01, 95% CI: 0.056-0.064) and population G_{IS} values ranging
336 from 0.006 in 3-Saint-Affrique to 0.081 in 4-Valgañón (all p < 0.01). Average nucleotide
337 diversity (π) varies from 0.269 in 3-Saint-Affrique to 0.290 in 4-Valgañón and 5-Larraona, and
338 allelic richness (A_R) from 16696.70 (in 3-Saint-Affrique) to 17297.27 (in 5-Larraona). Based on

339 the filtering procedure we followed, the number of private alleles was relatively low but
340 consistently higher for *O. a.* subsp. *vitorica* (*i.e.* from 9 to 23) compared to *O. a.* subsp.
341 *aveyronensis* (*i.e.* from 0 to 2).

342 Overall *G*-statistics show that the whole set of populations is genetically structured:
343 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}
344 = 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).
345 The AMOVA confirms that although a major part of the genetic variance is found within
346 individual (81.8%, F_{IT} = 0.182, 95% CI: 0.177-0.186), significant proportions of the total
347 genetic variance are found among populations within subspecies/country (6.8%, F_{SC} = 0.072,
348 95% CI: 0.070-0.074) and between subspecies/country (6.1%, F_{CT} = 0.061, F_{CT} = 0.058-0.064).
349 Pairwise differentiation between sampling sites varies from F_{ST} = 0.049 to 0.082 within
350 subspecies/country and reaches F_{ST} = 0.111 to 0.147 between subspecies/country of origin
351 (all $p < 0.01$; see Supplementary Appendix S3 for details).

352 The clustering analysis carried out with sNMF suggests an optimal number of genetic
353 clusters of *K*=5 based on the cross-entropy criterion (Fig. 2B). We plotted the *Q*-matrices
354 (that contain individual admixture coefficients) corresponding to the best replicate runs for
355 *K*=2 (to investigate whether the program could distinguish the two subspecies) and *K*=5 (Fig.
356 2C). For *K*=2, the individuals of *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* were
357 unambiguously assigned to two distinct clusters. For the most likely combination of *K*=5, we
358 observe one cluster per sampling site except for the geographically close localities of 1-
359 Guilhaumard and 2-Lapanouse-de-Cernon that were grouped together in a unique cluster. In
360 the locality of 3-Saint-Affrique 4 out of 13 individuals (19-Oav-039, 19-Oav-040, 19-Oav-049
361 and 19-Oav-051) display >50% of ancestry proportion that are associated to the cluster
362 associated with 1-Guilhaumard and 2-Lapanouse-de-Cernon.

363

364 **Inference of past demographic history**

365 Reconstruction of historical dynamics of effective population size using STAIRWAYPLOT2
366 shows congruent trends consisting of progressive >95% decline in estimated N_e for all 6
367 populations, since the LGM, or at least, over the last 25000 years (see Fig. 3). Current
368 effective population sizes were estimated to vary between 302 (2-Lapanouse-de-Cernon)
369 and 642 individuals (3-St-Affrique) with this method. Despite the current geographical
370 discontinuity between *O. a.* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* populations, we
371 attempted to statistically compare speciation patterns with and without ongoing migration
372 using DILS. For each of the 9 trans-Pyrenean pairs tested, the Posterior Probabilities
373 associated with the models with ongoing migration vary between 0.45 and 0.69, which
374 represent cases of inferential ambiguities that do not allow strong conclusions to be drawn
375 from model comparisons with DILS. This can happen when the real history is an intermediate
376 model, or when the separation time between *O. a.* subsp. *aveyronensis* and *O. a.* subsp.
377 *vitorica* is recent compared to the current population sizes. Thus, based on biogeographical
378 knowledge of the species, we estimate demographic parameters assuming a strict isolation
379 model after the subdivision of the common ancestor into two daughter populations. The 9
380 parameter estimates made under the strict isolation model indicate similar scenarios: the
381 subdivision of an ancestral population of about 120 000 individuals (from 103916 to 129860
382 among the 9 analyses) into two daughter populations about 3060 generations ago (from
383 2050 to 6884 generations among the 9 analyses, which convert into 6150 to 20652 years
384 assuming a generation time of 3 years, 10250 to 33420 years assuming a generation time of
385 5 years and 14350 to 48188 years assuming a generation time of 7 years). This subdivision is
386 indeed considered recent as it is about six times smaller than current population sizes. This

387 suggests that most coalescence events are taking place more in the ancestral population and
388 not in the respective populations. Similar to what was inferred with STAIWAYPLOT2, our
389 analyses also point to a strong demographic decline. At the time of subdivision, we infer
390 population sizes for the *O. a.* subsp. *aveyronensis* lineages of about 13% of ancestral size,
391 and 15% of ancestral size for the *O. a.* subsp. *vitorica* lineages. On average, the sizes of the
392 current *O. a.* subsp. *aveyronensis* populations correspond to 15% of the ancestral
393 population, while the current *O. a.* subsp. *vitorica* sizes correspond to 28% of the ancestral
394 size. The goodness-of-fit test indicates that this estimated scenario reproduces the observed
395 data set with high fidelity (see Supplementary Appendix S4).
396
397 **Current and past Species Distribution Modelling**
398 ENMeval analyses identified RM = 1 and a mixture of Product (P), Quadratic (Q) and Hinge
399 (H) feature classes as the best-performing parameters for calibrating the final ENM. These
400 parameters yielded a single "best" candidate model with the lowest AICc score (925.02) and
401 an AICc weight of ~0.23. This model had mean omission rates of 0.165 (for the 10th
402 percentile) and a mean test AUC of 0.834 (see Supplementary Table S5, for details). With a
403 contribution of 41.65%, bio15 (the precipitation seasonality, *i.e.*, its coefficient of variation)
404 was found to be the bioclimatic variable the more prone to explain the *O. aveyronensis*
405 species complex distribution. It was followed by bio17 (the precipitation amount of the
406 driest quarter) with a contribution of 23.34% and bio10 (the mean temperature of the
407 warmest quarter), with a contribution of 11.31%, bio2 (the mean diurnal range) with 8.14%,
408 bio3 (the isothermality), with 5.32% and to a lesser extent bio8 (the mean temperature of
409 the wettest quarter) and bio9 (the mean temperature of the driest quarter) with 3.21% and
410 3.01%, respectively.

