A strong east–west Mediterranean divergence supports a new phylogeographic history of the carob tree (Ceratonia siliqua , Leguminosae) and multiple domestications from native populations

Juan Viruel, Nicolas Le Galliot, Samuel Pironon, Gonzalo Nieto Feliner, Jean-Pierre Suc, Fatma Lakhal-Mirleau, Marianick Juin, Marjorie Selva, Magda Bou, Dagher Kharrat, et al.

Published in the Journal of Biogeography: https://doi.org/10.1111/jbi.13726

A strong east-west Mediterranean divergence supports a new phylogeographic history of the carob tree (*Ceratonia siliqua*, Leguminosae) and multiple domestications from native populations

Authors: Juan Viruel [1,2], Nicolas Le Galliot [1], Samuel Pironon[2], Gonzalo Nieto Feliner[3], Jean-Pierre Suc[4], Fatma Lakhal-Mirleau[1], Marianick Juin[1], Marjorie Selva[1], Magda Bou Dagher Kharrat[5], Lahcen Ouahmane[6], Stefano La Malfa[7], Katia Diadema[8], Hervé Sanguin[9, 10], Frédéric Médail[1], Alex Baumel[1*]

[1]Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), Aix Marseille Univ, Avignon Université, CNRS, IRD, Faculté des Sciences et Techniques St-Jérôme, Av. Escadrille Normandie Niémen, 13 397 Marseille cedex 20, France
[2]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, United Kingdom
[3]Real Jardín Botánico (CSIC), Plaza de Murillo 2, 28014 Madrid, Spain
[4]Sorbonne Université, CNRS-INSU, Institut des Sciences de la Terre Paris, ISTeP UMR 7193, 75005 Paris,France
[5]Laboratoire Biodiversité et Génomique Fonctionnelle, Faculté des Sciences, Université Saint-Joseph, Campus Sciences et Technologies, Mar Roukos, Mkalles, BP: 1514 Riad el Solh, Beirut 1107 2050, Lebanon
[6]Université Cadi Ayyad Marrakech, Faculté des Sciences Semlalia, Laboratoire d'Ecologie et Environnement, Morocco
[7]Università degli Studi di Catania, Dipartimento di Agricoltura, Alimentazione e Ambiente (Di3A) Via Valdisavoia 5 - 95123 Catania - Italy
[8]Conservatoire Botanique National Méditerranéen de Porquerolles (CBNMed), 34 avenue Gambetta, 83400 Hyères, France
[9]CIRAD, UMR BGPI, F-34398 Montpellier, France

[10]CIRAD, UMR LSTM, F-34398 Montpellier, France

[*]Author for correspondence: alex.baumel@imbe.fr

Abstract

Aim: Phylogeography of fruit trees is challenging due to the recurrent exchanges between domesticated and wild populations. Here we tested the eastern refugium hypothesis (ERH) for the carob tree, *Ceratonia siliqua*, which supports its natural and domestication origins in the Eastern Mediterranean and a feral origin in the West.

Location: Mediterranean basin

Taxon: *Ceratonia siliqua* L., *Leguminosae*

Methods: A phylogenetic reconstruction based on two nuclear and one plastid sequences was performed to estimate the divergence time between the carob tree and its sister species, *Ceratonia oreothauma*. Variation from four plastid regions and 17 nuclear microsatellite loci were used to decipher genetic structure in the carob tree and to test coalescent-based models by an Approximate Bayesian Computation (ABC) approach. We assessed our hypotheses by examining palaeobotanical records and hindcasting the past distribution of the carob tree at the Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG) using species distribution modelling (SDM).

Results: The split between *C. oreothauma* and *C. siliqua* was estimated at 6.4 Ma, and a first divergence within *C. siliqua* at 1.3 Ma. After a presence since the Oligocene, *Ceratonia* was found in the Western and the Eastern Mediterranean in the fossil records during the Pleistocene. Plastid and nuclear markers, characterized by low allelic richness, revealed a strong west-east genetic structuring. Approximate Bayesian computation (ABC) analyses rejected the ERH.

Main conclusions: Our study supports a severe population decline during LIG. The strong west-east divergence and the occurrence of four lineages within *C. siliqua* provided support for a new hypothesis of multiple domestications of the carob tree from native populations throughout the Mediterranean basin.

Key words: Approximate Bayesian computation ABC, carob, coalescence, domestication, origin, palaeobotany, phylogeography, species distribution modelling.

Introduction

Phylogeography and palaeobotany have greatly contributed to our knowledge on species evolutionary histories and their ecological responses to environmental changes (Hu et al., 2009). Range contractions and expansions due to the climatic oscillations during the Pleistocene have been reported in temperate tree species (Gavin et al., 2014) as well as in thermophilous Mediterranean trees (Nieto-Feliner 2014; Migliore et al., 2018). The natural distribution ranges of the latter were also influenced by the cradle of several human societies that intensively modified forests due to forage, fruit harvesting or clearing for agriculture (Quézel & Médail 2003). The domestication of fruit trees began ca. 6,000 years ago in the Eastern Mediterranean and entailed recurrent exchanges between crops and their wild relatives (Meyer et al., 2012; Zohary & Hopf 2012).

Among the Mediterranean fruit trees, the carob (*Ceratonia siliqua* L., Leguminosae) is an evergreen dioecious tree characteristic of thermophilous vegetations (Zohary 2002; Quézel & Médail 2003; Baumel et al., 2018). It has been widely exploited since antiquity, and it is currently used for industrial, agricultural and soil restoration purposes. During its domestication, genotypes with larger and sweeter fruits have been selected and propagated by grafting (Zohary 2002). Despite its economic and cultural importance in the Mediterranean and the long-standing interest of botanists (e.g. de Candolle 1883), the status of the carob tree as a native species of the West Mediterranean vegetation is still source of debate (Ramon-Laca & Mabberley 2004; Baumel et al., 2018) and a phylogeographic study was still pending.

The genus *Ceratonia* has been present in the Mediterranean for millions of years and scattered occurrences from the Oligocene to the Pliocene suggest a large ancestral distribution (e.g. Palamarev 1989) followed by a strong decline due to Pleistocene climate changes (Ramon-Laca & Mabberley 2004). Archaeological, historical and philological evidences suggest a westward human-driven dissemination of *C. siliqua* (reviewed in Ramon-Laca & Mabberley 2004). This widely accepted "eastern refugium hypothesis" (ERH) postulates a single eastern refugium and a human-driven dissemination concomitant with its domestication based on three arguments. First, the discovery of a closely related species co-occurring with the carob in Yemen, *C. oreothauma* Hillc., G.P.Lewis & Verdc. (Hillcoat et al., 1980), suggests a center of origin in the Arabian Peninsula (Zohary 2002). Second, pollen and macro-remain records support the existence of *C. siliqua* in the East at the end of the Pleistocene and Early Holocene (Zohary 2002); meanwhile studies suggesting its presence in the western part of the Mediterranean during the Pleistocene are less abundant (Brenac 1984; Zapata et al., 2013; Servera-Vives et al., 2018). Third, the western scarcity of local names attributed to the carob tree would also be consistent with a domestication origin in the East (de Candolle 1883; Zohary 2002; Ramon-Laca & Mabberley 2004). Carob tree cultivars are exclusively propagated by grafting (Zohary 2002) and therefore the history of its cultivation is closely related to the development of grafting methods ca. 3,000 years ago (Meyer et al., 2012). According to the ERH, wild carob populations in the Western Mediterranean would have a recent and feral origin. The ERH, if valid, would be supported by a phylogeographic pattern in which western populations are derived from eastern ones and divergence times are congruent with the history of Mediterranean fruit tree domestication (Meyer et al., 2012). However, past records of carob tree from the palaeobotanical literature have never been the subject of a comprehensive review. In addition, a long-standing evolutionary history in the Mediterranean that precedes the origin of the agriculture could also explain the presence of *C. siliqua* in several types of thermophilous vegetation across the Mediterranean (Baumel et al., 2018). Therefore, alternative scenarios should be tested when addressing the origin of the carob tree in the Mediterranean.

Here we present the first phylogeographic study of the carob tree across its distribution range in the Mediterranean basin to test the ERH *versus* an alternative scenario, which also has supporting independent evidences. First, we applied phylogenetic methods to estimate divergence time between the two *Ceratonia* species and to determine whether current populations of *C. siliqua* could have originated from domestication of *C. oreothauma*. Second, we investigated plastid DNA polymorphisms to document carob maternally-inherited lineages. Third, nuclear microsatellite (SSR) markers were used to identify the main carob genetic pools across the Mediterranean and analyzed under a statistical phylogeographic frame using Approximate Bayesian computation (ABC, Knowles 2009). Finally, we

assessed our phylogeographic inferences by conducting a comprehensive review of palaeobotanical information and reconstructing the potential historical range dynamics of *C. siliqua* using species distribution modelling (SDM).

Material and Methods

Plant material

Leaves of 1067 carob trees from 78 populations were sampled across the Mediterranean between 2015 and 2017. Localities were selected to cover the geographical and ecological ranges of *C. siliqua* as described in Baumel et al. (2018; Table S1.1). Although discriminating wild carob trees from abandoned orchards was not always possible, most efforts were done to collect leaves of individuals from natural habitats. On the field, each sampled population was identified as cultivated or uncultivated and the possibility of human-mediated admixture was considered throughout the analyses (see below, Table S1.1). *Ceratonia oreothauma* was sampled from Edinburgh herbarium and Oman botanical garden material. DNA extraction method is described in Appendix S2.

Divergence time analyses

Molecular dating analyses were based on available data from Genbank database for two nuclear sucrose synthase (SUSY) copies and the plastid maturase K (*mat*K) region (see Appendix S2). Divergence times were estimated using a Bayesian relaxed-clock approach implemented in BEAST v. 1.8.3 (Drummond & Rambaut, 2007). The concatenated dataset was run using a GTR + G model, a Yule speciation process prior, and an uncorrelated lognormal molecular clock. A fossil attributed to *Ceratonia emarginata* Heer from Central East Europe and dated at mid to Late Oligocene (Palamarev 1989) was used to calibrate the *Ceratonia* stem node (lognormal prior distribution, mean in real space = 28.1 ± 0.2 Ma). Three secondary calibrations from Lavin et al. (2005) were applied to deeper nodes of the tree using normal prior distributions: i) 54 ± 3.4 Ma assigned to the Umtiza stem node, ii) 58.6 ± 0.25 Ma applied to the papilionoid stem node and, iii) 59 ± 0.2 Ma set to the legume stem node.

Plastid DNA haplotype variation within C. siliqua

Four plastid markers (*mat*K gene, *psb*D-*trn*T spacer, *rpl32-trn*L spacer and *ccSA-ndh*D spacer) were sequenced for 424 *C. siliqua* and ten *C. oreothauma* individuals. A haplotype network was constructed with the concatenated dataset using TCS (Clement et al., 2000). Nucleotide diversity (π) and Tajima's D (D) were computed with DNAsp (Rozas et al., 2017).

Phylogeography of the carob tree based on nuclear microsatellite markers

The development of 18 new SSR markers for carob tree and a method for genotyping microsatellites by high-throughput sequencing are described in Viruel et al. (2018). This genotyping improvement allows verifying the presence of paralogous copies and scoring separately SSR and SNP variation. One microsatellite marker was removed due to the difficulty to filter alleles between two paralogue copies and therefore 17 markers were finally used in our study. Because carob varieties are usually propagated by grafting, some sampled genotypes were found to be clones in the dataset and they were excluded from the analysis. The final dataset included 1037 multilocus genotypes. Genetic diversity was summarized at the level of the main genetic groups obtained from STRUCTURE analyses (see below) and the allelic richness (N_A), Shanon index, observed (H_O) and unbiased expected heterozygocity (H_E), and private allele rate (number of alleles by locus unique to a single genetic group) were computed with GENEALEX software (Peakall & Smouse 2012). A map showing N_A and H_E per population was elaborated to document the geographical patterns of genetic diversity.

STRUCTURE software (Pritchard et al., 2000) was used to investigate the genetic structure based on SSR data, in which one to ten genetic groups (K = 1-10) were tested allowing admixture at individual level, with correlated allele frequencies and without prior information on geographical origin. Ten replicates were performed, each run having a burn-in period of 200,000 simulations and a chain length of 2,000,000 simulations. STRUCTURE HARVESTER (Earl & vonHoldt 2012) was used to obtain likelihood values across the multiple values of K as well as to apply the delta K criterion to select the optimal K (Table S3.1). However, acknowledging the discrepancies between methods in choosing the optimal K (Janes et al., 2017), all the results of STRUCTURE from K=1 to K=10 computed with CLUMPAK (Kopelman et al., 2015) are shown in Appendix S3. In addition, to estimate genetic structure with a

model-free method, we performed a hierarchical classification analysis of pairwise population genetic differentiation. A Jost's D differentiation matrix was analyzed with the Ward method based on within group variance criteria (Murtagh & Legendre 2014). The population pairwise Jost's D differentiation matrix was computed with the MMOD package (Winter 2012) using 58 populations that contain at least 10 genotypes.

Inferences on phylogeography by ABC

We aimed to test the ERH versus an alternative hypothesis by conducting coalescent simulations and comparing models using an ABC approach. Models were constructed with a reduced number of genetic groups (K=4 from STRUCTURE analyses) used as populations in coalescent simulations. We aimed to minimize the confusing effects of recent admixture by assigning each population to a group according to their probability of membership obtained in STRUCTURE. For this purpose, a reduced dataset was used containing 38 populations (542 genotypes) with membership values above the median of each group (see Appendix S1).

The two candidate models for ABC analyses are shown in Figure 1. According to the ERH, current populations of carob tree would descend from an eastern refugium, and western populations would be more recently derived than central populations. The alternative model is supported by the plastid haplotype network, the palaeobotanical data, the SDM results and the K=2 clustering from STRUCTURE (see Results): it assumes a phylogeographic split between eastern and western populations associated to two main refugia that led to two main genetic groups. This model, hereafter TRH (two refugia hypothesis), is also well supported by several phylogeographic studies focusing on widespread Mediterranean plants (Nieto Feliner 2014).

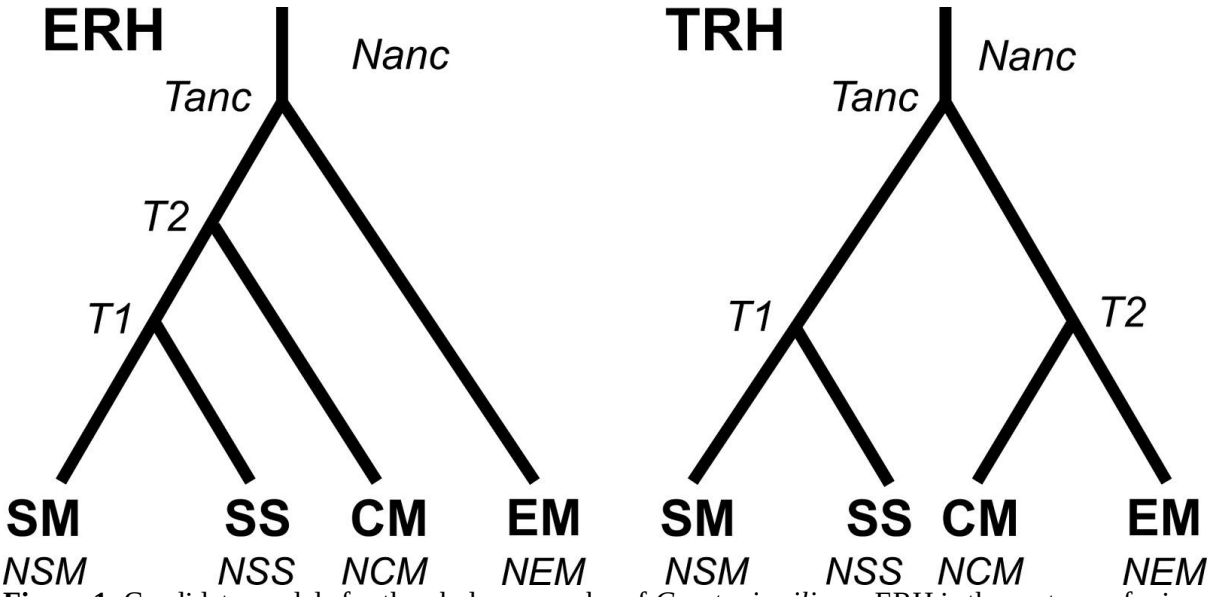


Figure 1: Candidate models for the phylogeography of *Ceratonia siliqua*. ERH is the eastern refugium hypothesis and TRH the two refugia hypothesis. SM, SS, CM and EM correspond to the Southern Morocco, South Spain, Central Mediterranean and Eastern Mediterranean genetic groups, respectively.

All coalescent simulations were done with FASTSIMCOAL 2.6 (Excoffier et al., 2011). For model selection $1x10^5$ simulations were generated. The present and ancestral population effective sizes were drawn from uniform distributions. After the ancestral node (Tanc) the effective sizes remained constant. The details on coalescent simulations and prior distributions are given in Appendix S2. Allelic richness, expected heterozygosity, pairwise and overall F_{st}, as well as pairwise $\delta\mu^2$ were used as summary statistics for ABC and were computed using ARLSUMSTAT with ARLEQUIN (Excoffier & Lischer 2010). Summary statistics used for ABC are given in Appendix 2. Model selection and parameter estimation was done according to Aoki et al. (2019). Confusion matrix, misclassification rate and model selection were performed in R with the Random forest approach implemented in the ABCRF package (Pudlo et

al., 2015). One thousand trees were constructed, and the best model was selected by classification votes of random forest. Parameter inference was done for the selected model performing one million simulations, with the neural network method (R package ABC; Csilléry et al., 2012) and a logit transformation of parameter values, considering the closest 1000 simulations to the observed data (tolerance = 0.1%). The 95% highest posterior densities (HPD) were obtained with the HPDinterval function (R coda package, Plummer et al., 2006). The quality of parameters estimation was controlled in two different ways. First, posterior distributions of the parameters were compared to prior distributions based on density curves. We chose the mode of the posterior distribution after having controlled for the presence of a peak in posterior density curves. The posterior distributions of summary statistics were obtained from 1000 datasets simulated by using parameter values randomly sampled from parameters posterior distribution. The goodness-of-fit was checked by graphical inspection of the distance between observed data and posterior distribution of simulated data. To convert node ages in years we fixed a generation time for the carob tree, the details of our approach are detailed in Appendix S2.

Environment Niche Modelling

We included 650 occurrence points to model the distribution range of the carob tree using a consensus approach modelling procedure (Marmion *et al.*, 2009) based on climatic variables at a 2.5 arc-min resolution from version 1.4 of the Worldclim database (Hijmans et al., 2005). We projected the species distribution under current, Mid-Holocene (~6 ka), Last Glacial Maximum (~22 ka), and Last Interglacial (~130 ka) climatic conditions. More details about carob distribution modelling are given in Appendix 2.

Palaeobotanical information

We performed a comprehensive survey of fossils in the literature resulting in a tentative summary of 55 occurrences of *Ceratonia* in the past (including macro-remains such as leaves, fruits, seeds, woods; and pollen grains; Table S2.5). A map was elaborated using GeoMapApp (Ryan *et al*., 2009).

Results

Divergence time analyses

Divergence time estimates (Figure S3.1) support a Late Miocene divergence between *C. siliqua* and its sister species *C. oreothauma* at 6.4 Ma (95%, HPD: 2.9-10.9). The divergence between the two main *C. siliqua* pDNA groups (see below) was estimated to occur during the Early Pleistocene (1.3 Ma, 95%, HPD: 0.4-2.2).