411 The spatial projection of the current bioclimatic inferred for the *O. aveyronensis* species complex corresponds to its actual geographic distribution but show that some additional areas may be climatically suitable for this orchid (Fig. 5). Spatial projections considering paleoclimatic conditions show that the extent of suitable habitat for the *O. aveyronensis* species complex may have reached its apex during the Younger Dryas cold period (12.9-11.7 kyrs BP) before starting to decrease across the Holocene. The LGM conditions (21 kyrs BP) were not suitable for the establishment of viable populations in the studied area but important suitable habitats may have existed during the Heinrich Stadial 1 event (17-14.7 kyrs BP) and the Bøllering-Allerød warming (14.7 – 12.9 kyrs BP). Altogether, these results may be consistent with the 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).

426

427 **Discussion**

428 *Ophrys* orchids exhibit among the higher rate of speciation reported so far (Breitkopf et al., 429 2015) and are being increasingly mentioned as biological model of interest to investigate 430 ecological speciation and adaptive radiations (e.g. Nosil et al., 2017; Baguette et al., 2020; 431 Nürk et al., 2020). The selection pressures exerted by pollinators and the high degree of 432 specificity necessary to ensure the success of the sexual swindling strategy certainly explains 433 the overwhelming importance of ecological and sometimes sympatric speciation in *Ophrys* 434 (see Baguette et al., 2020). This does not preclude for a significant role of allopatric

435 speciation in *Ophrys*, even in non-insular systems (see Bertrand et al., 2021a). Our study, is
436 to our knowledge, the first to document a scenario of allopatric speciation in a continental
437 system event if divergence is probably too recent to accurately date the actual cessation of
438 gene flow. The *Ophrys aveyronensis* species complex therefore provide a promising
439 opportunity to investigate the relative contribution of biogeographic and ecological factors
440 in shaping evolutionary divergence in rapidly and intensely diversifying groups of organisms.
441

442 **Patterns of genetic diversity and differentiation based on population genomics**

443 Although distinguishing *Ophrys aveyronensis* subsp. *aveyronensis* and *O. a.* subsp. *vitorica* is
444 difficult but still overall possible based on phenotypic data (Gibert et al., 2022), the French
445 and Iberian counterparts of the *O. aveyronensis* complex can be unambiguously teased apart
446 based on population genomics data. Both the PCA, *F*-statistics (and/or their analogues) and
447 the clustering analysis confirm that the two subspecies form distinct genetic entities,
448 themselves exhibiting a genetic substructure consistent with geography (sampling sites).
449 Altogether, these results suggest that each sampled population of *O. aveyronensis* is more
450 genetically differentiated from populations from the other subspecies than from populations
451 of its own subspecies. This pattern *a priori* rules out a recent colonization and expansion
452 from one range to the other from a unique source population. Patterns of genetic diversity
453 also support this conclusion. We found that all the populations sampled displayed relatively
454 similar levels of allelic richness, heterozygosity, departure from panmixia and number of
455 private alleles. There is therefore no evidence for any of these six populations to appear as a
456 genetic subsample of the genetic diversity of a putative source population elsewhere.
457 Finally, the inference of the temporal dynamics of effective population sizes suggests that
458 each sampled population may be considered as kind of replicate of each other. They all

459 displayed consistent decreases in effective population size over the last tens of thousands of
460 years suggesting that they shared a common demographic history.

461 It has also been proposed that *Ophrys aveyronensis* may have a hybrid origin and
462 Delforge (2016) hypothesized that the Iberian *O. a.* subsp. *vitorica* may result from a hybrid
463 swarm (without mentioning the putative parental species). Our results do not support the
464 later hypothesis that consider that the two subspecies would have emerged from distinct
465 hybridization events. However, the demographic scenarios we tested remain relatively
466 simple and the Bayesian Posterior Probabilities we obtained with DILS show that the
467 approach implemented in this program could not strongly support the Ongoing Migration
468 over a Strict Isolation model. Events that are known to deceives demographic inferences
469 such as 'ghost introgression' from an unsampled lineage have a non-negligible probability in
470 *Ophrys*. Investigating whether the *Ophrys aveyronensis* complex has or not a hybrid origin
471 and at least whether the two taxa have been introgressed by other *Ophrys* taxa will thus
472 require the genomic analyses of additional species in a future study, including more
473 informative data such as transcriptomes.

474

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

476 The Ecological Niche Modelling results suggest that the LGM conditions were not suitable for
477 the establishment of viable populations of the *O. aveyronensis* species complex. This
478 suggests that either this species complex diversified from an ancestral *Ophrys* lineage in the
479 study area after this period, or that the populations of the complex occupied a relatively
480 remote glacial refugia during this period. Over the last 21000 years, the extent of the habitat
481 suitable for *O. aveyronensis* was rather large until the Holocene (about 12000 years ago)
482 (Fig. 5) where its area started to dramatically decrease while keeping an area of relatively

483 high suitability centered on the Southwest of France. The split and, to a lesser extent, the
484 shift Northward and Southward may have occurred during the Late Holocene (starting about
485 4000 years ago) (Fig. 5). This timing is on the same order of magnitude than the time of split
486 we inferred based on population genomic data and is consistent with a scenario of allopatric
487 divergence (without being able to fully exclude transient gene flow through climate-induced
488 oscillations) from an ancestral population with relatively high effective size.
489