Geographical structure of plastid polymorphism

Nine pDNA markers were tested for sequence variation and only four were polymorphic. The alignment of these four plastid markers has a length of 1472 bp. The presence of three substitutions and five indels identified six haplotypes in *C. siliqua*. Nucleotide diversity (0.00106), number of segregating sites (3) and Tajima's D (3.34, pval <0.01) indicated low nucleotide variation and a lack of singletons. Three haplotypes were identified in *C. oreothauma* due to the existence of two indels. The haplotypes of *C. siliqua* and *C. oreothauma* differ by fifteen substitutions and four indels.

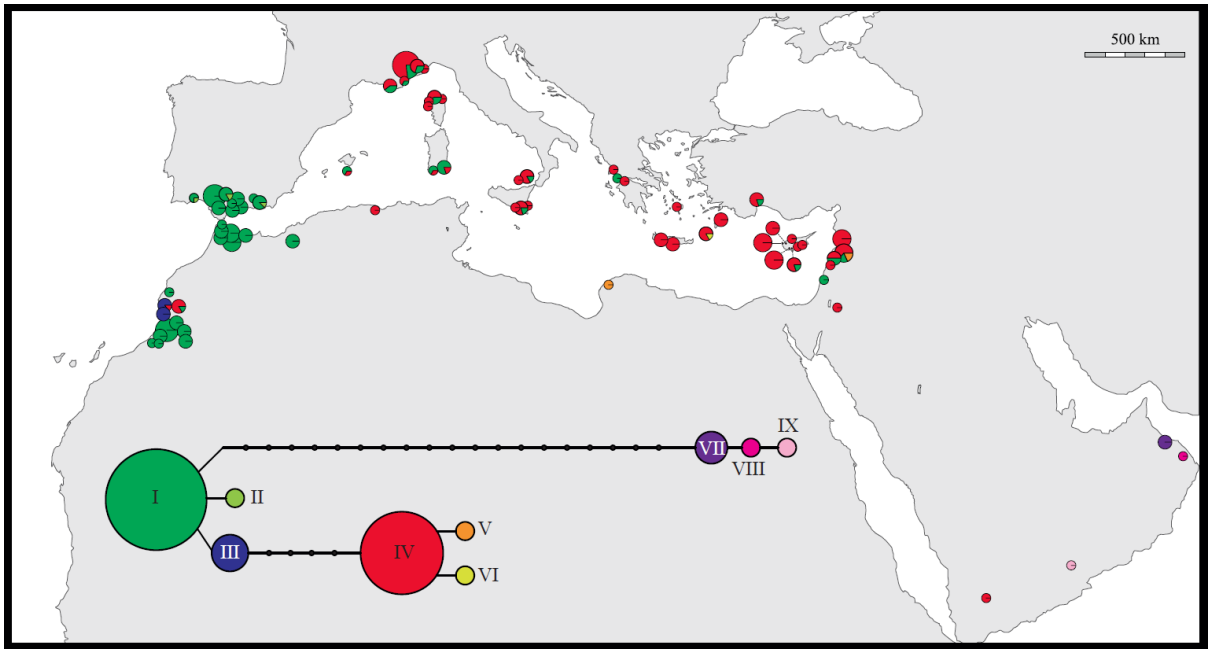


Figure 2: Geographic distribution of the plastid haplotypes identified in *Ceratonia siliqua* (I to VI) and *C. oreothauma* (VII to IX). In the haplotype network, the size of the circles is representative of sample size. Four markers were sequenced to obtain these haplotypes for a total of 1472 bp.

The network arranged the haplotypes of *C. siliqua* in two major groups mainly represented by haplotypes I and IV, respectively (Figure 2). The first group conformed by the haplotypes I, II and III is connected to *C. oreothauma* and its frequency is higher in the western part of the distribution range (Figure 2). Haplotypes IV, V and VI, forming the second group, differ by only one indel. This second group is more frequent in the East and is almost absent in the West except for Southern Morocco, in which the samples collected north of Agadir included the haplotypes III and IV. Most of the samples throughout the Mediterranean had either the haplotype I (219 samples) or IV (179 samples), whereas haplotype III was found in 11 samples all from South Morocco, haplotype II in three samples from Spain, and V and VI in three samples from the Eastern Mediterranean. In *C. oreothauma*, the most frequent haplotype VII was retrieved in 11 samples whereas haplotypes VIII and IX appeared in one and two samples, respectively.

Structure of genetic diversity

The 17 SSR markers revealed a total of 105 alleles with a mean of 6.2 alleles per locus. On average, 1033 genotypes were obtained per locus, and the overall mean observed heterozygosity was 0.47, the expected heterozygosity 0.55 and the fixation index 0.14.

The results of STRUCTURE from K=1 to K=10 are shown in Appendix 3, Figure S3.2. According to STRUCTURE HARVESTER results (Table S3.1), the best solutions are two or four genetic groups. No multimodality was observed for these solutions (Figure 3) and they are congruent with the Ward's clustering tree based on population pairwise Jost's D differentiations (Figure 3 and Figure S3.3) which clearly shows four main clusters that are grouped into two lineages. Although sub-optimal K solutions (e.g. K= 3, 5 or 7, Figure S3.2) present additional patterns, we focused here on two or four groups solutions that provide the best trade-off between STRUCTURE results and simplicity to test the ERH.

For K=2, the first genetic group includes South Morocco and South Spain (in green, Figure 3) whereas the second group includes the genotypes from the remaining populations. The K=4 clustering identified the following groups (Figure 3): SM, Southern Morocco; SS, Southern Spain; CM, Central Mediterranean including genotypes from Portugal, Algeria, France, Sardinia, Sicily and Balearics; EM, Eastern Mediterranean (i.e. Greece, Cyprus and Lebanon).

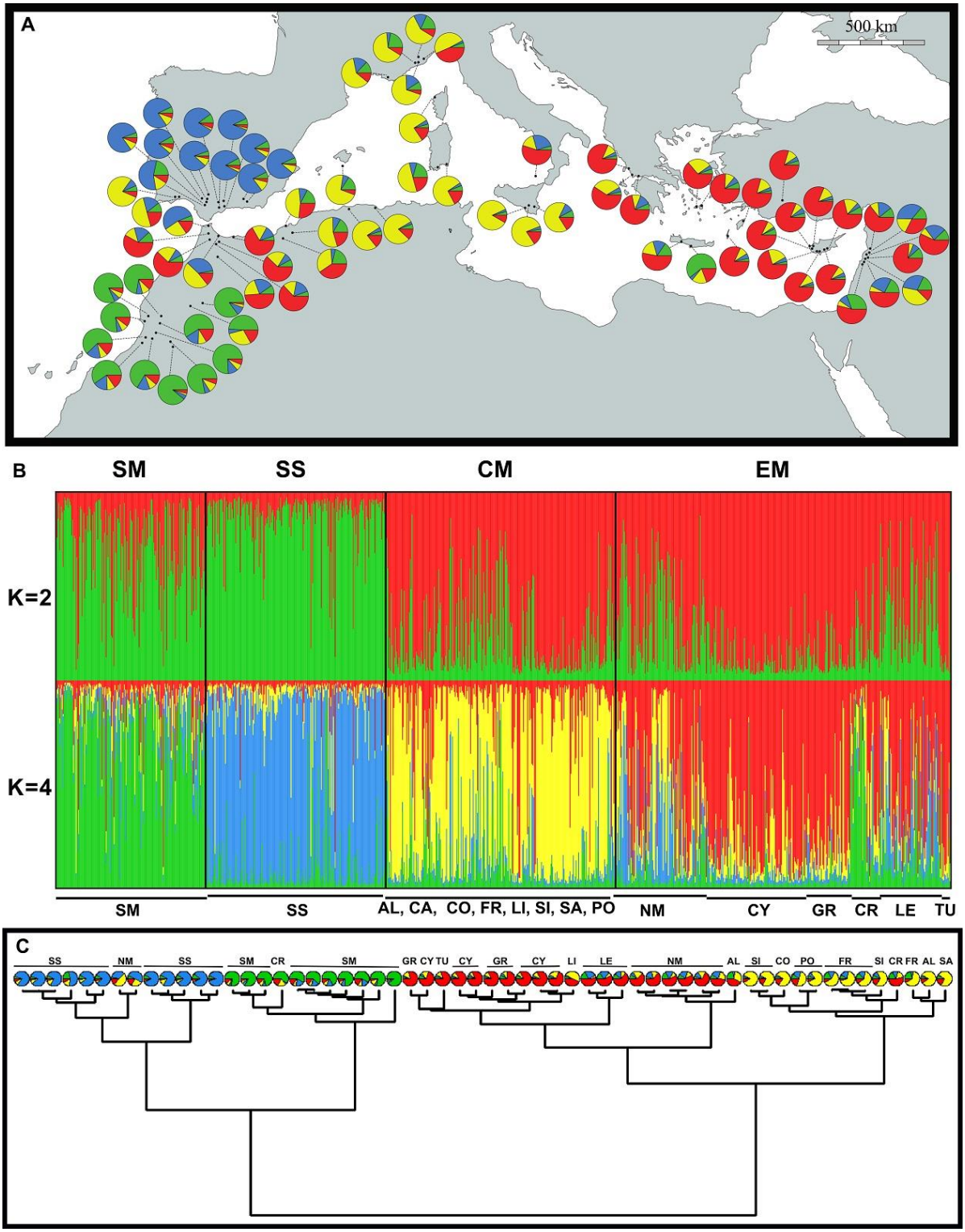


Figure 3: Genetic groups inferred for carob tree using 17 microsatellite markers. *SM, Southern Morocco; SS Southern Spain; CM, Central Mediterranean; EM, Eastern Mediterranean.* (A) Map of the membership frequencies of 1037 carob genotypes from 67 populations according to the K=4 clustering solution of STRUCTURE. (B) Admixture proportions for K=2 and K=4 obtained from ten replicates in STRUCTURE (each 2 000 000 chain length after 200 000 burn-in simulations, admixture model, no population prior) analyzed by CLUMPAK. Each sample is shown by a vertical bar partitioned according to its membership to one of the K groups. (C) Clustering tree obtained with Ward's method and pairwise Jost's D differentiation between populations having at least ten individuals (the tree with population names is shown in Figure S3.3). The pie charts corrrespond to admixture plots obtained with STRUCTURE for K=4 groups. *AL, Algeria; CA, Cabrera; CO, Corsica; CR, Crete; CY, Cyprus; FR France; GR, Greece; LE, Lebanon; LI, Liguria; NM, North Morocco; PO, Portugal; SA, Sardinia; SI, Sicily.*

Admixture is observed in most populations and it is particularly strong in North Morocco, France, Crete and Lebanon (Figure 3). Six populations from North Morocco and one from Algeria were mostly assigned to a fourth group (EM) even though they show admixture with the other groups, whereas two population of North Morocco are assigned to South Spain (Figure 3). In Crete, one population is mostly assigned to the SM group (GRFAR) whereas the second is either assigned to the EM or CM group depending on the analysis (Figure 3). For ABC analysis, populations from North Morocco, Crete and some from Lebanon were excluded from the second analysis (see details of selection Table S1.1). The overall F_{st} between the four groups is 6.7 % when all populations are included and it is 10% when admixed populations are removed.

Table 1: Genetic diversity for the four main genetic groups of *Ceratonia siliqua*. *N* is the number of trees screened, *MLG* is the number of multilocus genotype obtained, *Na* is the allelic richness, *Na(>=5%)* is the allelic richness restricted to allele having a minimum frequency of 5%, *I* is the Shannon index, *Ho* is the observed heterozygosity, H_E is the unbiased expected heterozygosity, *F* is the fixation index and *Priv* is the private allelic rate.

Genetic group	**N**	**MLG**	**Na**	**Na(>=5%)**	**I**	**Ho**	**H_E**	**F**	**Priv**
South Morocco (SM)	176	167	5.5	3	1.08	0.51	0.58	0.11	0.47
South Spain (SS)	207	206	4.6	3.2	1.03	0.53	0.56	0.06	0.06
Central Med. (CM)	282	275	5.1	2.7	0.88	0.43	0.48	0.08	0.12
East Med. (EM)	402	392	5.2	3.1	0.98	0.46	0.53	0.12	0.18
Total population	1067	1037	6.2	3.8	1.05	0.47	0.55	0.14	

Summary of genetic diversity within the four main genetic groups (Table 1) revealed small differences in allelic richness but higher H_E and Shannon values in groups SS and SM, with the lowest values in group CM. Number of private markers are more than twice higher in SM than in EM but low in SS and CM. The map of N_A and H_E per population revealed higher scores in SM and SS and few populations in the East with high diversity (Figure S3.4).

Phylogeographic inferences by ABC

The model selected by the random forest ABC analyses is graphically summarized in Figure 4. Classification error rates of ERH and TRH models were 14 and 19 %, respectively, for an overall error rate of 16.5 %. The TRH model received 951 votes from the 1000 classification trees, which allowed rejecting the ERH model. ABC supports TRH with a posterior probability of 0.92, in which two divergent lineages derived from the ancestral node subsequently diverged into the current four genetic groups (Figure 4). Parameters inferred for the most probable model are indicated in Table S3.2. The mode was used as a point estimate for all parameters after graphical inspection of posterior distributions (Figure S3.5). Comparisons of distribution of simulated summary statistics (1000 datasets) with values of the observed summary statistics showed a good congruence between simulated and observed statistics confirming the goodness of fit for the selected model (Figure S3.6).

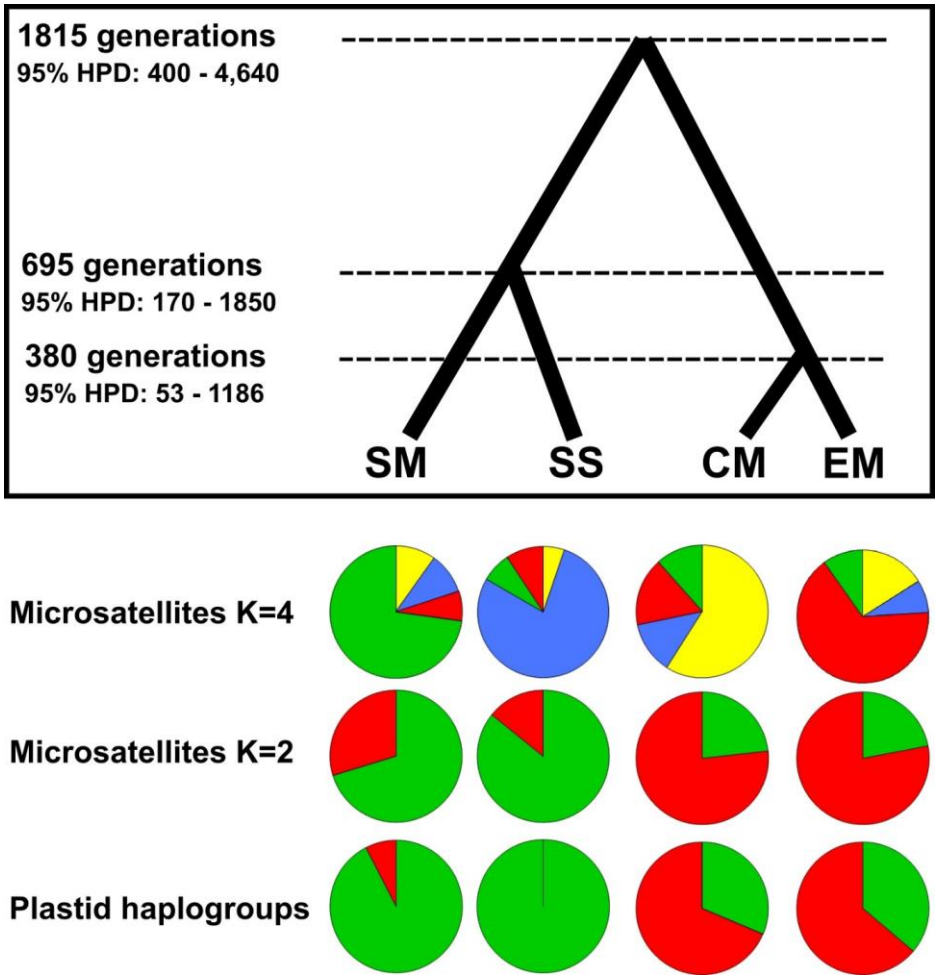


Figure 4: The best fitting model inferred by ABC and admixture proportions from microsatellites data (STRUCTURE models for K= 2 and K= 4, full data set of 1037 carob genotypes from 67 provenances) and plastid markers data. SM = Southern Morocco. SS = South Spain. CM = Central Mediterranean. EM = Eastern Mediterranean. Plastid haplogroups, in green haplotypes I, II and III, in red haplotypes IV, V and VI.

ABC analyses estimated 1815 generations to the ancestral node, 695 generations to the node between SM and SS and 380 generations to the node between CM and EM (Table S3.2 and Figure 4). Assuming a generation time of 64 years for carob tree (see Appendix S2), the ancestral divergence of the carob lineages would be dated to 116 ka (95% HPD 25 – 297 ka), the split between South Morocco and South Spain to 44 ka (95% HPD 10 – 118 ka) and the split between the central and eastern groups to 24 ka (95% HPD 3 ka – 76 ka). These nodes ages, despite their large 95% HPD, have well peaked distribution modes (Figure S3.5) supporting that the main genetic groups of the carob tree diverged well before the onset of crop tree domestication in the Mediterranean.

Distribution hindcasting and past occurrences of the carob

Distribution modelling for past periods infers a potential contraction towards the south during the Last Glacial Maximum (LGM) at about 22 ka, with suitable environmental conditions persisting below 41° of latitude. This result is consistent for three Global Circulation Models used here (Appendix 3, Figure S3.7) although only the results from the CCSM4 model are shown in Figure 5. The strongest change is inferred during the Last Interglacial (LIG) at about 130 ka (Figure 5) with suitable conditions for the species restricted to the Atlantic coasts of Morocco, Spain and Portugal, which suggests extinction or very strong contraction elsewhere.