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

491 Selection pressures exerted by insect pollinators are likely to be the main driver of ecological
492 and putatively sympatric speciation in *Ophrys* orchids (see Baguette et al., 2020). In this way,
493 populations of a given *Ophrys* species being pollinated by distinct insect pollinator species
494 may be considered as engaged in a speciation process. So far, field observations support that
495 *O. aveyronensis* from the Grands Causses and their Iberian counterparts are pollinated by
496 the same pollinator species: the solitary bee *Andrena hattorfiana* (see Benito Ayuso, 2019).
497 In a recent study on this system, however, we showed that *O. aveyronensis* from the Grands
498 Causses and from the Iberian Peninsula present subtle differences at a set of morphological
499 and coloration traits that may be involved in pollination (Gibert et al., 2022). Although
500 additional work is required to address whether this variation may be adaptive and involved
501 in evolutionary divergence, these results call for an assessment of i) which phenotypic traits
502 may be under divergent selection across the *O. aveyronensis* species complex geographic
503 distribution and ii) an investigation of putative genomic regions associated with phenotypic
504 trait values, genetic structure or ecological features, something that remains largely
505 underexplored in *Ophrys* (but see Sedeek et al., 2014, for a noticeable exception).
506

507 **Taxonomic and conservation implications**

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

519 Recent studies elaborated from some empirical evidence that delineating the so-
520 called 'micro-species' in *Ophrys* is practically cumbersome (Bateman et al., 2021; Bateman
521 and Rudall, 2023), especially in the section *Sphegodes* to which the *O. aveyronensis* species
522 complex belongs. Some genomic data such as whole plastomes that are useful to delineate
523 the main *Ophrys lineages*, present sequence similarity degree of more than 99.5% between
524 'microspecies' and therefore do not present enough variation to allow to discriminate them
525 (Bertrand et al., 2019; Bertrand et al., 2021b; Bateman et al., 2021). Here, however, our
526 results demonstrate that an approach consisting of sampling populations (*i.e.* a sufficient
527 number of individuals at a given location) and appropriate genetic markers (e.g. SNPs
528 genotyped on populations based on RAD-seq protocols) give signals that are consistent with
529 taxonomy. We therefore claim that such approaches of population genomics within an
530 integrative taxonomy framework (relying on flower morphology, pollinator identity,

ecological features…) is by far the most fruitful to better understand the process of evolutionary divergence and the most relevant to improve at the same time our knowledge on taxonomy, systematics, ecology and evolution of *Ophrys* orchids.

Author contributions

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

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

Data availability statement

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

Acknowledgements

This work was primarily supported by an inter-LabEX TULIP-CEMEB initiative (n°197310) to Joris Bertrand and Bertrand Schatz. This research was also funded by an Agence Nationale pour la Recherche Jeune Chercheur Jeune Chercheuse (ANR JCJC) grant to Joris Bertrand, grant number ANR-21-CE02-0022-01, and is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP [ANR-10- LABX-41]. Lastly, this research was also funded by the Observatoire de REcherche Méditerranéen de l'Environnement (SO Ocove & SO PolliMed, OSU OREME) to Bertrand Schatz. We thank Jean-Luc Roux and Daniel Vizcaïno for the

554 information provided about sampling sites in Spain. We thank Christian Moliné for his
555 contribution to fieldwork.

556

557 **Conflict of interest**

558 The authors declare no conflict of interest

559

560 References

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

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

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

571 Bateman, R.M., Rudall, P. J., Murphy, A.R.M., Cowan, R.S., Devey, D.S. & Peréz-Escobar, O.A.
572 (2021) Whole plastomes are not enough: phylogenomic and morphometric
573 exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect.
574 *Sphegodes*. *Journal of Experimental Botany*, 72, 654-681. doi: 10.1093/jxb/eraa467

575 Bateman, R.M. & Rudall, P.J. (2023) Morphological continua make poor species: genus-wide
576 morphometric survey of the European Bee Orchids (*Ophrys* L.). *Biology*, 12, 136. doi:
577 10.3390/ biology12010136

Beatty, G.E., & Provan, J. (2013) Post-glacial dispersal, rather than *in situ* glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericacea). *Journal of Biogeography*, 40, 335-344. doi: 10.1111/j.1365-2699.2012.02789.x

Beatty, G.E., & Provan, J. (2014) Phylogeographical analysis of two cold-tolerant plants with disjunct Lusitanian distributions does not support *in situ* survival during the last glaciation. *Journal of Biogeography*, 41, 2185-2193. doi: 10.1111/jbi.12371

Beatty, G.E., Lennon, J.J., O'Sullivan, C.J., & Provan, J. (2015) The not-so-Irish spurge: *Euphorbia hyberna* (Euphorbiaceae) and the Littletonian plant 'steeplechase'. *Biological Journal of the Linnean Society*, 114, 249-259. doi: 10.1111/bij.12435

Benito Ayuso, J. (2019) Estudios sobre polinización en el género *Ophrys* (Orchidaceae), I, *Flora Montiberica*, 74, 32-37.

Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2019) Characterization of the complete plastome of *Ophrys aveyronensis*, a Euro-Mediterranan orchid with an intriguing geographic distribution. *Mitochondrial DNA Part B*, 4, 3256-3257. doi: 10.1080/23802359.2019.1670748

Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2021) Whole plastid genome-based phylogenomics supports an inner placement of the *O. insectifera* group rather than a basal position in the rapidly diversifying *Ophrys* genus (Orchidaceae). *Biology Letters*, 168, 452-457. doi: 10.1080/23818107.2021.1893216

Bertrand, J. A. M., Baguette, M., Joffard, N., & Schatz, B. (2021) Challenges inherent in the systematics and taxonomy of genera that have recently experienced explosive radiation: the case of orchids of the genus *Ophrys*. In Grancolas, P. & Maurel, M.-C. *Systematics and the Exploration of Life* (pp 113-134). Hoboken, New Jersey: Wiley.