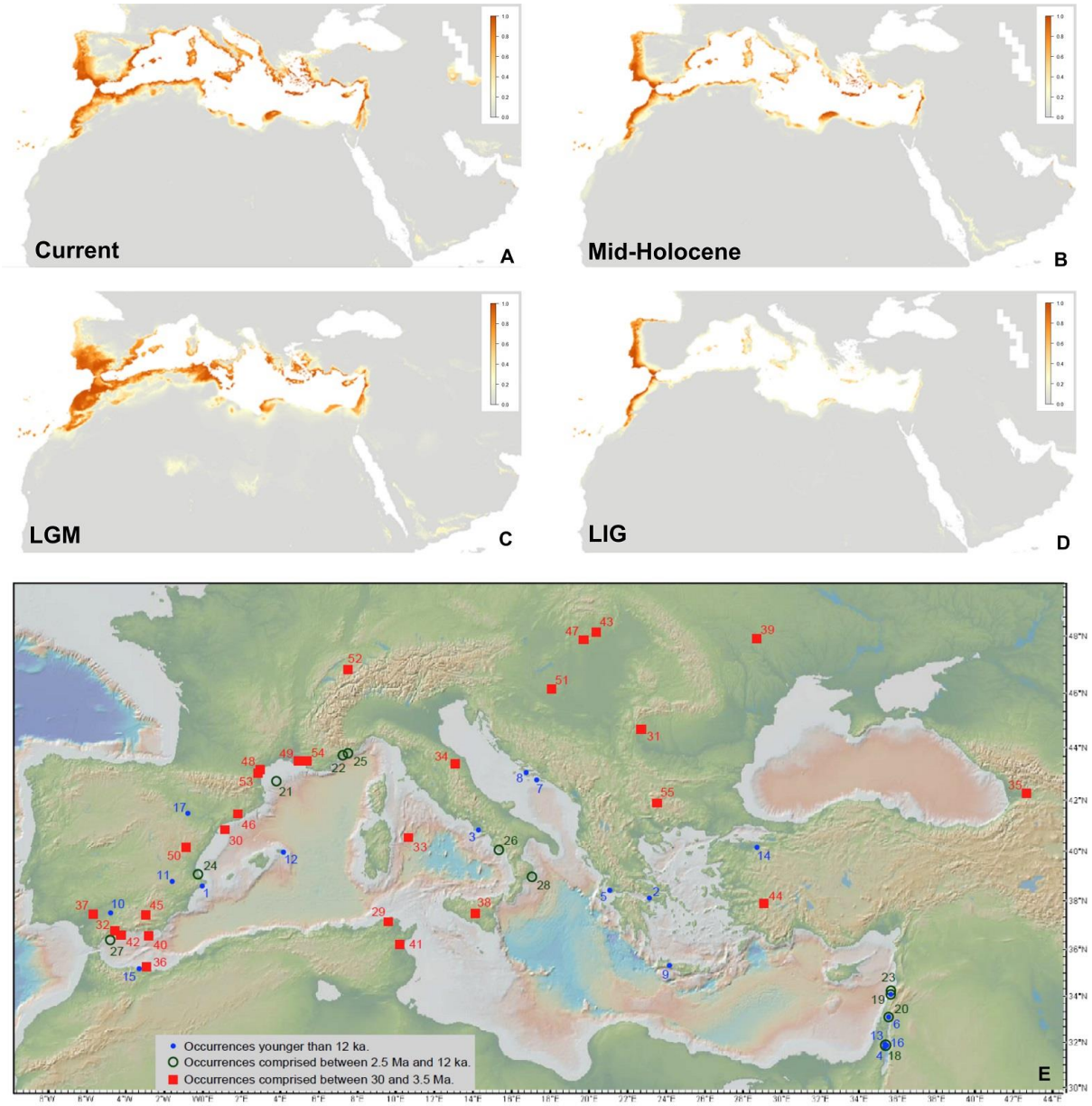


Figure 5: Species distribution modelling of the carob in four periods (A, Present; B, Mid-Holocene; C, Last Glacial Maximum (LGM); D, Last Inter-Glacial (LIG)), and E, Map of palaeobotanical data of *Ceratonia* from Oligocene to Middle Ages (macroremains and pollen grains). Numbers correspond to the input order of data in Table S2.5 from the youngest (No.1, Middle Ages) to the oldest data (No.55, Early Oligocene).

Forty-four references allowed compiling 55 past occurrences of *Ceratonia* mostly based on pollen records (Table S2.5). Mapped fossil records (Figure 5) indicate a mostly continuous presence of *Ceratonia* from Oligocene to Late Pliocene around the palaeo-Mediterranean Sea (including the former Paratethys) which progressively reduced from 20 Ma up to its modern physiography reached at *ca*. 3.6 Ma. The Early Pleistocene occurrences are restricted to southwestern Europe (localities 25 – 28) but Middle and Late Pleistocene occurrences (18-24) are located in the East and the West. Although pollen grains and macro-remains records of Early Holocene were reported mainly from the Middle East, two western occurrences for this period (dated at ca. 11 ka, localities 17 and 15), indicate the probable presence of the carob tree before domestication in the Western Mediterranean.

Discussion

Reconstructing the phylogeography of cultivated fruit trees is challenging due to their long generation time, the vegetative propagation conducted by farmers and the recurrent exchanges between cultivated and wild populations. All these aspects are found in the carob tree, but our multidisciplinary study has unveiled relevant evidences that allow a better understanding of its natural history and domestication.

Divergence time analyses estimated an origin for *Ceratonia* near the Oligocene/Miocene boundary and a split between *C. siliqua* and *C. oreothauma* around the Late Miocene. The fossil pollen records found during the Miocene in the Mediterranean are morphologically similar to modern *C. siliqua* and differ from the pollen of its sister species, *C. oreothauma*, by the number of apertures and surface ornamentation (Figure S2.3). The geographic distribution of the *Ceratonia* fossil records prior to 3.5 Ma was linked to the tropical-subtropical wide northern margin of the paleo-Mediterranean Sea (Bessedik et al., 1984), corresponding to the pre-Mediterranean sclerophyllous vegetation (Axelrod, 1975; Jiménez-Moreno and Suc, 2007). In the Pleistocene, after the onset of the Mediterranean climate, it is noteworthy that a blatant thermophilous forest community with *Ceratonia* and *Olea* was identified by 1.4 Ma pollen records at Camerota (S Italy) (*Olea*: 10% and *Ceratonia*: 12% of the pollen assemblage; Brenac, 1984). Fossil data and our phylogenetic results identified carob tree as a "pre-Mediterranean" lineage (Herrera, 1992) and its ancestors probably were widely distributed around the Tethys Sea during the Paleogene (Palamarev, 1989) in tropical forests that were impoverished by the successive extinctions during the transition towards the Mediterranean climate (Suc, 1984; Mijarra et al., 2009; Suc et al., 2018). Fossil records of the carob dated during the Pleistocene were found both in the western and eastern Mediterranean, therefore not supporting the ERH.

The overall SSR allelic richness of the carob tree (N_A = 6.2; and 3.8 when rare alleles are excluded) is low compared to values published in recent studies of Mediterranean trees using microsatellites, such as *Olea europaea* L. (N_A= 16, Diez et al., 2015), *Prunus dulcis* (Mill.) D.A. Webb (N_A = 17, Delplancke et al., 2012), or *Phoenix dactylifera* L. (N_A = 13, Zehdi-Azouzi et al., 2015); but higher than in the Tertiary relict *Platanus orientalis* L. (N_A= 1.9, Rinaldi et al., 2019). La Malfa et al. (2014) reported a very similar N_A value of 3.1 for nine SSR markers screened in carob cultivars from Italy, Malta and Spain. Plastid allele richness is also very low with only six haplotypes and most of the range is represented by only two of them. The positive Tajima's D result obtained with plastid data is supporting a strong demographic decline in the past. These results contrast with the higher plastid diversity observed in other Mediterranean trees like *Erica arborea* L., *Olea europaea* or *Myrtus communis* L. (Désamoré et al., 2011; Besnard et al., 2017; Migliore et al., 2018); but are congruent with the low plastid variation observed in *Laurus nobilis* L. (Rodriguez-Sanchez et al., 2009) and even higher than in *Celtis occidentalis* L. and *Nerium oleander* L. (Mateu-Andres et al., 2015). Nuclear and plastid markers therefore indicate low levels of allelic richness for the carob tree which are consistent with a strong decline scenario. If an expansion followed this decline, it should have been too recent to allow for the recovery of plastid variation. As indicated above, this past decline of the carob tree in consistent with palaeobotanical data and SDM results.

Far from confirming the ERH, genetic data supports an alternative hypothesis for the carob tree history. First, the strong west-east plastid differentiation rejects an exclusive eastern origin (Figure 2). Second, clustering of microsatellite polymorphisms revealed that populations from Southern Morocco and Southern Spain form a distinct cluster and exhibit higher genetic diversity (Figure S3.4). Therefore, an extended presence of the carob tree in the Western Mediterranean is strongly supported by the simple description of genetic diversity structure. Finally, ABC analyses also rejected the ERH. A similar ERH was also proposed in the past for the olive tree but genetic data sampled throughout the Mediterranean support a different scenario (Breton et al., 2006; Besnard et al., 2017).

The most likely scenario for the phylogeography of the carob tree depicts two lineages (Figure 4), one being restricted to the westernmost part of its range. A subsequent divergence found within this western lineage is older than the split between the central and eastern genetic groups, supporting an older and larger persistence in the western margin of the range. Such an old western occurrence is also supported by the higher genetic diversity and originality (private markers) scored in the western lineage for both plastid and SSR markers. Calculating a specific timing for this scenario is highly dependent on the

generation time, which remains difficult to be estimated for trees (e.g. Cavender-Bares et al., 2011; Tsuda et al., 2015). However, despite uncertainties on generation time and the large confidence intervals for divergence times, the origin of the four lineages recognized within *C. siliqua* (116 ka) substantially preceded its domestication, which may have started ca. 3 ka (Meyer et al., 2012), as well as the first tree crop domestication records ca. 6 ka (Zohary & Hopf 2012). Phylogeography and palaeobotany thus support a native status of the carob tree throughout the Mediterranean.

Independently from molecular data, SDM results are also consistent with a range contraction towards a western refugium along the Atlantic coast during the LIG (Fig. 5). This unexpected pattern for a thermophilous tree could be due to an increased climatic continentality (i.e. colder winters and warmer as well as drier summers), which was more pronounced in the East during LIG times (Felis et al., 2004). This modelled response of the carob tree to the drastic reduction of its suitable range is consistent with the confidence interval for its likely ancestral node age coinciding with the Eemian period (130 - 115 ka, LIG) and would explain the strong population decline estimated by molecular data. A similar range contraction during the LIG was recently inferred for the thermophilous shrub *Lavatera maritima* (Villa-Machío et al., 2018).

The genetic diversity of the carob tree is also characterized by introgression events between the main four lineages. Carob trees from Northern Morocco, which belong to the Western plastid haplogroup, are mostly assigned to the EM group while having memberships for the SS and CM groups. This incongruence between nuclear and plastid markers, also observed in one population of Crete, could be explained by long-term effects of asymmetric introgressions during the expansion of cultivated carob trees, as proposed in the theoretical study of Currat et al. (2008). According to these authors, plastid markers are particularly prone to such process. In North Morocco, this incongruence strongly suggests that the invading genetic pool was largely introgressed by the resident wild carobs bearing the western haplotype, which was fixed in the admixed populations. Further studies will explore this hypothesis and whether such expansions and introgressions could have taken place in the Roman or Arab times in Northern Morocco. In Lebanon and France, a strong admixture was observed but in this case for both plastid and nuclear markers, which may suggest a more recent contact between the two main lineages.

Conclusion

Our multidisciplinary study based on phylogeography, palaeobotany and species distribution modelling clarified an enduring controversy on the origin of the carob, an important fruit tree for the Mediterranean cultures. We provide strong support for the existence of four lineages within *C. siliqua* that evolved before the major civilizations of the Fertile Crescent and the Mediterranean.

Our results are very supportive of the persistence of the carob tree in Moroccan and Iberian refugia but they do not discard the existence of smaller refugia in the Eastern Mediterranean. SDM predicts small suitable areas in this region during the LIG and both plastid and microsatellite patterns of variation indicate a strong east-west divergence. We thus propose a new hypothesis according to which current *C. siliqua* populations originated from two disjunct refugia after LIG.

The previously accepted eastern domestication hypothesis is also difficult to reconcile with our data because cultivated carobs are present within the four lineages discovered here. The role of Greek, Roman and Arab farmers in the history of the carob was certainly important for the dissemination of carob cultivation practices and cultivars, but our results support a local use and domestication of the carob tree from native populations throughout the Mediterranean.

Acknowledgement:

This study is part of the DYNAMIC project supported by French national agency of research (ANR-14-CE02-0016) and benefited from equipment and services from the molecular biology facility (SCBM) at IMBE (Marseille, France) and from the genotyping and sequencing core facility (iGenSeq) at ICM (Hôpital Pitié Salpêtrière, Paris, France). J.V. benefited from a Postdoc Fellowship funded by DYNAMIC and a Marie Skłodowska-Curie Individual Fellowship (704464 - YAMNOMICS - MSCA-IF-EF-ST). The simulations and ABC were done on the "High Performance Computing Cluster" OSU Institut Pythéas (Aix Marseille Univ, INSU-CNRS). The authors thank for providing information or their help to complete our sampling: Annette Patzelt (Oman Botanic Garden), Minas Papadopulos

(Department of Forests of Republic of Cyprus), Zahra Djabeur (Oran University), Nabil Benghanem (Tizi-Ouzou University), Gianluigi Bacchetta (Cagliari University), Sonja Yakovlev (Paris-Sud University), Errol Vela (CIRAD), Severine Fauquette (ISEM), Omar Boudouma (ISTeP), Maria Panitsa (Patras University), and the services of Junta de Andalucia.
The data used in this study is available on Genbank and Dryad, see the supporting information at the end of this document.

REFERENCES

Aoki, K., Tamaki, I., Nakao, K., Ueno, S., Kamijo, T., Setoguchi, H., ... Tsumura, Y. (2019). Approximate Bayesian computation analysis of EST-associated microsatellites indicates that the broadleaved evergreen tree *Castanopsis sieboldii* survived the Last Glacial Maximum in multiple refugia in Japan. *Heredity*, 122, 326-340.

Axelrod, D. I. (1975). Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden*, 62, 280–334.

Baumel, A., Mirleau, P., Viruel, J., Boudagher Kharrat, M., La Malfa, S., Ouahmane, L., ... Médail, F. (2018). Assessment of plant species diversity associated with the carob tree (*Ceratonia siliqua*, Fabaceae) at the Mediterranean scale. *Plant Ecology and Evolution*, 151, 185–193.

Besnard, G., Terral, J.-F., & Cornille A. (2017). On the origins and domestication of the olive: a review and perspectives. *Annals of Botany*, 121, 385–403.

Bessedik, M., Guinet, P., & Suc, J.-P. (1984). Données paléofloristiques en Méditerranée nord-occidentale depuis l'Aquitanien. *Revue de Paléobiologie*, special volume, 25–31.

Brenac, P. (1984). Végétation et climat de la Campanie du Sud (italie) au Pliocène final d'après l'analyse pollinique des dépôts de Camerota. *Ecologia mediterranea*, 10, 207–216.

Breton, C., Tersac, M., & Bervillé, A. (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of Biogeography*, 33, 1916-1928.

Candolle, A. de (1883). *Origine des plantes cultivées*. Germer Baillière et Cie.

Cavender-Bares, J., Gonzalez-Rodriguez, A., Pahlich, A., Koehler, K., & Deacon, N. (2011). Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. *Journal of Biogeography*, 38, 962-981.

Clement, M., Posada, D., & Crandall, K. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657-1660.

Csilléry, K., François, O., & Blum, M. G. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475-479.

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. *Evolution*, 62, 1908-1920.

Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E., & Arrigo, N. (2012). Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications*, 5, 317-329.

Désamoré, A., Laenen, B., Devos, N., Popp, M., González- Mancebo, J. M., Carine, M. A., & Vanderpoorten, A. (2011). Out of Africa: north- westwards Pleistocene expansions of the heather *Erica arborea*. *Journal of Biogeography*, 38, 164-176.

Diez, C. M., Trujillo, I., Martinez- Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, 206, 436-447.

Drummond, A.J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214.

Earl, D. A., & vonHoldt B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359-361.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567

Excoffier, L., & Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27, 1332-1334

Felis, T., Lohmann, G., Kuhnert, H., Lorenz, S. J., Scholz, D., Pätzold, J., ... Al-Moghrabi, S. M. (2004). Increased seasonality in Middle East temperatures during the last interglacial period. *Nature*, 429, 164-168.

Gavin D.G., Fitzpatrick M.C., Gugger P.F., Heath K.D., Rodríguez-Sánchez F., Dobrowski S.Z., … Williams J.W. (2014). Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204, 37–54.

Herrera, C. M. (1992). Historical effects and sorting processes as explanations for contemporary ecological patterns: character syndromes in Mediterranean woody plants. *The American Naturalist*, 140, 421-446.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones & Jarvis A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25,1965-1978.
Hillcoat D., Lewis G., & Verdcourt, B. (1980). A New Species of *Ceratonia* (Leguminosae-Caesalpinioideae) from Arabia and the Somali Republic. *Kew Bulletin*, 35, 261.
Hu F.S., Hampe A., & Petit R.J. (2009). Paleoecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment*, 7, 371–379.
Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K= 2 conundrum. *Molecular Ecology*, 26, 3594-3602.
Jiménez-Moreno, G., & Suc, J.-P. (2007). Middle Miocene latitudinal climatic gradient in Western Europe: evidence from pollen records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 253, 224–241.
Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15, 1179-1191
Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, 40, 593-612.
La Malfa, S., Currò, S., Douglas, A. B., Brugaletta, M., Caruso, M., & Gentile, A. (2014). Genetic diversity revealed by EST-SSR markers in carob tree (*Ceratonia siliqua* L.). *Biochemical Systematics and Ecology*, 55, 205-211.
Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. *Systematic Biology*, 54, 575-594.
Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15, 59-69.
Mateu-Andrés, I., Ciurana, M. J., Aguilella, A., Boisset, F., Guara, M., Laguna, E., ... Pedrola-Monfort, J. (2015). Plastid DNA homogeneity in *Celtis australis* L. (Cannabaceae) and *Nerium oleander* L.(Apocynaceae) throughout the Mediterranean Basin. *International Journal of Plant Sciences*, 176, 421-432.
Meyer R.S., Duval A.E., & Jensen H.R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, 196, 29–48.
Migliore J., Baumel A., Leriche A., Juin M., & Médail F. (2018). Surviving glaciations in the Mediterranean region: an alternative to the long-term refugia hypothesis. *Botanical Journal of the Linnean Society*, 187, 537–549.
Mijarra, J. M. P., Barrón, E., Manzaneque, F. G., & Morla, C. (2009). Floristic changes in the Iberian Peninsula and Balearic Islands (southwest Europe) during the Cenozoic. *Journal of Biogeography*, 36, 2025-2043.
Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification*, 31, 274-295
Nieto-Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, 16, 265–278.
Palamarev, E. (1989). Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendroflora. *Plant Systematics and Evolution*, 162, 93–107.
Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics* 28, 2537e2539.
Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, 6, 7-11
Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2015). Reliable ABC model choice via random forests. *Bioinformatics*, 32, 859-866.
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
Quézel, P. & Médail, F. (2003). *Écologie et biogéographie des forêts du bassin méditerranéen*. Elsevier, Paris, France.
Ramón-Laca, L. & Mabberley, D.J. (2004). The ecological status of the carob-tree (*Ceratonia siliqua*, Leguminosae) in the Mediterranean. *Botanical Journal of the Linnean Society*, 144, 431–436.
Rinaldi, R., Cafasso, D., Strumia, S., Cristaudo, A., Sebastiani, F., & Fineschi, S. (2019). The influence of a relict distribution on genetic structure and variation in the Mediterranean tree, *Platanus orientalis* L. *AoB PLANTS*, 11, plz002.
Rodríguez-Sánchez, F., Guzmán, B., Valido, A., Vargas, P., & Arroyo, J. (2009). Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *Journal of Biogeography*, 36, 1270-1281.
Rozas J., Ferrer-Mata A., Sánchez-Delbarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., & Sánchez-Gracia A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, 34, 3299–3302.

Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., Weissel, R. A., Ferrini, V., Goodwillie, A., Nitsche, F., Bonczkowski, J., & Zemsky, R. (2009). Global Multi-Resolution Topography synthesis. *Geochemistry Geophysics Geosystems*, 10, Q03014.
Servera-Vives G., Riera S., Picornell-Gelabert L., Moffa-Sánchez P., Llergo Y., Garcia A., … Trías M.C. (2018). The onset of islandscapes in the Balearic Islands: A study-case of Addaia (northern Minorca, Spain). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 498, 9–23.
Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, 307, 429.
Suc, J.-P., Popescu, S.-M., Fauquette, S., Bessedik, M., Jiménez-Moreno, G., Bachiri Taoufiq, N., Zheng, Z., & Médail, F. (2018) Reconstruction of Mediterranean flora, vegetation and climate for the last 23 million years based on an extensive pollen dataset. *Ecologia mediterranea*, 44, 53–85.
Tsuda, Y., Nakao, K., Ide, Y., & Tsumura, Y. (2015). The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture- like genetic structure is the result of simple population splitting not admixing. *Molecular Ecology*, 24, 1403-1418.
Villa-Machío, I., Fernández de Castro, A. G., Fuertes-Aguilar, J., & Nieto Feliner, G. (2018). Out of North Africa by different routes: phylogeography and species distribution model of the western Mediterranean *Lavatera maritima* (Malvaceae). *Botanical Journal of the Linnean Society* 187, 441-445.
Viruel J., Haguenauer A., Juin M., Mirleau F., Bouteiller D., BoudagherKharrat M., … Baumel A. (2018). Advances in genotyping microsatellite markers through sequencing and consequences of scoring methods for *Ceratonia siliqua* (Leguminosae). *Applications in Plant Sciences*, 6:e1201.
Winter, D. J. (2012). MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, 12, 1158-1160.
Zapata L., López-Sáez J.A., Ruiz-Alonso M., Linstädter J., Pérez-Jordà G., Morales J., ... Peña-Chocarro L. (2013). Holocene environmental change and human impact in NE Morocco: Palaeobotanical evidence from Ifri Oudadane. *The Holocene*, 23, 1286–1296.
Zehdi-Azouzi, S., Cherif, E., Moussouni, S., Gros-Balthazard, M., Abbas Naqvi, S., Ludeña, B., ... Si-Dehbi, F. (2015). Genetic structure of the date palm (*Phoenix dactylifera*) in the Old World reveals a strong differentiation between eastern and western populations. *Annals of Botany*, 116, 101-112.
Zohary D. (2002). Domestication of the carob (*Ceratonia siliqua* L.). *Israel Journal of Plant Sciences*, 50, 141-15.
Zohary D. & Hopf M. (2012). *Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley*. Oxford University Press.

BIOSKETCH

Juan Viruel is working on conservation genetics of plants and all the authors are interested on several aspect of plant ecology and evolution including crops and their wild relatives. This work is part of an international project on carob biogeography, microbiology and ecology (https://dynamic.cirad.fr).

Author contributions: H.S. and A.B. planned this project. J.V., J.P.S., N.L.G., G.N.F., S.P., F.M. and A.B. conceived the ideas; A.B., F.M., G.N.F., M.B.K., L.O., S.L.M., K.D., H.S. and J.V. conducted the fieldwork and collected the data with additional material from collaborators; F.L.M., M.S., and J.V. conducted the laboratory work; A.B., N.L.G., J.P.S., S.P. and J.V. analyzed the data; A.B. and J.V. led the writing with assistance from G.N.F., S.P., F.M. and J.P.S.. All authors gave final approval for publication.

DATA ACCESSIBILITY

- Sampling information, extended laboratory and statistical methods, as well as extended results are available in supporting information.
- GenBank accessions of plastid haplotypes and SUSY sequences acquired for *Ceratonia* are available in supporting information.
- In DRYAD repository (https://doi.org/10.5061/dryad.k7m020r) are available:
    - Sequences alignments (fasta files) used for phylogeography and divergence time analysis.
    - Microsatellites genotypes.
    - STRUCTURE results
    - Occurrence points used for SDM

SUPPORTING INFORMATION

APPENDIX S1 *Ceratonia* sampling information

APPENDIX S2 Extended material & methods.

APPENDIX S3 Extended results

APPENDIX S1 *Ceratonia* sampling information

TABLE S1: Locality information of the populations of carob tree included in this study. 'Pop_ID' refers to the population codes used in the tables available in the Dryad Digital Repository (https://doi.org/10.5061/dryad.k7m020r). 'N_µsat' is the number of samples genotyped by microsatellite marker, and 'N_pDNA' is the number of samples for which the four plastid markers were sequenced (Table S2.1). 'Admix' values show the mean membership of each population to one of the four clusters assigned by Structure results (SM, SS, CM and EM). For ABC analyses, only the populations with membership values higher than the median of the cluster were used. SM= South Morocco, SS = South Spain, CM = Central Mediterranean, EM = Eastern Mediterranean.

Pop_ID	Country	Location	N_µsat	N_pDNA	Longitude	Latitude	Status	admix_SM	admix_SS	admix_CM	admix_EM	ABC
ALGAB	Algeria	Sidi Bel Abbès	9	0	-0.543853	35.120683	uncultivated	0.1673	0.0866375	0.5430125	0.2030625	
ALGGOU	Algeria	Gouraya mountain near Bejaia	20	2	5.079751	36.764385	uncultivated	0.05552	0.02905	0.79511	0.12032	CM
ALGMESS	Algeria	Messerghine	9	0	-0.863352	35.670302	uncultivated	0.171825	0.0941125	0.4668875	0.2671875	
ALGTAB	Algeria	Tabouinèt, Blida	8	0	3.04335	36.5191833	uncultivated	0.045825	0.04455	0.7670375	0.1425875	CM
ALGTLEM	Algeria	Tlemcem, El Ourit	13	5	-1.32	34.89	uncultivated	0.211938462	0.051015385	0.332092308	0.404946154	
CHAKA	Cyprus	Akamas, Neo Chorio	24	11	32.3352	35.02555	uncultivated	0.04663913	0.059313043	0.080208696	0.813834783	EM
CHANO	Cyprus	Anogyra	32	12	32.74097	34.74728	cultivated	0.032281818	0.048540909	0.2143	0.704877273	EM
CHAVA	Cyprus	Avakas, Paphos	19	3	32.337821	34.919507	uncultivated	0.069917647	0.023011765	0.084747059	0.822323529	EM
CHKOF	Cyprus	Kofinou	11	5	33.396846	34.853211	cultivated	0.082863636	0.0489	0.157254545	0.710972727	EM
CHLYS	Cyprus	Lysos	11	5	32.53441	34.9995	uncultivated	0.023	0.037690909	0.127672727	0.811627273	EM
CHSAN	Cyprus	Sanida	10	2	33.22641	34.79309	uncultivated	0.03236	0.04392	0.08298	0.84074	EM
CHSOT	Cyprus	Sotira	9	5	32.85653	34.714946	cultivated	0.025433333	0.030744444	0.099744444	0.844066667	EM
CHSTE	Cyprus	Steni	7	1	32.47223	35.0037	cultivated	0.032842857	0.039028571	0.390728571	0.5374	
ESACE	Spain	El Acebuchal	13	5	-3.92639	36.81622	cultivated	0.070961538	0.780769231	0.108476923	0.0398	
ESALH	Spain	Alhaurin	20	7	-4.72978	36.61124	uncultivated	0.084035	0.8162	0.060505	0.039275	SS
ESBEN	Spain	Benaocaz	16	5	-5.43861	36.69741	uncultivated	0.08190625	0.8040375	0.05410625	0.05994375	SS
ESCAB	Spain	Cabrera Island	7	3	2.951	39.141	uncultivated	0.1589	0.066657143	0.717728571	0.056714286	SS
ESELB	Spain	El bosque	20	7	-5.47019	36.701	cultivated	0.069015	0.77751	0.10024	0.053245	
ESGRA	Spain	Grazalema	20	6	-5.41916	36.75605	uncultivated	0.0532	0.831765	0.08256	0.032475	SS
ESNER	Spain	Nerja	11	8	-3.85342	36.76867	uncultivated	0.067281818	0.823409091	0.086881818	0.022427273	SS
ESOJE	Spain	Ojen	20	7	-4.88065	36.57748	uncultivated	0.080125	0.84329	0.032065	0.04452	SS
ESPIC	Spain	El Picacho, Sierra del Aljibe	29	5	-5.649944	36.521861	uncultivated	0.215424138	0.561465517	0.134293103	0.088817241	
ESPIZ	Spain	Pizarra	21	7	-4.6997	36.76009	cultivated	0.070755	0.872135	0.02113	0.035985	SS
ESUBR	Spain	Ubrique	20	7	-5.44537	36.66859	uncultivated	0.054305	0.79426	0.08639	0.065035	
ESYUN	Spain	Yunquera	17	3	-4.90068	36.73414	cultivated	0.100076471	0.815570588	0.0235	0.060852941	SS
FRCOR	France	Isolated trees in villages between Piana and Calvi, Corsica	11	12	8.684197	42.463769	cultivated	0.04242	0.04674	0.75508	0.15576	CM

FREST	France	Isolated trees, along the road, Esterel	0	3	6.892866	43.430742	cultivated	NA	NA	NA	NA	
FREZE	France	Grande Corniche, Eze	20	11	7.38755	43.73848	uncultivated	0.09162	0.16468	0.69239	0.051315	CM
FRTLN	France	Mont Faron, Toulon	20	10	5.92595	43.1462	uncultivated	0.12842	0.153855	0.60873	0.10899	
FRTOUET	France	Touët de l'escarène	12	10	7.37336	43.85576	uncultivated	0.183016667	0.142675	0.587333333	0.086983333	
FRVIL	France	Villefrance sur mer	21	15	7.32025	43.71703	uncultivated	0.220719048	0.043366667	0.647652381	0.088252381	
GRAST	Greece	Askatos	6	1	21.070428	38.52721	uncultivated	0.065533333	0.121633333	0.0957	0.717116667	EM
GRFAR	Greece	Agio Farago, Crete	20	6	24.779095	34.928205	uncultivated	0.605275	0.03705	0.166035	0.191635	
GRKAR	Greece	Karpatos	6	6	27.1351111	35.6310806	uncultivated	0.034833333	0.054366667	0.09155	0.819266667	EM
GRKOU	Greece	Koumoulia mountain, Rhode island	15	6	28.12085	36.350304	uncultivated	0.0277	0.038323077	0.1412	0.792769231	EM
GRLEU	Greece	Leucada island	7	2	20.694513	38.674513	uncultivated	0.034728571	0.0417	0.108142857	0.815428571	EM
GRLOU	Greece	Loutro, Crete	19	5	24.076279	35.198983	uncultivated	0.153742105	0.140257895	0.176615789	0.529352632	
GRNAX	Greece	Naxos island	10	0	25.441915	36.94804	uncultivated	0.0702375	0.0694125	0.1041875	0.756175	EM
GRPAR	Greece	Paros island	5	1	25.15753	37.018631	uncultivated	0.05315	0.044025	0.2652	0.637625	
GRXIR	Greece	Xiromero	6	2	20.898655	38.714246	uncultivated	0.034188889	0.050311111	0.322877778	0.592622222	
ITAIRO	Italy	La Roya valley, between Trucco and Airole	12	3	7.576777	43.858197	uncultivated	0.054016667	0.02525	0.486366667	0.434358333	
ITMONU	Italy	Monumental carob trees between Raguza and Noto, Sicily	11	3	14.88955556	36.88525	cultivated	0.079054545	0.098409091	0.662763636	0.159772727	CM
ITPAN	Italy	Panarea island, Sicily	5	3	15.059	38.638	uncultivated	0.04514	0.25036	0.154	0.55048	
ITPIN	Italy	Pine Forest, Raguza, Sicily	19	3	14.47618	36.87884	cultivated	0.040478947	0.027526316	0.848410526	0.083584211	CM
ITROOT	Italy	Rootstock, Patrizio farm, Raguza, Sicily	35	12	14.610246	36.878135	uncultivated	0.026829412	0.03305	0.782241176	0.157873529	CM
ITSARR	Italy	Scillaras- Sarroca, Sardena	5	3	8.95	39.128	cultivated	0.206466667	0.0824	0.505433333	0.205733333	
ITSIN	Italy	Santa barbara, Sinna, Sardena	10	5	9.4473333	39.1613111	uncultivated	0.0603	0.033718182	0.733445455	0.172518182	CM
LIARA	Lebanon	Aramoun	8	3	35.700579	34.020618	uncultivated	0.178883333	0.288166667	0.40995	0.12305	
LIENF	Lebanon	Saydit el Nourieh, Anfeh	20	10	35.68203	34.30194	uncultivated	0.10534	0.1561	0.094145	0.6444	
LIFAN	Lebanon	Fanyou	5	3	35.67538	34.063	uncultivated	0.11648	0.07442	0.04834	0.76076	EM
LIJBE	Lebanon	Jbeïl	4	0	35.671	34.116947	uncultivated	0.14005	0.357025	0.178025	0.3249	
LIMAR	Lebanon	Mar Chaaya	6	3	35.647109	33.894394	uncultivated	0.307066667	0.0957	0.057683333	0.53955	
LINAH	Lebanon	Nahr Ibrahim	12	5	35.69636	34.0851	uncultivated	0.121216667	0.231766667	0.07755	0.569466667	
LIZEK	Lebanon	Zekrit valley	14	3	35.621232	33.945713	uncultivated	0.175307143	0.251114286	0.06335	0.510207143	
MAAIT	Morocco	Aït Mansour oasis	20	8	-8.88119	29.55149	uncultivated	0.89101	0.046705	0.023005	0.039285	SM
MAASNI	Morocco	Asni, Haut atlas	4	0	-7.972671	31.222189	uncultivated	0.48925	0.05135	0.291575	0.167825	
MADA	Morocco	Village de pêche de Dalia	11	3	-5.4772	35.90411	cultivated	0.060122222	0.533766667	0.252266667	0.153855556	
MAIMO	Morocco	Imouzzer des Ida-Outanane	11	6	-9.23713	30.73312	cultivated	0.722890909	0.069618182	0.087263636	0.120227273	SM

MAIMO2	Morocco	Imouzzer des Ida-Outanane	20	7	-9.4956	30.6557	cultivated	0.818785	0.06206	0.062775	0.05638	SM
MAKEL	Morocco	Oued Laou canyon	18	8	-5.22779	35.29865	cultivated	0.056188889	0.16455	0.149683333	0.629572222	
MAKHA	Morocco	Entre Dalia et Tetouan	10	7	-5.51225	35.6758	cultivated	0.11929	0.17397	0.12556	0.58119	
MALAO	Morocco	Oued Laou	21	6	-5.16468	35.39042	cultivated	0.027333333	0.352252381	0.478790476	0.141619048	
MALIL	Morocco	Oued Mliliah, entre Chefchaouen et Ouazanne	10	0	-5.3539	35.0429	cultivated	0.09656	0.21109	0.2047	0.48764	
MAOEJ	Morocco	Oued ElJehba, proximité village de Smih	10	5	-4.70325	35.16908	cultivated	0.04878	0.12083	0.16229	0.66809	
MAOUD	Morocco	Douar Oudjgal	10	4	-9.28211	29.75888	cultivated	0.62182	0.16007	0.07976	0.13836	
MAOUM	Morocco	Assif Oumarhouz canyon	23	9	-9.26054	29.7491	uncultivated	0.663609091	0.142127273	0.082231818	0.112036364	
MAOUR	Morocco	Ourika valley	2	3	-7.756	31.351	cultivated	0.85255	0.09255	0.03295	0.02195	SM
MAPAR	Morocco	Paradise Valley	20	5	-9.54509	30.56227	uncultivated	0.766205	0.056195	0.074855	0.10273	SM
MARTA	Morocco	Between Tafraout and Aïtmansour	11	7	-8.94341	29.63326	cultivated	0.780527273	0.059090909	0.101018182	0.059372727	SM
MATAL	Morocco	Oued Laou near Talembote village	15	6	-5.21348	35.29823	cultivated	0.07902	0.082806667	0.217173333	0.621	
MATAN	Morocco	Near Aït Baha village	19	0	-9.21047	29.94297	cultivated	0.5896	0.154463158	0.091889474	0.164042105	
MATAR	Morocco	Near Aouguenz village	19	10	-9.20208	29.84102	uncultivated	0.758536842	0.108142105	0.070226316	0.063105263	SM
MATAS	Morocco	Oued Laou	0	5	-5.16468	35.39042	cultivated	NA	NA	NA	NA	
MATIZ	Morocco	Tiznit	17	15	-9.683611	29.689722	cultivated	0.596652941	0.160264706	0.087876471	0.155211765	
MAZIN	Morocco	Zinat	11	6	-5.39234	35.43856	cultivated	0.164645455	0.167654545	0.105054545	0.562636364	
PORBEN	Portugal	Benagil	20	3	-8.423525	37.090124	cultivated	0.064806667	0.084986667	0.785793333	0.064413333	CM
PORVIL	Portugal	Vila Lara	15	1	-8.391454	37.095663	cultivated	0.105573333	0.185	0.50244	0.206966667	
TURKO	Turquie	Köprülü canyon, Anthalya	10	5	31.1808333	37.1922222	uncultivated	0.04691	0.04755	0.10796	0.79758	EM

APPENDIX S2 Extended material and methods, Genbank and Dryad data accessions.

A) Amplification of plastid markers
B) Amplification, cloning and sequencing of sucrose synthase gene (SUSY)
C) Divergence time estimation analyses
D) Microsatellite genotyping
E) Coalescent simulations and prior distribution
F) Survey of palaeobotanical literature
G) Carob tree generation time

A) Amplification of plastid markers

Total DNA was extracted from ca. 25 mg of dry leaves stored in silica gel using the NucleoSpin Plant II kit (MACHEREY-NAGEL SARL, France). DNA concentrations were normalized to 5ng/µl.
Specific primers for four plastid regions were designed for *Ceratonia* species: matK gene (F: CCTTCGATACTGGGTGAAAGAT, R: CCAGACCGGCTTACTAATGGG), ccSA-ndhD spacer (F: GTGGCCTCTATGGGCTTTCT, R: TCCGTTGACAAGGTCGAAGC), rpl32-trnL spacer (F: TCTCGACCCAATAAAAACCC, R: AAGGTACCGTTGGAACTAGA) and psbD-trnT spacer (F: ATTCAATGGGTTAGGTCCAC, R: CCCTTTTAACTCAGTGGTAG). The PCR mix contained 5 µL of 5x buffer, 2 µL MgCl2 25mM, 4 µL of dNTPs (1.25mM), 1 µL of each primer 10 uM, 0,2 µL of GoTaq, 1 µL of DNA and the total volume was 25 µL. PCR program consisted on 5 min at 94ºC; followed by 30 cycles of 30 s at 94ºC, 30 s at 52ºC and 1 min at 72ºC, and a final elongation step of 7 min at 72ºC. All PCRs were carried out in a Mastercycler nexus GSX1 thermocycler (Eppendorf) and the existence of amplicons was corroborated in agarose gels. All sequences were deposited in Genbank (Table S2.1).

TABLE S2.1: Genbank accessions of plastid haplotypes found within *Ceratonia siliqua* and *C. oreothauma* for each marker.

Haplotypes	ccsa-ndhD spacer	matk	psbD-trnT spacer	rpl32-trnL spacer	Taxon	Country	Provenance
1	MK564073	MK564064	MK532008	MK532022	*C. siliqua*	Spain	ESBEN
2	MK564074	MK564065	MK532009	MK532023	*C. siliqua*	Spain	ESGRA
3	MK564075	MK564066	MK532010	MK532024	*C. siliqua*	Morocco	MAPAR
4	MK564076	MK564067	MK532011	MK532025	*C. siliqua*	France	FRTLN
5	MK564077	MK564068	MK532012	MK532026	*C. siliqua*	Italy	SIPATT
6	MK564078	MK564069	MK532013	MK532027	*C. siliqua*	Lebanon	Jbeil
7	MK564079	MK564070	MK532014	MK532028	*C. oreothauma*	Oman	OMBG
8	MK564080	MK564071	MK532015	MK532029	*C. oreothauma*	Oman	E00023536
9	MK564081	MK564072	MK532016	MK532030	*C. oreothauma*	Yemen	E00277381

The concatened alignment of the four markers is available in Dryad Digital Repository (https://doi.org/10.5061/dryad.k7m020r). The order is ccsa-ndhD (387 bp long), the matk (624 bp long), the psbD-trnT (314 bp long) and rpl32-trnL (164 bp long).

B) Amplification, cloning and sequencing of sucrose synthase gene (SUSY)

The nuclear sucrose synthase (SUSY) gene was amplified in two samples of *Ceratonia siliqua* and one sample of *Ceratonia oreothauma* by using the primers designed by Manzanilla *et al.* (2012). The amplicon size was about 700 bp. The PCR mix contained 5 µL of 5x buffer, 1.5 µL MgCl2 25mM, 4 µL of dNTPs (1.25 mM), 0.4 1.25 µL of each primer (10 uM), 0,2 µL of GoTaq, 1 µL of DNA in a total volume was 25 µL. The PCR program consisted on 5 min at 94ºC; followed by 30 cycles of 30 s at 94ºC, 30 s at 54ºC and 45s at 72ºC, and a final elongation step of 7 min at 72ºC. All PCRs were carried out in a Mastercycler nexus GSX1 thermocycler (Eppendorf). Amplicons of the expected size were cut from an agarose gel and purified using the Wizard SV Gel kit (Promega). Purified bands were then cloned in a pGEMt A1360 plasmid vector (Promega) and eight clones per amplicon were sequenced. Reverse and forward chromatograms were assembled and revised with Sequencher (version 4.5, GeneCodes Corp., Ann Arbor, MI, USA) and alignments were performed in MEGA 7. Genbank accessions of SUSY sequences are indicated in Table S2.2.

TABLE S2.2: Genbank accessions of carob SUSY sequences used for divergence time analyses

Sample	Country	Provenance	SUSY 1	SUSY 2
Ceratonia siliqua sample 1	Spain	Ojen, ESOJ365	MK602656	MK634743
Ceratonia siliqua sample 2	Lebanon	Aramoun, LIARA205	MK602657	MK634744
Certonia oreothauma	Oman	E00023536	MK602658	MK634745

C) Divergence time estimation analyses

SUSY network was confidently split into two different paralogues and subsequently two matrices were elaborated corresponding with the copies already detected by Manzanilla (2012): SUSY1 and SUSY2. We used the available sequences at Genbank for matK and SUSY for other Leguminosae species to carry out divergence time estimation analyses. The alignment used for divergence time analysis is available in Dryad Digital Repository (https://doi.org/10.5061/dryad.k7m020r).

The best nucleotide substitution model was calculated with Jmodeltest2: GTR+G for matK, HKY+I+G for SUSY1 and SUSY2. The introns were removed due to difficulties with the alignments. Heterogeneity tests were calculated to estimate the incongruence between matrices following Viruel et al. (2018). Significant differences in topology were determined for both copies of SUSY1 and incongruences was tested by eye among the phylogenetic trees obtained for independent analyses.

Divergence times were estimated using a Bayesian relaxed-clock approach implemented in BEAST v. 1.8.3 (Drummond and Rambaut, 2007). Both concatenated and independent data sets were run under the selected substitution models, a Yule speciation process prior, and an uncorrelated lognormal molecular clock. The ucld.mean was assigned a uniform prior distribution of initial = 1e-4, lower = 1e-4, upper = 1e-1 value (Viruel et al. 2016). Two MCMC chains were run for 100 million generations, sampling parameters every 10,000th generations, and convergence was evaluated by the effective sampling size (ESS > 200) in TRACER v. 1.6. Post-burnin trees were summarized into a maximum clade credibility tree with mean values and 95% confidence intervals using TREEANNOTATOR v. 1.8.1.

We used four calibration points. A primary calibration was based on a fossil record of *Ceratonia* attributed to *Ceratonia emarginata* Heer from CE Europe and dated at Mid to Late Oligocene

(Palamarev 1989). This fossil was used to calibrate the *Ceratonia* stem node with a lognormal prior distribution (mean in real space = 28.1, stdev = 0.2). Three secondary calibrations based on a previous divergence time estimation study (Lavin et al. 2005) were also used with a normal prior distribution: i) 54.0 ± 3.4 assigned to the *Umtiza* stem node, ii) 58.6 ± 0.25 applied to the papilionoid stem node, and iii) 59 ± 0.2 set to the legume stem node.

References:

Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic biology*, 54, 575-594.

Manzanilla, V., & Bruneau, A. (2012) Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. *Molecular Phylogenetics and Evolution*, 65, 149-162

Palamarev E. (1989) Paleobotanical evidences of the Tertiary history and origin of the Mediterranean, sclerophyll dendroflora. *Plant Systematics and Evolution*, 162, 93–107.

Viruel, J., Segarra-Moragues, J. G., Raz, L., Forest, F., Wilkin, P., Sanmartin, I., & Catalan, P. (2016). Late Cretaceous–Early Eocene origin of yams (*Dioscorea*, Dioscoreaceae) in the Laurasian Palaearctic and their subsequent Oligocene–Miocene diversification. *Journal of Biogeography*, 43, 750-762.

Viruel, J., Forest, F., Paun, O., Chase, M. W., Devey, D., Sousa Couto, R., Segarra-Moragues, J. G., Catalan, P., & Wilkin, P. (2018). A nuclear *Xdh* phylogenetic analysis of yams (*Dioscorea*: Dioscoreaceae) congruent with plastid trees reveals a new Neotropical lineage. *Botanical Journal of the Linnean Society*, 187, 232-246.

D) Microsatellite genotyping

Viruel et al. (2018) described the laboratory and bioinformatic methods to amplify, sequence and genotype the 17 markers used in this study. The SSR data and results of STRUCTURE are available in Dryad Digital Repository (https://doi.org/10.5061/dryad.k7m020r).

References:

Viruel, Juan, et al. 2018 Advances in genotyping microsatellite markers through sequencing and consequences of scoring methods for *Ceratonia siliqua* (Leguminosae). *Applications in Plant Sciences* 6.12: e01201. https://doi.org/10.1002/aps3.1201

E) Coalescent simulations and prior distribution

All simulations were done in Fastsimcoal 2.6 (Excoffier et al., 2011), summary statistics used were computed using ArlSumStat in Arlequin (Excoffier & Lischer, 2010). Fifty diploid individuals per population were randomly selected to compute the summary statistics.

Population sizes (N) were drawn from a uniform distribution [lower bound 10; higher bound 10000 individuals]. Nodes timings from a uniform distribution [1; 10000]; Tanc is the most ancient node, T2 the second node and T1 the third node forward in time. We enforced Tanc>T2>T1 for the ERH, Tanc>T1 and Tanc>T2 for the TRH. Generalized stepwise mutation (GSM) model was used for microsatellite mutation model (Estoup et al 2002). SSR mutation rate (mutssr) was drawn from a LogUnif distribution [0.0001; 0.0008], and the probability that the SSR size change for more than one repetition per mutation (P) was drawn from a log uniform distribution [0.1; 0.3], according to reviews (Estoup et al., 2002; Petit & Hampe 2006).

References:

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology*, 11, 1591-1604.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources, 10, 564-567

Excoffier, L., & Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics, 27, 1332-1334

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, 37, 187-214.

Table S2.3: Parameters and prior used for coalescent simulation in FASTSIMCOAL 2.6

Parameter	Range	Distribution	Meaning
T1	1-10000	uniform	First divergence from present
T2	1-10000	uniform	Second divergence from present
Tanc	1-10000	uniform	Ancestral divergence
NSM	10-10000	uniform	Current effective size SM
NSS	10-10000	uniform	Current effective size SS
NCM	10-10000	uniform	Current effective size CM
NEM	10-10000	uniform	Current effective size EM
Nanc	10-10000	uniform	Ancestral effective size
mutssr	0.0001-0.0008	loguniform	Mutation rate per locus
p	0.1-0.3	loguniform	Parameter for the GSM model

Table S2.4: Summary statistics of genetic diversity based on SSR markers used for statistical phylogeography of *Ceratonia siliqua*. Only the population having membership above the median of the cluster were kept. SM= South Morocco, SS = South Spain, CM = central Mediterranean, EM = eastern Mediterranean. *K* is the mean number of alleles by locus, *H* is the expected heterozygosity, *Fst* and $\delta\mu^2$ are two different indices of differentiation.

	SM	**SS**	**CM**	**EM**	**mean**	
K	5,06	4,53	4,53	4,06	4,54	
H	0.58	0.55	0.46	0.50	0.52	
	SS_CM	**EM_CM**	**EM_SS**	**SM_CM**	**SM_SS**	**SM_EM**
Fst	0.14	0.06	0.12	0.10	0.08	0.11
$\delta\mu^2$	1.42	0.32	1.64	1.50	0.77	1.65
Overall Fst	10.2 %					

E) Carob Distribution Modelling

Occurrence records of *C. siliqua* were gathered across its native range in the Mediterranean basin from several sources. We extracted information from the survey of phytosociological study by Baumel et al. (2018), we also collected the data available at the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/) and we added the geographic coordinates of the populations sampled in this study (Table 1). We cleaned the dataset by discarding duplicates, records older than 1950, records located in the sea, and those having integer coordinates or coordinates equal to zero. We tried to eliminate all occurrences coming from carob orchards. To avoid duplicates in further analyses, we kept only one record for each occupied 2.5 arc-minutes (~5km) grid cell covering the species range. The final dataset consisted in 619 occurrence points (available in Dryad Digital Repository: https://doi.org/10.5061/dryad.k7m020r).

We extracted the climatic variables at a 2.5 arc-min resolution from the version 1.4 of the Worldclim database (http://worldclim.org) for each of the occurrence points. To avoid multi-collinearity among the 19 bioclimatic variables, we selected only six based on a principal component analysis (PCA) and a correlation matrix (results not shown): temperature seasonality (BIO4), maximum temperature of the warmest month (BIO5), mean temperature of the wettest (BIO8) and coldest quarters (BIO11), precipitation of both the warmest (BIO18) and coldest quarter (BIO19). As different species distribution modelling techniques may produce divergent predictions, we used a consensus approach (Marmion et al., 2009). We ran four different probabilistic models with the biomod2 package in R (Thuiller et al., 2009): Generalized Additive Model (GAM), Generalized Boosting Model (GBM), RandomForest (RF), and MaxEnt. These statistical models are calibrated with both presence and absence (or background for MaxEnt) points. No true absence information is available for *C. siliqua* across its range, so we ran the models with five different sets of 5,000 pseudo-absence/background points randomly sampled across the study area (Barbet-Massin et al., 2012). We assigned the same six climatic variables to these points as for occurrence records. We used 70% of the dataset to calibrate the models and the remaining 30% to evaluate them. This data-splitting strategy was repeated ten times and we evaluated the model predictive accuracy using two different indices: the True Skill Statistics (TSS, Allouche et al., 2006)

and the area under the Receiver Operating Characteristic curve (ROC, Hanley & McNeil, 1982). A total of 400 distribution models was ran (four algorithms, five pseudo-absence/background selections, two evaluation indices, and ten repetitions). The distribution of *C. siliqua* was projected under current, Mid-Holocene (~6,000 years ago), Last Glacial Maximum (LGM; ~22,000 years ago), and Last Inter-Glacial (LIG; ~120,000 – 140,000 years ago) climatic conditions. Mid-Holocene and LGM climate layers were generated by the Paleoclimate Modelling Intercomparison Project (Braconnot et al., 2007) and LIG ones by Otto-Bliesner et al. (2006). All were downloaded from Worldclim: Mid-Holocene and LGM data at a 2.5 arc-min resolution and LIG maps at a 30 arc-seconds resolution. The latter were then upscaled to 2.5 arc-min using the aggregate function from the raster package in R. We then projected the species distribution at Mid-Holocene and LGM according to three Global Circulation Models (GCMs); CCSM4, MIROC-ESM, and MPI-ESM-P and at the LIG according to projections from Otto-Bliesner et al. (2006). Finally, we generated one consensus map for each time slice and GCM by computing a mean of the 400 maps obtained previously weighted by their respective evaluation score (i.e. TSS score) and discarding models with TSS<0.6 and ROC<0.8. All the results are shown in Appendix 3 Figure S3.7.

References:

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43, 1223-1232.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3, 327-338.

Baumel A., Mirleau P., Viruel J., Kharrat M.B.D., Malfa S.L., Ouahmane L., ... Médail F. (2018) Assessment of plant species diversity associated with the carob tree (Ceratonia siliqua, Fabaceae) at the Mediterranean scale. *Plant Ecology and Evolution*, 151, 185–193.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J. Y., Abe-Ouchi, A., ... & Kageyama, M. (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and large-scale features. *Climate of the Past*, 3, 261-277.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15, 59-69.

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., & Hu, A. (2006). Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science*, 311, 1751-1753.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369-373

F) Survey of palaeobotanical literature

Concerning the oldest occurrences (i.e. those prior to 3 Ma), there are nine localities with *Ceratonia* macro-remains (leaf imprints) and eighteen localities with pollen grains (Table S2.5).

For most of the specimens older than 12 Ma, the leaf imprints have been attributed to the fossil species *C. emarginata* Al. Braun (localities n° 55, 51, 47, 44) or *C. vetusta* Sap. (locality n° 54), which are considered very close to the modern species *C. siliqua* (De Saporta, 1863; Palamarev *et al*., 2000; Akgün *et al*., 2004; Hably, 2006). Other leaf imprint of *C. emarginata* was found in Oeningen (Germany) dated in the 'Neogene' (Collection of Palaeontology of the National Museum of Natural History of Paris), the lack of precision in this dating prevented us to include it in Fig. 4 and Table S2.5. Those specimens younger than 4 Ma are all attributed to the modern *C. siliqua*, being wood charcoals or burned pods.

Considering the localities older than 1 Ma, fossil pollen grains ascribed to *Ceratonia* are very rare. Photographs from seven specimens are shown Fig. S2.1 (views 14 to 28), distributed from 15 (Langhian) to 1.4 Ma (Calabrian). They show the same morphology as the modern pollen of *C. siliqua* and a similar variability in size (Fig. S2.1: views 1 – 13). They significantly differ from the pollen of the sister species *C. oreothauma* Hillc., Lewis & Verde in the number of apertures and surface ornamentation (Ferguson, 1980).

Scarcity of carob pollen grains is explained by the insect pollination prevalent mode of the plant that results in its under-representation as observed in modern pollen floras (Rossignol, 1969; Weinstein, 1981; Bottema & Sarpaki, 2003). This handicap is accentuated by the pollen transport supplied by rivers in most of the localities deposited in marine coastal conditions. Among the two alone lacustrine localities, Rubielos de Mora (n° 50) and Camerota (n° 26), the latter shows very propitious conditions to have recorded fossil pollen grains of *Ceratonia* by the very small size of the palaeo-lake and its deposition context at a glacial-interglacial transition. These peculiarities made of Camerota an exceptional site with the expression of an initial *Oleo-Ceratonion* palaeo-vegetation as shown in the main text of the paper. All the pollen localities younger than 1 Ma (Table S2.5 n° 24 to 1) show very low percentages of *C.siliqua* except a sample from the Bolomar Cave (locality n° 24).

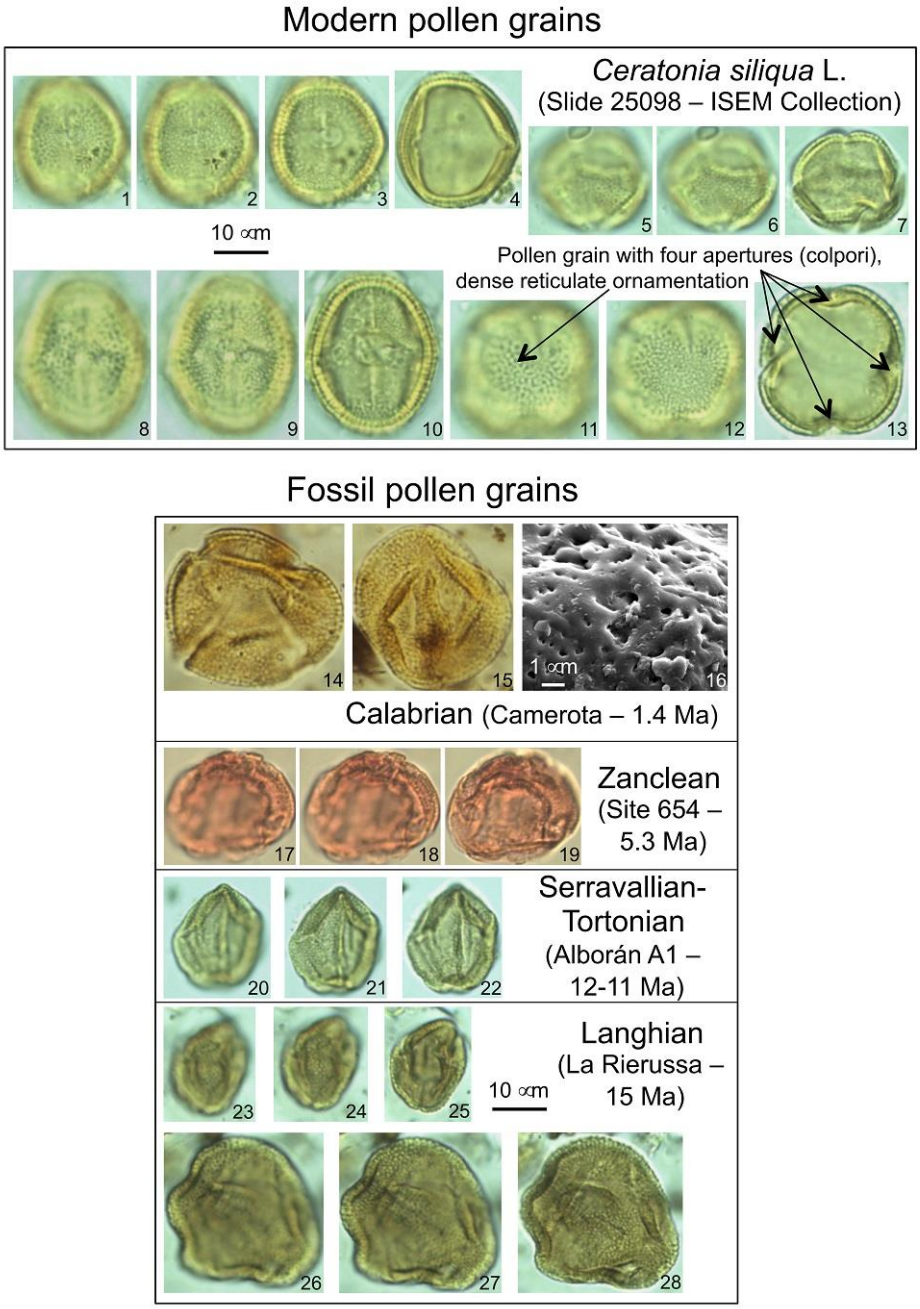


Figure S2.1: Modern pollen grains of *Ceratonia siliqua* L. and examples of fossil pollen grains ascribed to *Ceratonia*. The scale bar (10 µm) is the same for all the photographs at light microscope (X 1000), it differs (1 µm) for view 16 at scanning electron microscope (X 10000). Description of morphological terms refers to Punt *et al*., (2007).

Images 1 – 13, Modern pollen grains of *Ceratonia siliqua* L. (slide 25098 from the Collection of the Institut des Sciences de l'Evolution de Montpellier, ISEM – specimen from Lebanon at the Herbarium of the National Museum of Natural History in Paris).