602 Breitkopf, H., Onstein, R.E., Cafasso, D., Schlüter, P.M., & Cozzolino, S. (2015) Multiple shifts
603 to different pollinators fuelled rapid diversification in sexually deceptive *Ophrys*
604 orchids. *New Phytologist*, 207, 377-389. doi: 10.1111/nph.13219
605 Catchen, J., Amores, A., Cresko, W., & Postlewait, J. (2011) Stacks: building and genotyping
606 loci de novo from short-read sequences. *G3: genes, Genomes, Genetics*, 1, 171-182.
607 doi: 10.1534/g3.111.000240
608 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2013) Stacks: an
609 analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140. doi:
610 /10.1111/mec.12354
611 Coyne, J.A., & Orr, H.A. (2004). *Speciation*. Sunderland, Mass: Sinauer Associates.
612 Claessens, J., & Kleynen, J. (2016) *Orchidées d'Europe : fleurs et pollinisation*. Biotope, Mèze,
613 France : Biotope.
614 Davison, A., Birks, J.D.S., Brookes, R.C., Messenger, J.E., & Griffiths, H.I. (2001) Mitochondrial
615 phylogeography and population history of pine martens *Martes martes* compared
616 with polecats *Mustela putorius*. *Molecular Ecology*, 10, 2479-2488. doi: 10.1046/j.1365-
617 294x.2001.01381.x.
618 Delforge, P. (2016) *Orchidées d'Europe, d'Afrique du Nord et du Proche Orient*. Lausanne,
619 Switzerland: Delachaux & Niestlé.
620 Fagúndez, J. & Díaz-Tapia, P. (2023) Comparative phylogeography of a restricted and a
621 widespread heather: genetic evidence of multiple independent introductions of Erica
622 mackayana into Ireland from northern Spain. *Botanical Journal of the Linnean Society*.
623 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

Frichot, E., Mathieu, F., Trouillon T., Bouchard, G., & François O. (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196, 973-983. doi: 10.1534/genetics.113.160572

Frichot, E. & François O. (2015) LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925-929. doi: 10.1111/2041-210X.12382

Gibert, A., Louty, F., Buscail, R., Baguette, M., Schatz, B. &, Bertrand, J.A.M. (2022) Extracting quantitative information from images taken in the wild: a case study of two vicariants of the *Ophrys aveyronensis* species complex. *Diversity*, **14**, 400. doi: 10.3390/d14050400

Grindon, A. J., & Davison, A. (2013) Irish *Cepea nemoralis* land snails have a cryptic Franco-Iberian origin that is most easily explained by the movements of Mesolithic humans. *PLoS ONE*, 8, e65792. doi: doi.org/10.1371/journal.pone.0065792

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

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913. doi: 10.1038/35016000/

Hijmans, R. J., Cameron, S.E., Parra, J. -L., Jones, P. J., & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978. doi: 10.1002/joc.1276

Inda, L.A., Pimentel, M., & Chase, M.W. (2012) Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Annals of Botany.*, 110, 71-90. doi: 10.1093/aob/mcs083

Joffard, N., Massol, F., Grenié, M., Montgellard, C. & Schatz, B. (2019) Effect of pollination strategy, phylogeny and distribution niches of Euro-Mediterranean orchids. *Journal of Ecology*, 107, 478-490. doi: 10.1111/1365-2745.13013

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., & Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wideSNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr521

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021): CHELSA-TraCE21k v1. 0. Downscaled transient temperature and precipitation data

since the last glacial maximum. *Climate of the Past Discussions*, 1-27. doi: 10.5194/cp-2021-30

Krasovec, M., Chester, M., Ridout, K., & Filatov, D.A. (2018) The mutation rate and the age of the sex chromosomes in *Silene latifolia*. *Current Biology*, 28, 1832-1838. doi: 10.1016/j.cub.2018.04.069

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

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

Liu, X., & Fu, Y.-X. (2015) Exploring population changes using SNP frequency spectra. *Nature Genetics*, 47, 555-559. doi: 10.1038/ng.3254.

Liu, X., & Fu, Y.-X. (2020) Stairway Plot 2: demographic inference with folded SNP frequency spectra. *Genome Biology*, 21, 280. doi: 10.1186/s13059-020-02196-9

Mascheretti, S., Rogatcheva M. B., Gündüz, I., Fredga, K., & Searle, J.B. (2003) How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome *b* sequences. *Proceedings of the Royal Society B*, 270, 1593-1599. doi: doi.org/10.1098/rspb.2003.2406

Meirmans, P. G. (2020) GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, 20, 1126-1131. doi: 10.1111/1755-0998.13145

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.

Phillips, S.J., Anderson, R.P., & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231-259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., & Dudík, M. (2008). Modelling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31, 161–175. doi: doi.org/10.1111/j.0906-7590.2008.5203.x

719 Philipps, S.J., Anderson, R.P., Dudík, M., Shapire, R.E., & Blair, M.E. (2017) Opening the black
720 box: an open-source release of Maxent. *Ecography*, 40, 887-893. doi:
721 doi.org/10.1111/ecog.03049
722 Reich, I., Gormally, M., Allcock, A.L., McDonnel, R., Castillejo, J., Iglesias, J., … Smith, C.J.
723 (2015) Genetic study reveals close link between Irish and Northern Spanish
724 specimens of the protected Lusitanian slug *Geomalacus maculus*. *Biological Journal of the*
725 *Linnean Society*, 116, 156-168. doi: 10.1111/bij.12568
726 Rochette, N.C., & Catchen J.M. (2017) Deriving genotypes from RAD-seq short-read data
727 using Stacks. *Nature Protocols*, 12, 2640-2659. doi:10.1038/nprot.2017.123
728 Salvado, P., Aymerich Boixader, P., Vila Bonfill, A., Martin, M., Quélennec, C., Lewin, J.-M.,
729 Hinoux, V., & Bertrand, J.A.M. (2022) Little hope for the polyploid endemic Pyrenean
730 Larkspur (*Delphinium montanum*): evidences form population genomics and
731 Ecological Niche Modeling. *Ecology and Evolution*, 12, e8711. doi: 10.1002/ece3.8711
732 Sedeek, K.E.M., Scopece, G., Staedler, Y.M., Schönenberger, J., Cozzolino, S., Schiestl, F. P., &
733 Schlüter, P.M. (2014) Genic rather than genome-wide differences between sexually
734 deceptive *Ophrys* orchids with different pollinators. *Molecular Ecology*, **23**, 6192-
735 6205. doi: 10.1111/mec.12992
736 Weir, B.S. & Goudet, J. (2017) A unified characterization of population structure and
737 relatedness. *Genetics*, 206, 2085-2103. doi:
738 https://doi.org/10.1534/genetics.116.198424
739

740 Table 1. Sampling site, sample size (*n*), geographic coordinates and elevation, population
741 summary statistics: number of private alleles, expected and observed heterozygosity (H_E and
742 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									