1 – 4, LO-analysis of a small pollen in equatorial view.

1, Focus on the reticulate ornamentation of ectexine.

2, Focus on the base of luminae.
3, Focus on a colporus (front view).
4, Focus on two colpori (profile view) and optical section. The fourth colporus is visible in transparency in the middle part of the pollen.

5 – 7, LO-analysis of a small pollen in polar view.
5, Focus on the reticulate ornamentation of ectexine.
6, Focus on the base of luminae.
7, Focus on the four colpori and optical section.

8 – 10, LO-analysis of a large pollen in equatorial view.
8, Focus on the reticulate ornamentation of ectexine.
9, Focus on the base of luminae and a colporus (front view).
10, Focus on two colpori (profile view) and optical section. The fourth colporus is visible in transparency in the middle part of the pollen.

11 – 13, LO-analysis of a large pollen in polar view.
11, Focus on the reticulate ornamentation of ectexine.
12, Focus on the base of luminae.
13, Focus on the four colpori and optical section.

Images 14 – 27, Fossil pollen grains ascribed to *Ceratonia*.

14 – 16, Large pollen from Camerota in Southern Italy (Calabrian, Early Pleistocene), old of 1.4 Ma (Suc & Popescu, 2005).
14, Polar view of a pollen grain showing the reticulate ornamentation of ectexine, the four colpori and optical section.
15, Equatorial view showing the reticulate ornamentation of ectexine, the base of luminae, and the four colpori.
16, Surface of a pollen grain at scanning electron microscope (for comparison with modern pollen of *C. siliqua*, see Fig. A from Ferguson, 1980).

17 – 19, LO-analysis of a medium-size pollen in equatorial view from Site 654 in the Western Tyrrhenian Sea (early Zanclean, Pliocene), old of 5.3 Ma (Popescu, unpublished).
17, Focus on the reticulate ornamentation of ectexine.
18, Focus on the base of luminae.
19, Focus on two colpori (front view).

20 – 22, LO-analysis of a small pollen in equatorial view from hole Alborán A1 in the Alborán Sea (Serravallian-Tortonian, Mid-Late Miocene), old of 12-11 Ma (Jiménez-Moreno & Suc, 2007).
20, Focus on the reticulate ornamentation of ectexine.
21, Focus on the base of luminae and a colporus (front view).
22, Focus on two colpori (profile view).

23 – 25, LO-analysis of a small pollen in equatorial view from La Rierussa section in Catalunya (Langhian, Mid-Miocene), old of 15 Ma (Jiménez-Moreno & Suc, 2007).
23, Focus on the reticulate ornamentation of ectexine.
24, Focus on the base of luminae.
25, Focus on two colpori (profile view), the two other colpori are visible in transparency.

26 – 28, LO-analysis of a large pollen in a slantwise polar view from La Rierussa section in Catalunya (Langhian, Mid-Miocene), old of 15 Ma (Jiménez-Moreno & Suc, 2007).
26, Focus on the reticulate ornamentation of ectexine.
27, Focus on the base of luminae.
28, Survey of the four colpori.

On the basis of the LO-analysis method (i.e., successive focuses according to Lux vs. Obscuritas) at light microscope (Erdtman, 1952), the pollen grain of *C. siliqua* shows a homogenous morphology despite important variations in size[1] (Figs. 1 – 13). As described by Ferguson (1980), it is spheroidal

[1] Description of morphological terms refers to Punt *et al*. (2007).

somewhat longiaxe (= prolate) with a 4-lobed outline in polar view, because of the occurrence of four apertures. Each aperture is a colporus made of a relatively long colpus in the ectexine (ectoaperture) and a relatively large porus elongated along the equatorial axis in the endexine (endoaperture). Each porus is bordered by a thin annulus. Endexine shows the same thickness than the semitectate ectexine. Ectexine ornamentation is a dense homobrochate reticulum with muri thinner than luminae. There is no margo along the colpi. *C. siliqua* has predominantly a tetracolporate pollen (i.e., with 4 colpori) whereas *C. oreothauma* has predominantly a tricolporate pollen (i.e., with 3 colpori) with a reticulum showing equal size of muri and luminae (Ferguson, 1980).

Pollen grain of *Citrus* (Rutaceae) is also tetracolporate and reticulate. It can be distinguished from that of *Ceratonia siliqua* by its thicker endexine, the occurrence of a thin margo and costae along the colpi, a thicker annulus around the pori, and an equal thickness of muri and luminae of the reticulum.

Table S2.5. Summary of known fossil occurrences of *Ceratonia* from Oligocene to Middle Ages. Ages are indicated according to the chronostratigraphic charts in thousand years (ka) or in million years (Ma). p = pollen grains; m = macroremains (leaves, fruits, seeds, wood).

No (Fig.4)	Age	ka	Fossil type	Locality	Coordinates	Reference
1	Middle Ages	0.7	m	Rock of Ifach, Calp, SE. Spain	38.6381°N – 0.0736°E	Ntinou *et al*., 2013
2	Middle Ages	1.9	p	Psatha, Corinth Gulf, Greece	38.1068°N – 23.2215°E	Jahns, 2003
3	Antiquity	2	m	Herculanum, S. Italy	40.8060°N – 14.3474°E	Meyer, 1980
4	Antiquity	3	p	Ein Feshkha, Jericho, West Bank	31.7221°N – 35.4612°E	Neumann *et al*., 2007
5	Antiquity	3	p	Katochi, W. Greece	38.4296°N – 21.1621°E	Fouache *et al*., 2005
6	Antiquity	3	p	Lake Hula, N. Israel	33.1065°N – 35.6057°E	Van Zeist *et al*., 2009
7	Antiquity	3.1	p	Mljet, S. Dalmatian Islands	42.7792°N – 17.3472°E	Jahns & van den Bogaard, 1998
8	Antiquity	3.3	p	Dalmatian Islands	43.0541°N – 16.7964°E	Šoštarić, 2005
9	Antiquity	3.5	p	Kournas, Crete	35.32°N – 24.2909°E	Bottema & Sarpaki, 2003
10	Holocene	4	p	Laguna de Zoñar, S. Spain	37.4814°N – 4.6938°W	Martín-Puertas *et al*., 2008
11	Holocene	4	m	Bastida de Totana, SE. Spain	38.7592°N – 1.5597°W	Núñez *et al*., 1988
12	Holocene	6	p	Addaia lagoon, Minorca, E. Spain	39.9898°N – 4.2060°E	Servera-Vives *et al*., 2018
13	Holocene	7.5	m	Cave V/49, Jebel Quruntul, N. Israel	31.8718°N – 35.4310°E	Melamed, 2002
14	Holocene	7.5	m	Aktopraklık, NW. Turkey	40.1744°N – 28.7701°E	Schroedter & Nelle, 2015
15	Holocene	11	p	Ifri Oudadane, N. Morocco	35.2056°N – 3.2978°W	Zapata *et al*., 2013
16	Holocene	11	m	Tell es-Sultan, West Bank	31.8046°N – 35.4856°E	Asouti *et al*., 2015
17	Earliest Holocene	11.3	p	Salada Mediana, Ebro Valley, NE. Spain	41.5018°N – 0.7326°W	González-Sampériz *et al*., 2004
18	Late Pleistocene	12	m	Jericho, West Bank	32.1304°N – 35.7632°E	Western, 1971
19	Late Pleistocene and Antiquity	35 to 7, and ca. 2	p	Nahr Ibrahim Cave, Lebanon	34.0662°N – 35.6452°W	Leroi-Gourhan, 1973, 1980
20	Late Pleistocene	65	p	Nahal Mahanayeem Outlet, N. Israel	33.0332°N – 35.6313°E	Aharonovich *et al*., 2014
21	Mid- to Late Pleistocene	350 to 40	p	Core PRGL1-4, NW. Mediterranean	42.6930°N – 3.8416°E	Suc *et al*., 2018
22	Late Pleistocene	112 to 62	p	Lazaret Cave, SE. France	43.6903°N – 7.2948°E	Lebreton *et al*., 2007
23	End of the Mid-Pleistocene	ca. 125	p	Batroum, Lebanon	34.2617°N – 35.6605°W	Leroi-Gourhan, 1973, 1980
24	Mid-Pleistocene	600 to 120	p	Bolomar Cave, SE. Spain	39.0594°N – 0.2416°W	Ochando *et al*., 2019
	Age	**Ma**				
25	Early Pleistocene	1	p	Vallonnet Cave, SE. France	43.7646°N – 7. 4702°E	Renault-Miskovsky & Girard, 1978
26	Calabrian, Early Pleistocene	1.4	p	Camerota, S. Italy	40.0386°N – 15.3707°E	Brenac, 1984
27	Gelasian, Early Pleistocene	2.3	p	WellAndalucia G1, SW. Mediterranean	36.3963°N – 4.7522°W	Feddi *et al*., 2011
28	Gelasian, Early Pleistocene	2.5 to 1.8	p	Crotone, S. Italy	39.0347°N – 17.1524°E	Combourieu-Nebout, 1990
29	Piacenzian, Late Pliocene	3.5	m	Ichkeul Lake, N. Tunisia	37.2028°N – 9.6623°E	Arambourg *et al*., 1952
30	Zanclean, Early Pliocene	3.8	p	Well Tarragona E2, NW. Mediterranean	40.8425°N – 1. 1458°E	Bessais & Cravatte, 1988
31	Zanclean, Early Pliocene	4.95	p	Husnicioara, Turnu Severin, SW. Romania	44.6689°N – 22.7599°E	Popescu *et al*., 2006

32	Zanclean, Early Pliocene	5.3	p	Malaga, S. Spain	36.7197°N – 4.4972°W	Suc, unpublished
33	Zanclean, Early Pliocene	5.3	p	Site 654, W. Tyrrhenian Sea	40.5793°N – 10.6966°E	Popescu, unpublished
34	Messinian, Late Miocene	5.37	p	Maccarone, Apiro, E. Italy	43.4090°N – 13.1042°E	Bertini, 1994
35	Early Pontian, Messinian, Late Miocene	6	m	Kodor River, Kutaisi, Georgia	42.2967°N – 42.7164°E	Axelrod, 1975
36	Messinian, Late Miocene	6	p	Ifounassene, N. Morocco	35.3270°N – 2. 9587°E	Bachiri Taoufiq, unpublished
37	Messinian, Late Miocene	6.2	p	Carmona, SW. Spain	37.4577°N – 5.6502°W	Fauquette *et al*., 2006
38	Tortonian and Messinian, Late Miocene	8 - 6	p	Capodarso, Enna, Sicily, S. Italy	37.5167°N – 14.1453°E	Suc *et al*., 1995
39	Early Sarmatian, Late Miocene	10 - 9	m	Bursuc, Moldavia	47.9724°N – 28.7414°E	Iamandei *et al*., 2005
40	Tortonian, Late Miocene	11	p	Well Andalucia A1, SW. Mediterranean	36.5833°N – 2.7166°W	Jiménez-Moreno & Suc, 2007
41	Serravallian-Tortonian, Mid-Late Miocene	12 - 11	p	Saouaf Basin, Tunisia	36.2281°N – 10. 1717°E	Ben Moktar & Mannaï-Tayech, 2012
42	Serravallian-Tortonian, Mid-Late Miocene	12 - 11	p	Well Alborán A1, SW. Mediterranean	36.6333°N – 4.2230°W	Jiménez-Moreno & Suc, 2007
43	Early Sarmatian, Mid-Miocene	12.6	m	Dédestapolcsány, NE. Hungary	48.1801°N – 20.4850°E	Hably, 2006
44	Serravallian, Mid-Miocene	13	m	Soma, W. Turkey	37.8869°N – 29.1434°E	Akgün *et al*., 2007
45	Langhian, Mid-Miocene	14	p	Gor, SE. Spain	37.3766°N – 2.9666°W	Jiménez-Moreno & Suc, 2007
46	Langhian, Mid-Miocene	15	p	La Rierussa, Gelida, NE. Spain	41.4527°N – 1.8655°E	Jiménez-Moreno & Suc, 2007
47	Lower Badenian, Mid-Miocene	16	m	Szurdokpüspöki, N. Hungary	47.8639°N – 19.7154°E	Hably, 2006
48	Burdigalian, Early Miocene	16	p	V. Hugo College, Narbonne, S. France	43.1880°N – 3.0060°E	Bessedik *et al*., 1984
49	Burdigalian, Early Miocene	17	p	Estagel, Istres, S. France	43.5187°N – 4.9975°E	Jiménez-Moreno & Suc, 2007
50	Burdigalian, Early Miocene	17.5	p	Rubielos de Mora, E. Spain	40.1855°N – 0.7913°W	Jiménez-Moreno *et al*., 2010
51	Ottnangian-Karpatian, Early Miocene	18	m	Abaliget, S. Hungary	46.1427°N – 18.1143°E	Hably, 2006
52	Burdigalian, Early Miocene	19	p	Häutligen, Bern, Switzerland	46.8572°N – 7.6052°E	Jiménez-Moreno, 2005
53	Aquitanian, Early Miocene	23	p	Portel, S. France	43.0519°N – 2.9286°E	Bessedik, 1984
54	Aquitanian, Early Miocene	23	m	Aix en Provence, SE. France	43.5431°N – 5.4317°E	De Saporta 1863
55	Early Oligocene	30	m	Eleshnitsa, SW. Bulgaria	41.8702°N – 23.6189°E	Palamarev *et al*., 2000

References.

Aharonovich, S., Sharon, G., & Weinstein-Evron, M. (2014) Palybnological investigations at the Middle Palaeolithic site of Nahal Mahanayeem Outlet, Israel. *Quaternary International*, 331, 149–166.

Akgün, F., Kayseri, M. S., & Akkiraz, M. S. (2007) Palaeoclimatic evolution and vegetational changes during the Late Oligocene–Miocene period in Western and Central Anatolia (Turkey). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 253, 56–106.

Arambourg, C., Arenes, J., & Depape, G. (1952) Sur deux flores fossiles quaternaires d'Afrique du Nord. *Comptes Rendus de l'Académie des Sciences de Paris*, 234, 128–130.

Asouti, E., Kabukcu, C., White, C. E., Kuijt, I., Finlayson, B., & Makarewicz, C. (2015) Early Holocene woodland vegetation and human impacts in the arid zone of the southern Levant. *The Holocene*, 25, 1565–1580.

Axelrod, D. I. (1975) Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden*, 62, 280–334.

Ben Moktar, N. & Mannaï-Tayech, B. (2012) Reconstitution de la végétation et du climat durant le Miocène dans le bassin de Saouaf (Tunisie centro-nord-orientale). *Geodiversitas*, 34, 445–456.

Bertini, A. (1994) Messinian-Zanclean vegetation and climate in north-central Italy. *Historical Biology*, 9, 3–10.

Bessais, E., & Cravatte, J. (1988) Les écosystèmes végétaux pliocènes de Catalogne méridionale. Variations latitudinales dans le domaine nord-ouest méditerranéen. *Geobios*, 21, 49–63.

Bessedik, M. (1984) The early Aquitanian and upper Langhian-lower Serravallian environments in the Northwestern Mediterranean region. *Paléobiologie continentale*, 14, 153–179.

Bessedik, M., Aguilar, J.-P., Cappetta, H., & Michaux, J. (1984) Le climat du Néogène dans le sud de la France (Provence, Languedoc, Roussillon) d'après l'analyse des faunes (rongeurs, sélaciens) et des flores polliniques. *Paléobiologie continentale*, 14, 191–204.

Bottema, S., & Sarpaki, A. (2003) Environmental change in Crete: a 9000-year record of Holocene vegetation history and the effect of the Santorini eruption. *The Holocene*, 13, 733–749.

Brenac, P. (1984) Végétation et climat de la Campanie du Sud (Italie) au Pliocène final d'après l'analyse pollinique des dépôts de Camerota. *Ecologia mediterranea*, 10, 207–216.

Combourieu-Nebout, N. (1990) Les cycles glaciaire-interglaciaire en région méditerranéenne de 2,4 à 1,1 Ma : analyse pollinique de la série de Crotone (Italie méridionale). *Paléobiologie continentale*, 17, 35–59.

Fauquette, S., Suc, J.-P., Bertini, A., Popescu, S.-M., Warny, S., Bachiri Taoufiq, N., Perez Villa, M.-J., Chikhi, H., Subally, D., Feddi, N., Clauzon, G., & Ferrier, J. (2006) How much did climate force the Messinian salinity crisis? Quantified climatic conditions from pollen records in the Mediterranean region. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 238, 281–301.

Feddi, N., Fauquette, S., & Suc, J.-P. (2011) Histoire plio-pléistocène des écosystèmes végétaux de Méditerranée sud-occidentale: apport de l'analyse pollinique de deux sondages en mer d'Alboran. *Geobios*, 44(1), 57–69.

Fouache, E., Dalongeville, R., Kunesch, S., Suc, J.-P., Subally, D., Prieur, A., & Lozouet, P. (2005) The Environmental Setting of the Harbor of the Classical Site of Oeniades on the Acheloos Delta, Greece. *Geoarchaeology*: *An International Journal*, 20, 285–302.

González-Sampériz, P., Valero-Garcés, B. L., & Carrión, J. S. (2004) Was the Ebro valley a glacial refugium for temperate trees? *Anales de Biologia*, 26, 13–20.

Hably, L. (2006). Catalogue of the Hungarian Cenozoic leaf, fruit and seed floras from 1856 to 2005. *Studia Botanica Hungarica.*, 37, 41–129.

Iamandei, S., Iamandei, E., Tusa, L., & Obada, T. (2005) Sarmatian petrified wood within "Bursuc Flora"(Moldova Rep.). *Acta Palaeontologica Romaniae*, 5, 223–229.

Jahns, S. (2003) A late Holocene pollen diagram from the Megaris, Greece, giving possible evidence for cultivation of Ceratonia siliqua L. during the last 2000 years. *Vegetation History and Archaeobotany*, 12, 127–130.

Jahns, S., & van den Bogaard, C. (1998) New palynological and tephrostratigraphical investigations of two salt lagoons on the island of Mljet, south Dalmatia, Croatia. *Vegetation History and Archaeobotany*, 7, 219–234.

Jiménez-Moreno, G. (2005) Utilización del análisis polínico para la reconstrucción de la vegetación, clima y estimación de paleoaltitudes a lo largo de arco alpino europeo durante el Mioceno (21-8 Ma). PhD thesis, University of Grenade and University Lyon 1, 312 p.

Jiménez-Moreno, G., Fauquette, S., & Suc, J. P. (2010) Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. *Review of Palaeobotany and Palynology*, 162, 403–415.

Jiménez-Moreno, G., & Suc, J.-P. (2007) Middle Miocene latitudinal climatic gradient in Western Europe: evidence from pollen records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *253*, 224–241.

Lebreton, V., Lartigot, A.-S., Karatsori, E., Messager, E., Marquer, L., & Renault-Miskovsky, J. (2007) Potentiels et limites de l'analyse pollinique de spéléothèmes quaternaires : applications à la reconstitution de l'environnement végétal de l'homme préhistorique sur le pourtour nord-méditerranéen. *Quaternaire*, 18, 153–174.

Leroi-Gourhan, A. (1973) Les possibilités de l'analyse pollinique en Syrie et au Liban. *Paléorient*, 1(1), 39–47.

Leroi-Gourhan, A. (1980) Les analyses polliniques au Moyen-Orient. *Paléorient*, 6, 79–91.

Martín-Puertas, C., Valero-Garcés, B. L., Pilar Mata, M., González-Sampériz, P., Bao, R., Moreno, A., & Stefanova, V. (2008) Arid and humid phases in southern Spain during the last 4000 years: the Zonar Lake record, Cordoba. *The Holocene*, 18, 907–921.