743

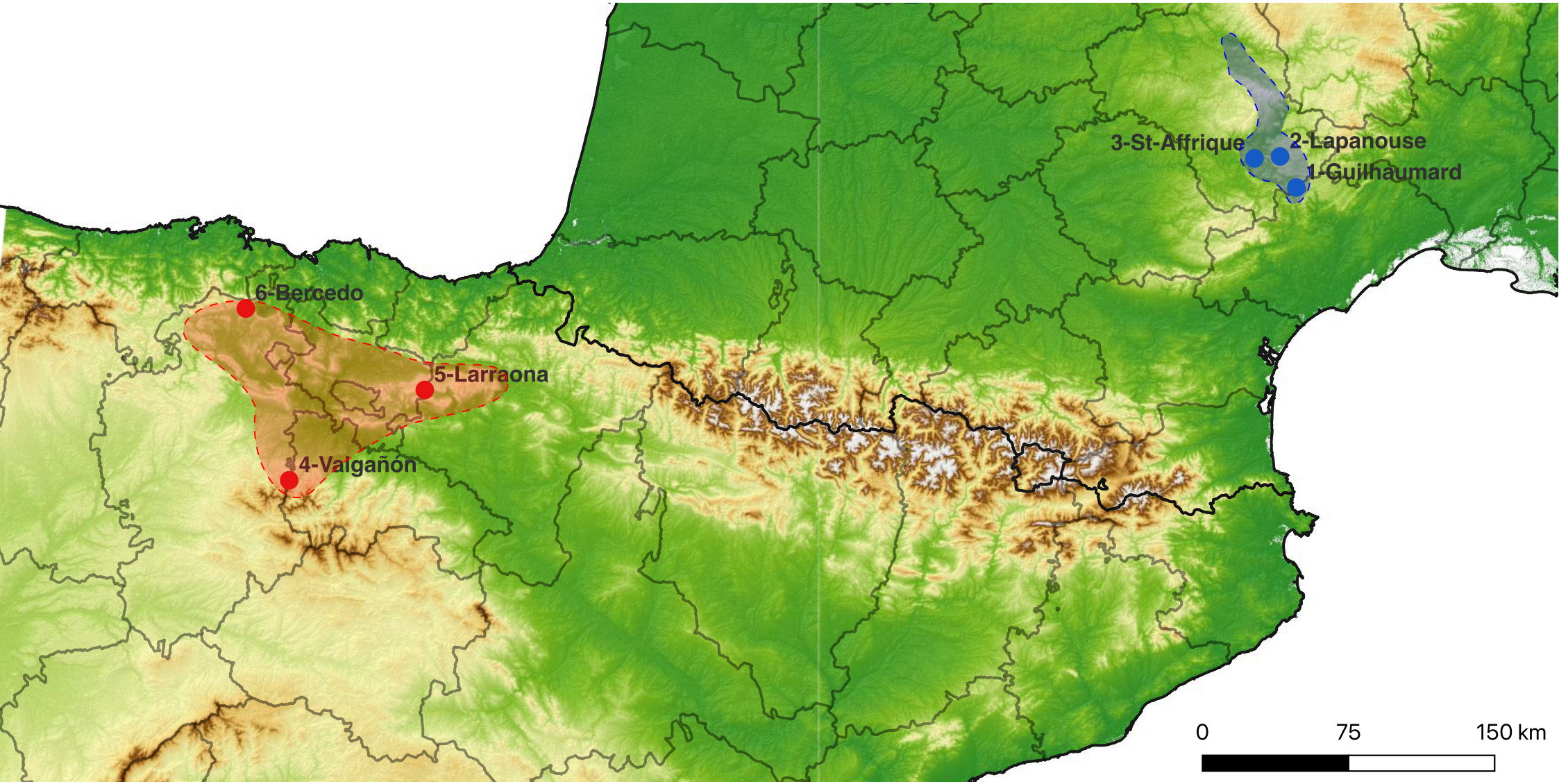
744 Table 2. Summary of the main demographic parameters inferred with DILS for each *Ophrys*
745 *aveyronensis* subsp. *aveyronensis* and *Ophrys aveyronensis* subsp. *vitorica* population pairs
746 (all details are provided in Table S4). N_e 1, N_e 2, N_e anc. correspond to effective population
747 sizes estimated for population 1, population 2 and the ancestral population, t_{split} to the time
748 of split between population 1 and population 2 in number of generations, respectively.

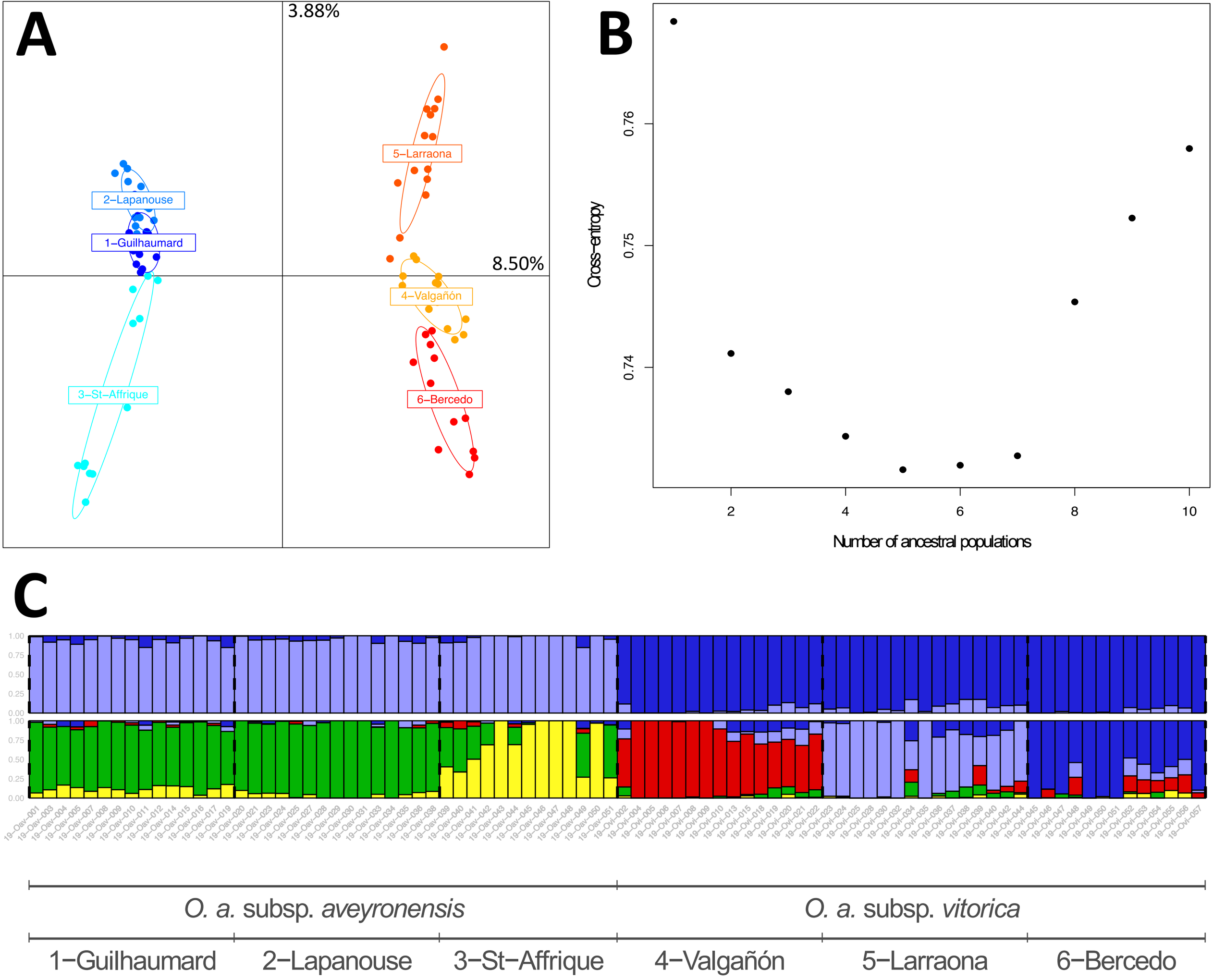
Population 1	Population 2	N_e 1	N_e 2	N_e anc.	t_{split}
Guilhaumard	Valgañón	19479	10389	129860	3837
Guilhaumard	Larraona	37410	44684	103916	2568
Guilhaumard	Bercedo	7302	26772	121691	2050
Lapanouse	Valgañón	20495	89214	120560	2476
Lapanouse	Larraona	29891	9197	114966	2853
Lapanouse	Bercedo	15855	7318	121961	3023
St-Affrique	Valgañón	16668	31253	104177	6884
St-Affrique	Larraona	4830	13282	120745	4432
St-Affrique	Bercedo	2540	66049	127017	2107