Melamed, Y. (2002) Chalcolithic and Hellenistic plant remains from cave V/49. *Atiqot*, 41, 101–108.

Meyer, F. G. (1980) Carbonized food plants of Pompeii, Herculaneum, and the Villa at Torre Annunziata. *Economic Botany*, 34, 401–437.

Neumann, F. H., Kagan, E. J., Schwab, M. J., & Stein, M. (2007) Palynology, sedimentology and palaeoecology of the late Holocene Dead Sea. *Quaternary Science Reviews*, 26, 1476–1498.

Ntinou, M., Badal, E., Carrión, Y., Fueyo, J. L. M., Carrión, R. F., & Mira, J. P. (2013) Wood use in a medieval village: the contribution of wood charcoal analysis to the history of land use during the 13th and 14th centuries a. d. at Pobla d'Ifach, Calp, Alicante, Spain. *Vegetation History and Archaeobotany*, 22, 115–128.

Núñez, D. R., de Castro, C. O., & Martínez, A. A. (1988) Arqueobotánica y Paleoetnobotánica en el Sureste de España, datos preliminares. *Trabajos de Prehistoria*, 45, 317–334.

Ochando, J., Carrión, J.S., Blasco, R., Fernández, S., Amorós, G., Munuera, M., Sañudo, P., & Peris, J.-F. (2019) Silvicolous Neanderthals in the far West: the mid-Pleistocene palaeoecological sequence of Bolomor Cave (Valencia, Spain). *Quaternary Science Reviews*, in press.

Palamarev, E., Kitanov, G., Staneva, K., & Bozukov, V. (2000) Fossil flora from Paleogene sediments in the northern area of the Mesta Graben in the Western Rhodopes. II. Analysis and stratigraphic importance of the flora. *Phytologia Balcanica*, 6, 3–11.

Popescu, S.-M., Krijgsman, W., Suc, J.-P., Clauzon, G., Mărunțeanu, M., & Nica, T. (2006) Pollen record and integrated high-resolution chronology of the early Pliocene Dacic basin (southwestern Romania). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 238, 78–90.

Renault-Miskovsky, J., & Girard, M. (1978) Analyse pollinique du remplissage pléistocène inférieur et moyen de la grotte du Vallonnet (Roquebrune-Cap-Martin, Alpes-Maritimes). *Géologie Méditerranéenne*, 5, 385–402.

Saporta, G. D. de (1863) Études sur la vegetation du sud-est de la France a l'époque Tertiaire. Première partie, V. *Annales des Sciences Naturelles Botanique,* 4th series, 19, 159–278.

Schroedter, T. M., & Nelle, O. (2015) New insights into Mid-Holocene vegetation in the southern Marmara region: Charcoal from the Late Neolithic to Early Chalcolithic settlement site Aktopraklık, northwestern Turkey. *Quaternary International*, 366, 81–95.

Servera-Vives, G., Riera, S., Picornell-Gelabert, L., Moffa-Sánchez, P., Llergo, Y., Garcia, A., Mus-Amezquita, M., García-Álvarez, S., & Trías, M. C. (2018) The onset of islandscapes in the Balearic Islands: A study-case of Addaia (northern Minorca, Spain). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 498, 9–23.

Suc, J.-P., Popescu, S.-M., Fauquette, S., Bessedik, M., Jiménez-Moreno, G., Bachiri Taoufiq, N., Zheng, Z., & Médail, F. (2018) Reconstruction of Mediterranean flora, vegetation and climate for the last 23 million years based on an extensive pollen dataset. *Ecologia mediterranea*, 44(2), 53–85.

Šoštarić, R. (2005) The development of postglacial vegetation in coastal Croatia. *Acta Botanica Cro*atica, 64, 383–390.

Suc, J.-P., Violanti, D., Londeix, L., Poumot, C., Robert, C., Clauzon, G., Gautier, F., Turon, J.-L., Ferrier, J., Chikhi, H., & Cambon, G. (1995) Evolution of the Messinian Mediterranean enviuronments: the Tripoli Formation at Capodarso (Sicily, Italy). *Review of Palaeobotany and Palynology*, 87, 51–79.

Van Zeist, W., Baruch, U., & Bottema, S. (2009) Holocene palaeoecology of the Hula area, northeastern Israel. A Timeless Vale. Archaeological and Related Essays on the Jordan Valley in Honour of Gerrit Van Der Kooij on the Occasion of his Sixty-Fifth Birthday, edited by: Kaptijn, K. & Petit, L. P., Leiden University Press, Leiden, 29–64.

Western, A. (1971) The ecological interpretation of ancient charcoals from Jericho. *Levant*, 3, 31–40.

Zapata, L., López-Sáez, J. A., Ruiz-Alonso, M., Linstädter, J., Pérez-Jordà, G., Morales, J., Kehl, M., & Peña-Chocarro, L. (2013) Holocene environmental change and human impact in NE Morocco: Palaeobotanical evidence from Ifri Oudadane. *The Holocene*, 23, 1286–1296.

G) Carob tree generation time

The review of 28 publications of phylogeography conducted on tree species (Table S2.6) revealed an average generation time of 69 years and values ranging from 10 to 220 years, with sometimes various values for the same species (e.g. 25 and 100 years for olive tree). Assuming an age of maturity (α) of 10 years (the first fruits are observed for 7 years old carob trees in cultivation) the formula $T = \alpha + [s/\lambda\text{-}s)]$ (Cavenders-Bare et al 2011) allows to estimate the generation time in various demographic contexts. It gives a maximum generation time of 109 years for a stable population ($\lambda = 1$) and a high survival rate (0.99) and a minimum generation time of 18 years for an expanding population ($\lambda = 1.01$) and a low survival rate (0.90), leading to a median generation time of 64 years which is close to the average value from literature. We used this value to convert node ages in years ago.

Table S2.6. Generation times used for tree species in phylogeographical studies

Taxa	Generation time in years	Reference
Eight neotropical tree species	50	Barthe et al. 2017
Afzelia quanzensis	150	Jinga and Ashley 2018
Betula maximowicziana	100	Tsuda et al. 2015
Castanopsis fargesii	25	Sun et al. 2014
Euterpe globosa	101	Van Valen 1975
Juglans regia	80 - 110	Pollegioni et al. 2017
Juniperus microsperma	50	Shang et al. 2014
Malus sylvestris	7.5	Cornille et al 2015
Myrceugenia correifolia	5	Perez et al. 2017
Neolitsea sericea	10 - 30	Cao et al .2018
Two *Nothofagus* species	30 - 50	Soliani et al. 2015
Olea europea	100	Besnard et al. 2013
Olea europea	20	Diez et al. 2015
Phoenix canariensis	25	Saro et al. 2015
Populus adenopoda	40	Fan 2018
Populus tremula	15	Ingvarssson 2008
Two *Populus* species	20 – 60	Macaya-Sanz 2012
Prunus dulci	50	Delplancke et al. 2013
Prunus sibirica	20	Wang 2017
Quercus sub sec. *Virentes*	120 -220	Cavender-Bares 2011
Quercus cerris	30 -50	Bagnoli 2016
Quercus chrysolepis	50	Bemmels 2016
Two *Quercus* species	80	Yang 2016
Nine *Quercus* species	50	Ortego et al .2015
Two *Quercus* species	50 -220	San Jose-Maldia et al. 2017
Sequoiadendron giganteum	300	Dodd and DeSilva 2016
Taxus baccata	100	Mayol et al 2015
Xylocarpus granatum	10	Tomizawa 2017

References:

Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., ... & Vendramin, G. G. (2016). Combining molecular and fossil data to infer demographic history of *Quercus cerris*: insights on European eastern glacial refugia. *Journal of Biogeography*, 43, 679-690.

Barthe, S., Binelli, G., Hérault, B., Scotti-Saintagne, C., Sabatier, D., & Scotti, I. (2017). Tropical rainforests that persisted: inferences from the Quaternary demographic history of eight tree species in the Guiana shield. *Molecular Ecology*, 26, 1161-1174

Bemmels, J. B., Title, P. O., Ortego, J., & Knowles, L. L. (2016). Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative

distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, 25, 4889-4906.

Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, A., Arrigo, N., ... & Savolainen, V. (2013). The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122833.

Cao, Y. N., Wang, I. J., Chen, L. Y., Ding, Y. Q., Liu, L. X., & Qiu, Y. X. (2018). Inferring spatial patterns and drivers of population divergence of *Neolitsea sericea* (Lauraceae), based on molecular phylogeography and landscape genomics. *Molecular Phylogenetics and Evolution*, 126, 162-172.

Delplancke, M., Alvarez, N., Benoit, L., Espindola, A., Joly, H. I., Neuenschwander, S., & Arrigo, N. (2013). Evolutionary history of almond tree domestication in the Mediterranean basin. *Molecular Ecology*, 22, 1092-1104.

Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, 206, 436-447.

Dodd, R. S., & DeSilva, R. (2016). Long-term demographic decline and late glacial divergence in a Californian paleoendemic: *Sequoiadendron giganteum* (giant sequoia). *Ecology and Evolution*, 6, 3342-3355.

Fan, L., Zheng, H., Milne, R. I., Zhang, L., & Mao, K. (2018). Strong population bottleneck and repeated demographic expansions of *Populus adenopoda* (Salicaceae) in subtropical China. *Annals of Botany*, 121, 665-679.

Ingvarsson PK (2008) Multilocus patterns of nucleotidepolymorphism and the demographic history of *Populus tremula*. *Genetics*, 180, 329–340

Jinga, P., & Ashley, M. V. (2018). A mountain range is a strong genetic barrier between populations of *Afzelia quanzensis* (pod mahogany) with low genetic diversity. *Tree Genetics & Genomes*, 14, 4.

Macaya-Sanz, D., Heuertz, M., López-de-Heredia, U., de-Lucas, A. I., Hidalgo, E., Maestro, C., ... & González-Martínez, S. C. (2012). The Atlantic–Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. *Molecular Ecology*, 21, 3593-36

Mayol, M., Riba, M., González-Martínez, S. C., Bagnoli, F., Beaulieu, J. L., Berganzo, E., ... & Romšáková, I. (2015). Adapting through glacial cycles: insights from a long-lived tree (*Taxus baccata*). *New Phytologist*, 208, 973-986.

Ortego, J., Noguerales, V., Gugger, P. F., & Sork, V. L. (2015). Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology*, 24, 6188-6208.

Pérez, F., Hinojosa, L. F., Peralta, G., Montenegro, P., Irarrázabal, C., & Cossio, M. (2017). Genetic Patterns of *Myrceugenia correifolia*, a Rare Species of Fog-Dependent Forests of Mediterranean Chile: Is It a Climatic Relict? *Frontiers in Plant Science*, 8, 1097.

Pollegioni, P., Woeste, K., Chiocchini, F., Del Lungo, S., Ciolfi, M., Olimpieri, I., ... & Malvolti, M. E. (2017). Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions. *PloS one*, 12, e0172541.

San Jose-Maldia, L., Matsumoto, A., Ueno, S., Kanazashi, A., Kanno, M., Namikawa, K., ... & Tsumura, Y. (2017). Geographic patterns of genetic variation in nuclear and chloroplast genomes of two related oaks (*Quercus aliena* and *Q. serrata*) in Japan: implications for seed and seedling transfer. *Tree Genetics & Genomes*, 13, 121.

Saro, I., González-Pérez, M. A., García-Verdugo, C., & Sosa, P. A. (2015). Patterns of genetic diversity in *Phoenix canariensis*, a widespread oceanic palm (species) endemic from the Canarian archipelago *Tree Genetics & Genomes*, 11, 815.

Shang, H. Y., Li, Z. H., Dong, M., Adams, R. P., Miehe, G., Opgenoorth, L., & Mao, K. S. (2015). Evolutionary origin and demographic history of an ancient conifer (Juniperus microsperma) in the Qinghai-Tibetan Plateau. *Scientific Reports*, 5, 10216.

Soliani, C., Tsuda, Y., Bagnoli, F., Gallo, L. A., Vendramin, G. G., & Marchelli, P. (2015). Halfway encounters: Meeting points of colonization routes among the southern beeches *Nothofagus pumilio* and *N. antarctica*. *Molecular Phylogenetics and Rvolution*, *85*, 197-207.

Sun, Y., Hu, H., Huang, H., & Vargas-Mendoza, C. F. (2014). Chloroplast diversity and population differentiation of *Castanopsis fargesii* (Fagaceae): a dominant tree species in evergreen broad-leaved forest of subtropical China. *Tree Genetics & Genomes*, 10, 1531-1539.

Tomizawa, Y., Tsuda, Y., Saleh, M., Wee, A., Takayama, K., Yamamoto, T., ... & Ardli, E. (2017). Genetic structure and population demographic history of a widespread mangrove plant *Xylocarpus granatum* J. Koenig across the Indo-West Pacific region. Forests, 8, 480.

Tsuda, Y., Nakao, K., Ide, Y., & Tsumura, Y. (2015). The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. *Molecular Ecology*, 24, 1403-1418.

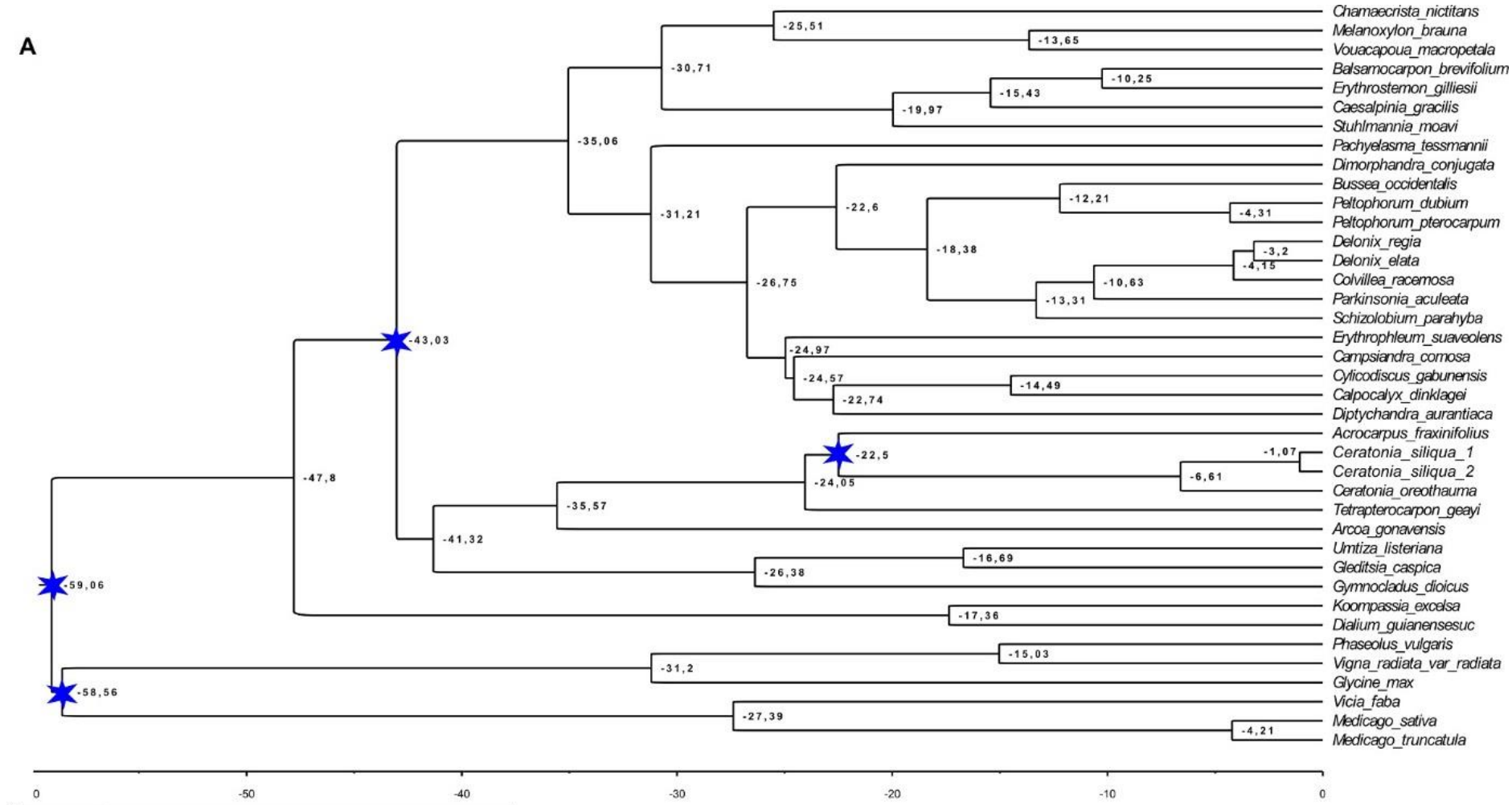
Van Valen, L. (1975). Life, death, and energy of a tree. *Biotropica*, 259-269.

Wang, Z., Zeng, Y., Zhang, Z., Sheng, S., Tian, J., Wu, R., & Pang, X. (2017). Phylogeography Study of the Siberian Apricot (*Prunus sibirica* L.) in Northern China Assessed by Chloroplast Microsatellite and DNA Makers. *Frontiers in plant science*, 8, 1989.

Yang, J., Di, X., Meng, X., Feng, L., Liu, Z., & Zhao, G. (2016). Phylogeography and evolution of two closely related oak species (*Quercus*) from north and northeast China. *Tree Genetics & Genomes*, 12, 89.

APPENDIX S3 Extended results

- Divergence time analysis
- Results of STRUCTURE HARVESTER
- Plot of admixture proportion from STRUCTURE analysis from K=2 to K=10 clusters
- Ward clustering tree based on Jost'D differentiations
- Genetic diversity map
- ABC parameter estimates and distributions
- ABC Posterior predictive simulations plots
- SDM maps for three Global Circulation Models



C

Age (Ma)	Median	95% CI
(a) *Ceratonia siliqua* crow node	1.07	0.15 - 2.6
(b) *Ceratonia* crown node	6.6	2.7 - 12
(c) *Ceratonia* stem node	22.5	16.5 - 29.5

Figure S3.1: Divergence time estimates from BEAST analysis of a concatenated matrix of matK, SUSY 1 and 2 markers (two runs of 50 M simulations). (A) Time tree, Calibrations were fixed for nodes indicated by a star. *Ceratonia siliqua* 1 and 2 are coming from Spain (Ojen) and Lebanon (Aramoun) and they are belonging to plastid haplogroup 1 and 2, respectively. (B) *Ceratonia* divergence time distributions. (C) Node ages medians and their 95% CI

A) Structure analyses

Table S3.1: Result of STRUCTURE HAVERSTER based on 10 replicates of STRUCTURE analysis for each K (2 000 000 chain length after 200 000 burn-in simulations by replicate).