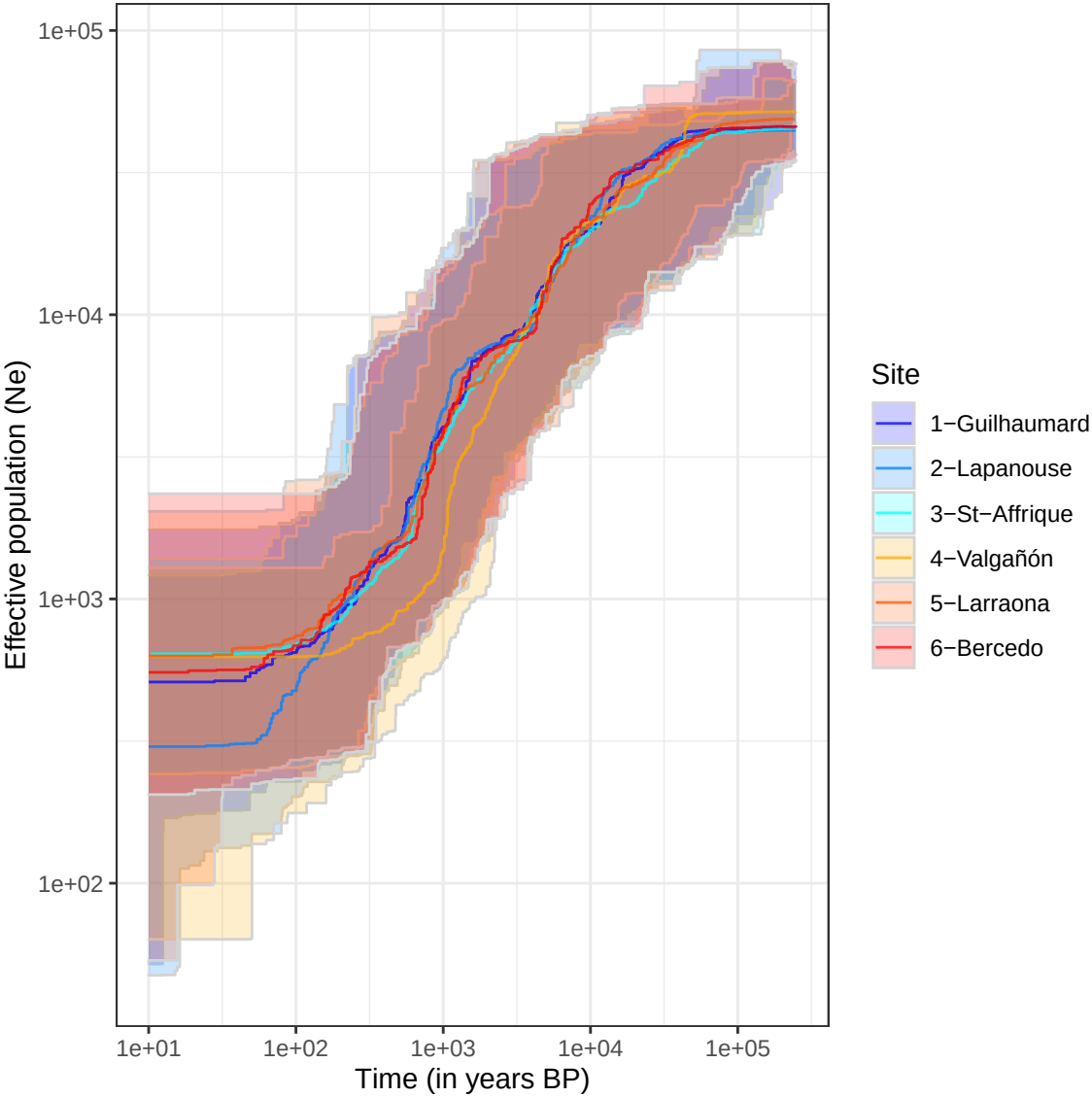
749

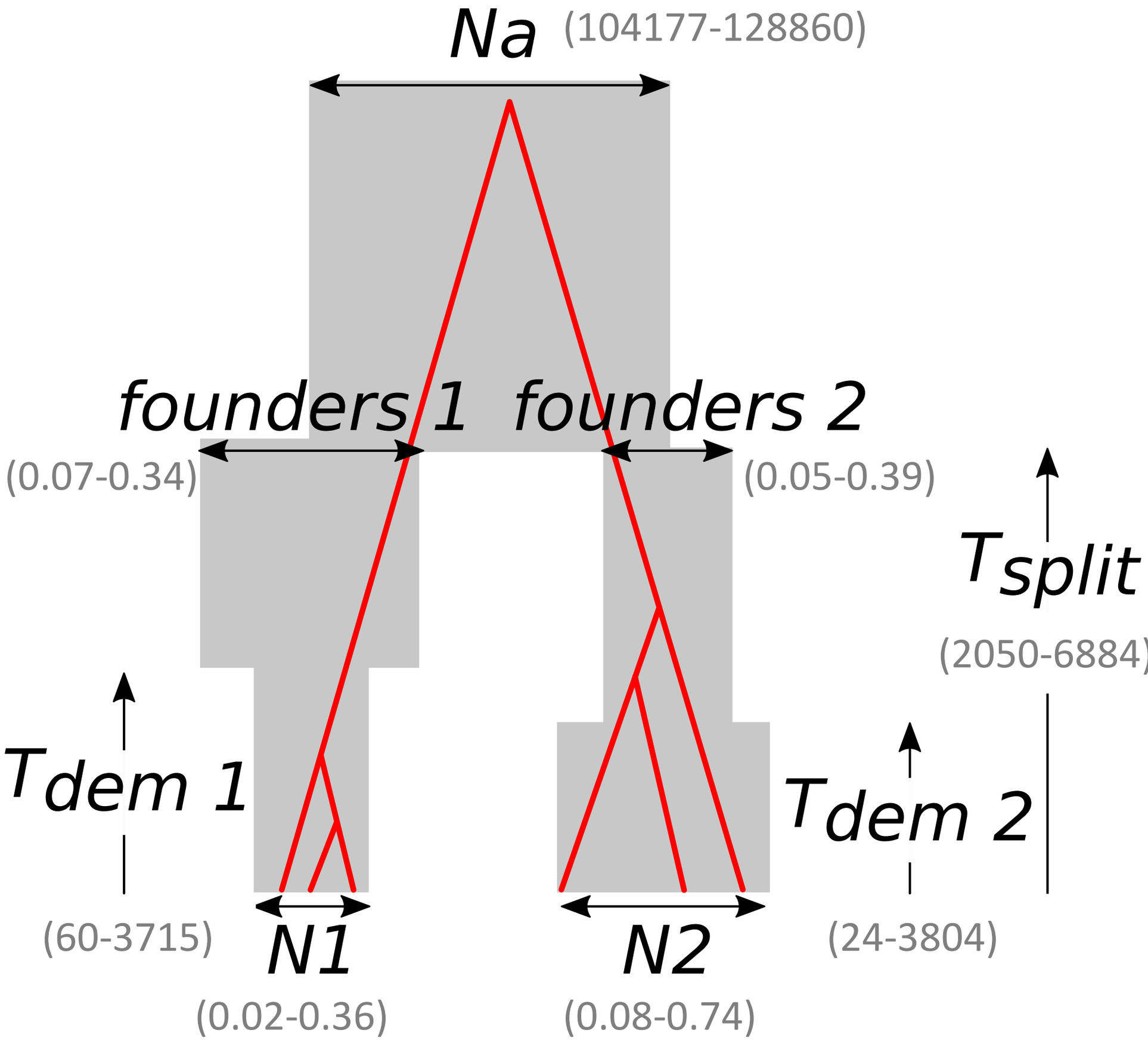
750 **Figure 1** Map displaying the 6 sites sampled: 3 for *Ophrys aveyronensis* subsp. *aveyronensis*:
751 1-Guilhaumard, 2-Lapanouse-de-Cernon, 3-St-Affrique (in blue) and 3 for *O. a.* subsp.
752 *vitorica*: 4- Valgañón, 5-Larraona, 6-Bercedo (in red). The approximate known geographic
753 distribution is highlighted for both taxa with the same color code.
754
755 **Figure 2** A) Principal Component Analysis (PCA) displaying the two first axes (PC1 and PC2)
756 representing 8.50% and 3.88% of the total genetic variance. PCA was computed based on 86
757 individuals genotyped at 9,301 SNP and colors depict sampling localities. B) Values of the
758 cross-entropy criterion for a number of clusters ranging from K = 1 to 10 (10 sNMF runs
759 each). The optimal number of K was found to be 5. C) Barplots of ancestry coefficients
760 obtained from sNMF for 86 individuals for K = 2 and K = 5.
761
762 **Figure 3** Demographic history of the 6 studied populations of *O. a.* subsp. *aveyronensis*
763 (shades of blue) and *O. a.* subsp. *vitorica* (shades of red) depicting the mean effective
764 population size N_e (and associated 95 confidence intervals) over time as inferred from
765 STAIRWAYPLOTv2, assuming a mutation rate of 7 x10^{-9} and a generation time of 5 years.
766
767 **Figure 4** Schematic overview of the retained demographic scenario and range of associated
768 parameters inferred for each pair of populations (*O. aveyronensis* subsp. *aveyronensis/ O. a.*
769 subsp. *vitorica*). N_a: ancestral effective population size, N_1: (current) effective population size
770 for the *O. a.* subsp. *aveyronensis* population (given in number of individuals), N_2: Current)
771 effective population size for the *O. a.* subsp. *vitorica*, *founders1*: effective population size of
772 *O. a.* subsp. *aveyronensis* before T_{dem1} and *founders2*: effective size of *O. a.* subsp. *vitorica*
773 before T_{dem2} (each of them given in proportion of N_a). T_{split}: time of split between the

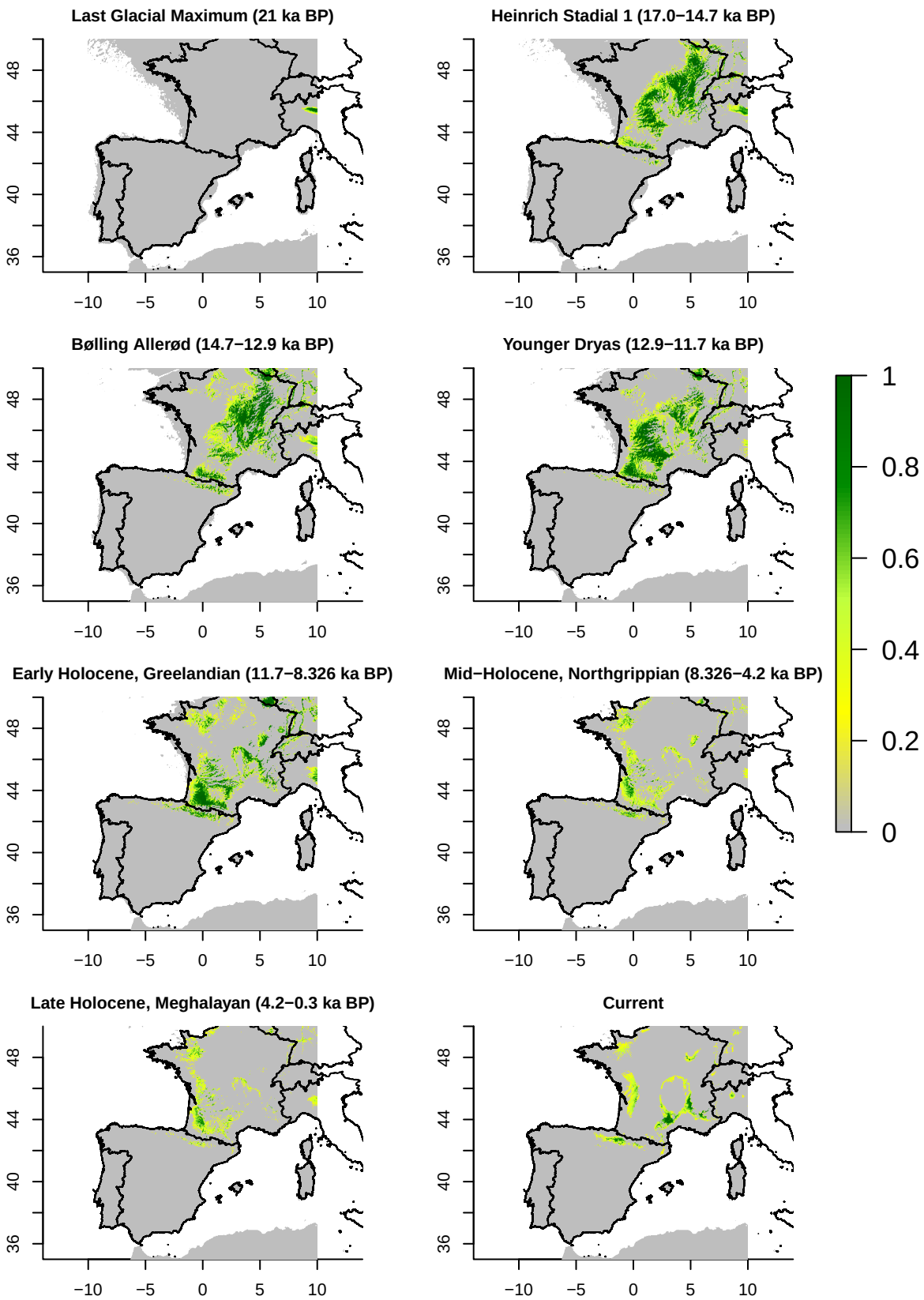
774 populations, T_{dem1}: time of *O. a.* subsp. *aveyronensis* effective population size change and
775 T_{dem2}: time of *O. a.* subsp. *vitorica* effective population size change (all in number of
776 generations).
777
778 **Figure 5** Spatial projections of the bioclimatic niches inferred for the *Ophrys aveyronensis*
779 species complex depicting climate suitability at different time periods from the Last Glacial
780 Maximum (LGM, about 21 000 kyrs ago) and with projections for the Heinrich Stadial1 (17.0-
781 14.7 kyrs BP), the Bølling-Allerød (14.7-12.9 kyrs BP) the Younger Dryas Stadial (12.9-11.7
782 kyrs BP) and the Holocene: early-Holocene, Greenlandian (11.7-8.326 kyrs BP), mid-
783 Holocene, Northgrippian (8.326-4.2 kyrs BP) and late-Holocene, Meghalayan (4.2-0.3 kyrs
784 BP) and currently.


3-St-Affrique
2-Lapanouse
1-Guilhaumard
6-Bercedo
5-Larraona
4-Valgañón
0
75
150 km


A
3.88%
8.50%
2–Lapanouse
1–Guilhaumard
3–St–Affrique
5–Larraona
4–Valgañón
6–Bercedo
B
Cross–entropy
0.74
0.75
0.76
2
4
6
8
10
Number of ancestral populations
C
1.00
0.75
0.50
0.25
0.00
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


Na
(104177-128860)
founders 1
founders 2
(0.07-0.34)
(0.05-0.39)
Tsplit
(2050-6884)
Tdem 1
Tdem 2
(60-3715)
N1
N2
(24-3804)
(0.02-0.36)
(0.08-0.74)


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