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-38397.990000	0.031623	—	—	—
2	10	-37098.560000	1.963670	1299.430000	668.610000	340.489995
3	10	-36467.740000	35.077096	630.820000	8.980000	0.256008
4	10	-35845.900000	5.846176	621.840000	275.990000	47.208634
5	10	-35500.050000	25.124766	345.850000	58.510000	2.328778
6	10	-35212.710000	46.442376	287.340000	9.200000	0.198095
7	10	-34934.570000	52.053883	278.140000	62.180000	1.194531
8	10	-34718.610000	43.083986	215.960000	8.880000	0.206109
9	10	-34511.530000	37.374086	207.080000	2.940000	0.078664
10	10	-34301.510000	57.785897	210.020000	—	—

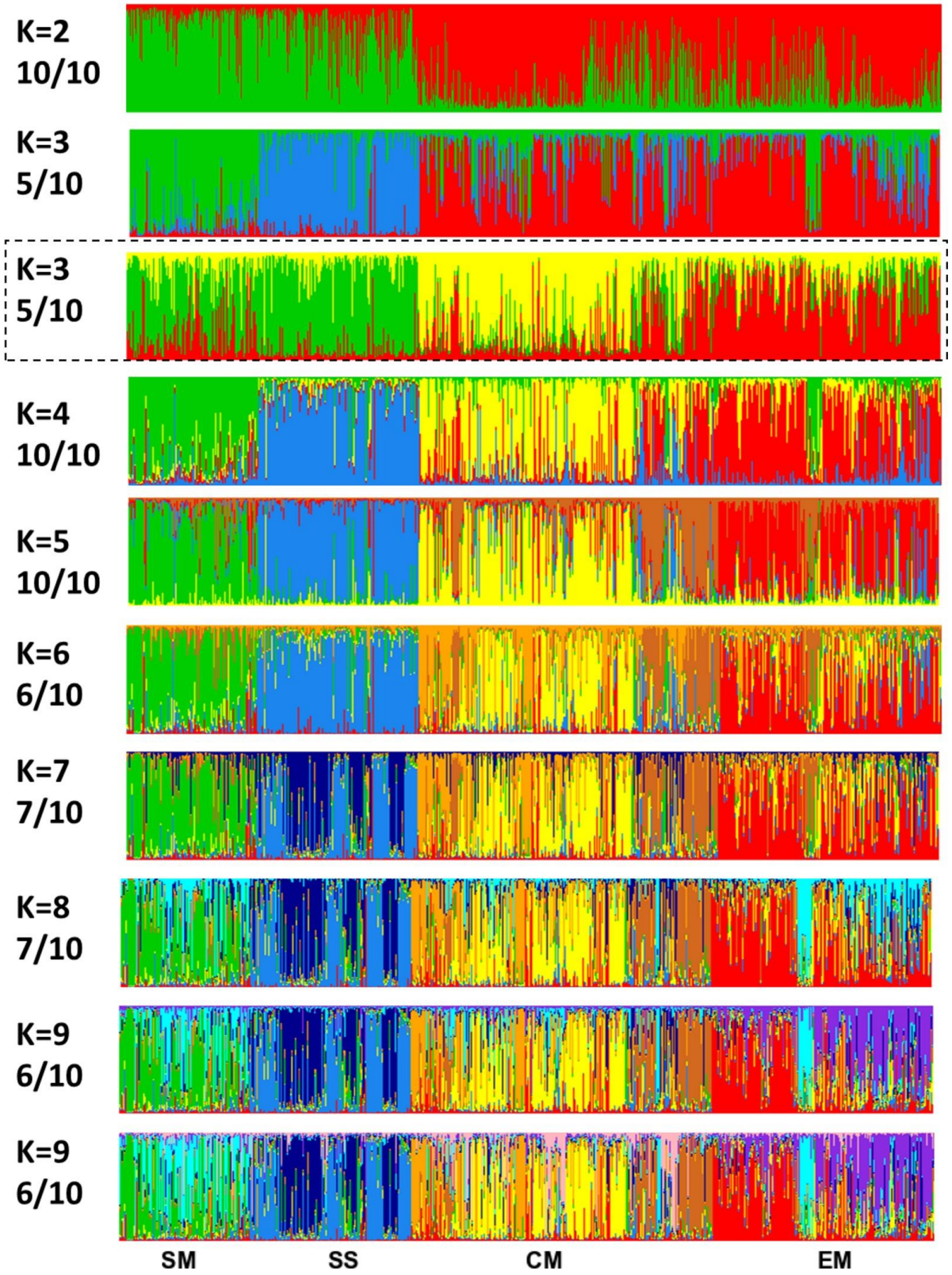


Figure S3.2. Admixture proportions inferred for 1037 carob genotypes from 67 provenances for K= 1 to K = 10 by 10 replicates of STRUCTURE (each 2 000 000 chain length after 200 000 burn-in simulations, admixture model, no population prior) analyzed by CLUMPAK. For each K only the majority solution is shown with its support for 10 replicates, excepted for K=3. The population are in the same order than in Figure 2.

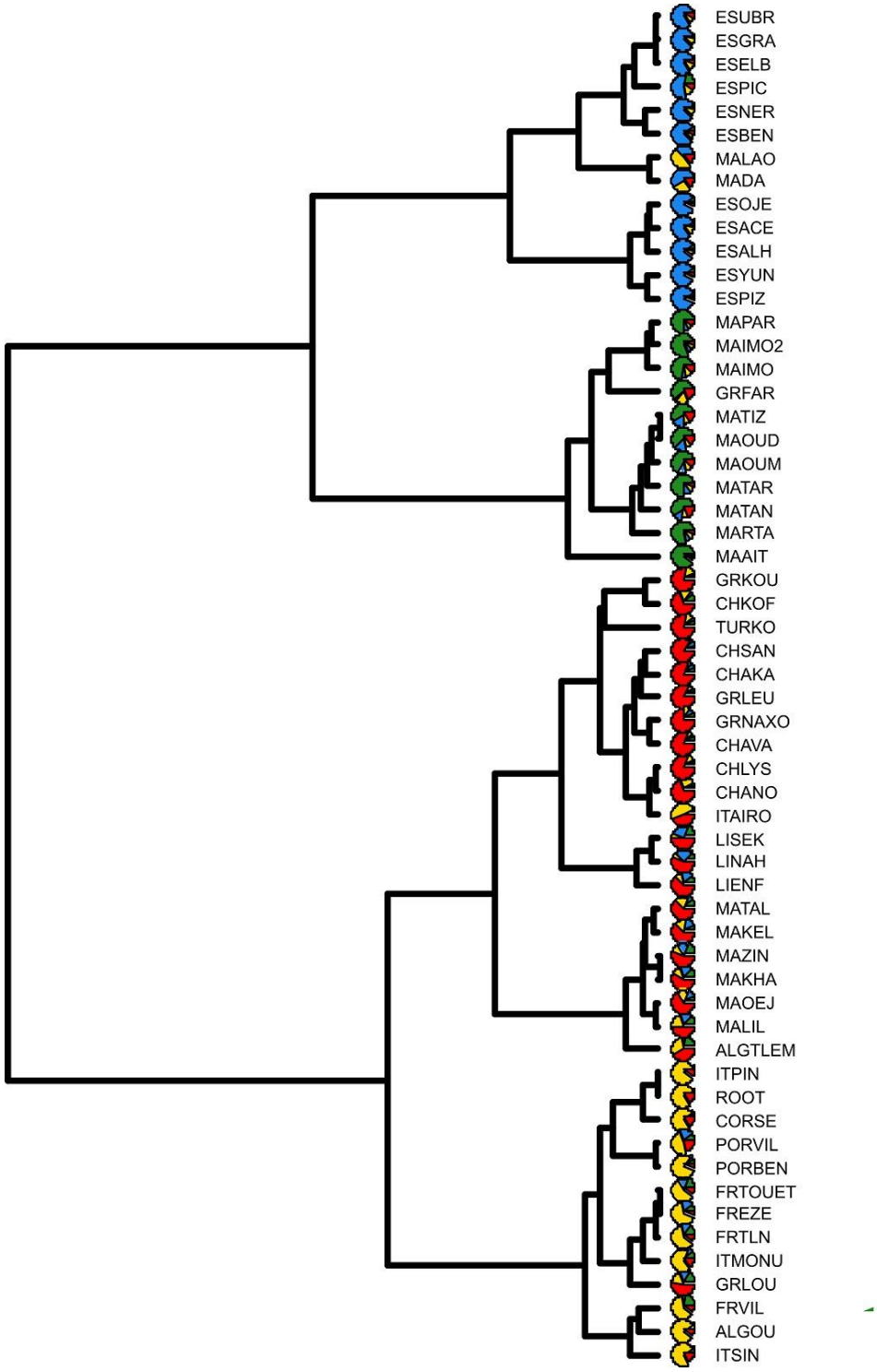


Figure S3.3. Clustering tree obtained with the Ward method and pairwise Jost's D differentiations between populations having at least 10 individuals. Population names are corresponding to locality names of Table S1.1. The pie charts are corresponding to admixture plots obtained with STRUCTURE for K=4 groups.

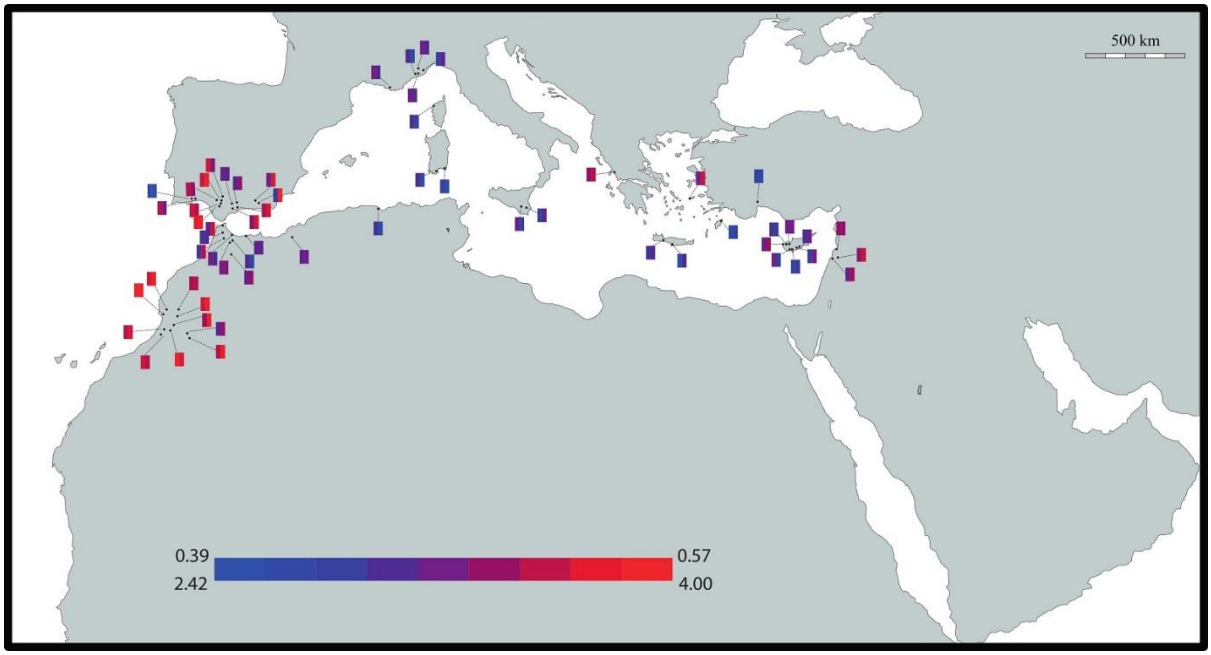


Figure S3.4: Genetic diversity scores per population of *Ceratonia siliqua* genotyped for 17 microsatellite loci, allelic richness on the left (from 2.42 to 4) and expected heterozygosity on the right.

Table S3.2: Parameters inferred by the ABC neuralnet method for TRH model.

Parameter	Prior Range	Distribution	Meaning	Mode and 95% HPD		
				2.5%	**Mode**	**97.5%**
T1	1-10,000	uniform	Divergence between SM and SS	170	695	1,850
T2	1-10,000	uniform	Divergence between EM and CM	53	380	1,186
Tanc	1-10,000	uniform	Ancestral divergence	400	1,815	4,640
NSM	10-10,000	uniform	Current effective size SM	4,295	9,350	9,985
NSS	10-10,000	uniform	Current effective size SS	2,460	5,840	9,990
NCM	10-10,000	uniform	Current effective size CM	3,460	6,910	9,985
NEM	10-10,000	uniform	Current effective size EM	3,145	7,245	9,990
Nanc	10-10,000	uniform	Ancestral effective size	15	910	7,510
mutssr	0.0001-0.0008	loguniform	Mutation rate per locus	1.0 E-04	1.9 E-04	3.0 E-04
p	0.1-0.3	loguniform	Parameter for the GSM model	1.1 E-01	1.5 E-01	2.9 E-01

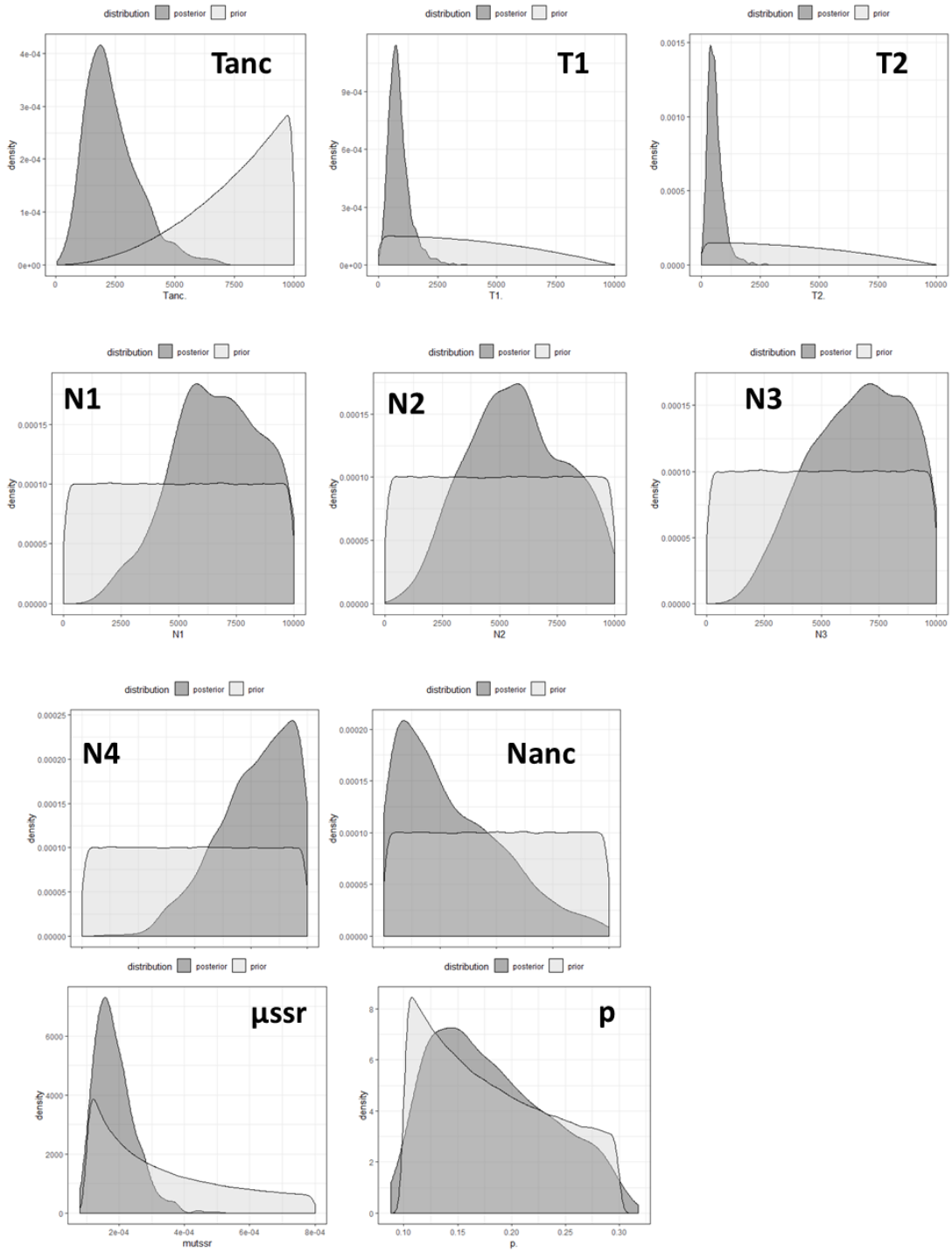


Figure S3.5: Parameter distributions obtained from model TRH. Prior and posterior in light and dark grey respectively.

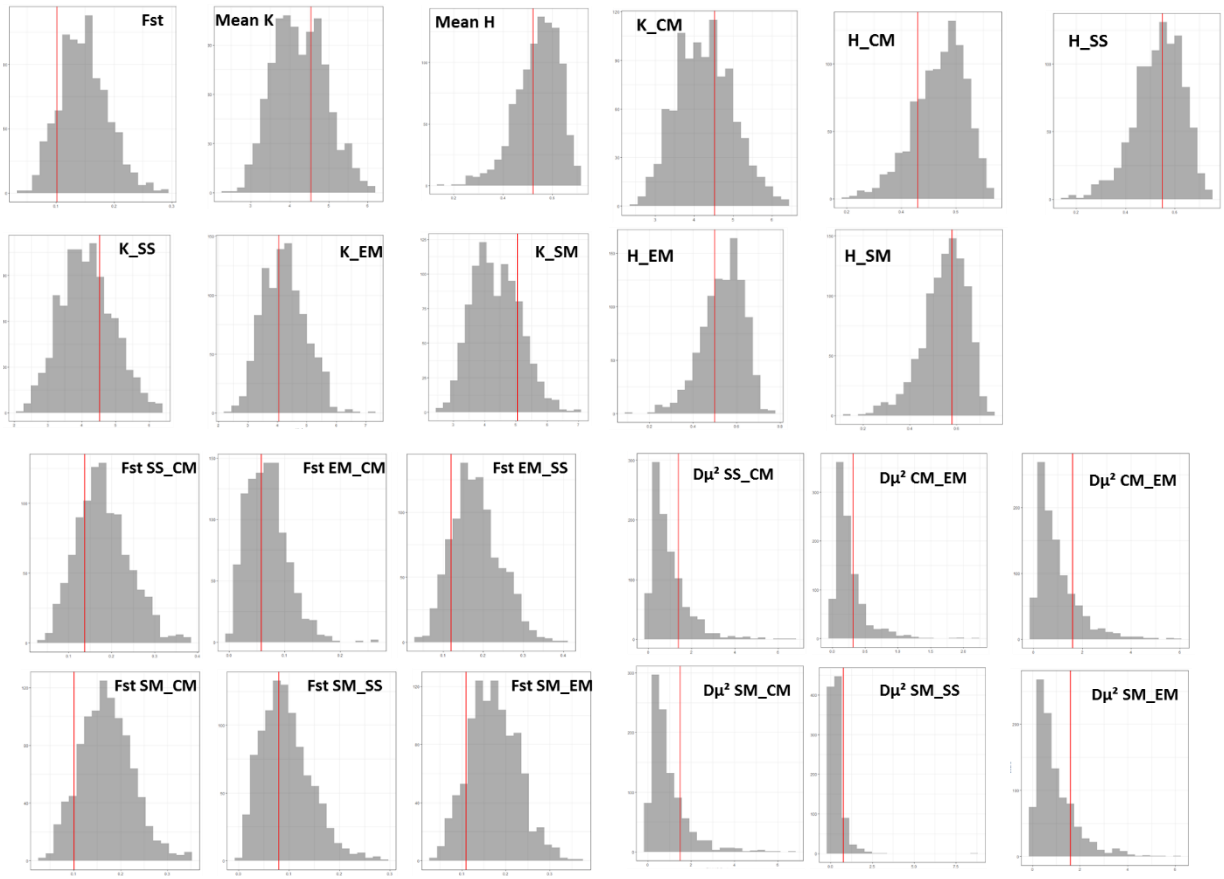


Figure S3.6: Posterior predictive simulations of the TRH model. Histograms show the distribution of 1,000 simulated values and the red vertical lines the observed value.

B) Environmental niche modelling

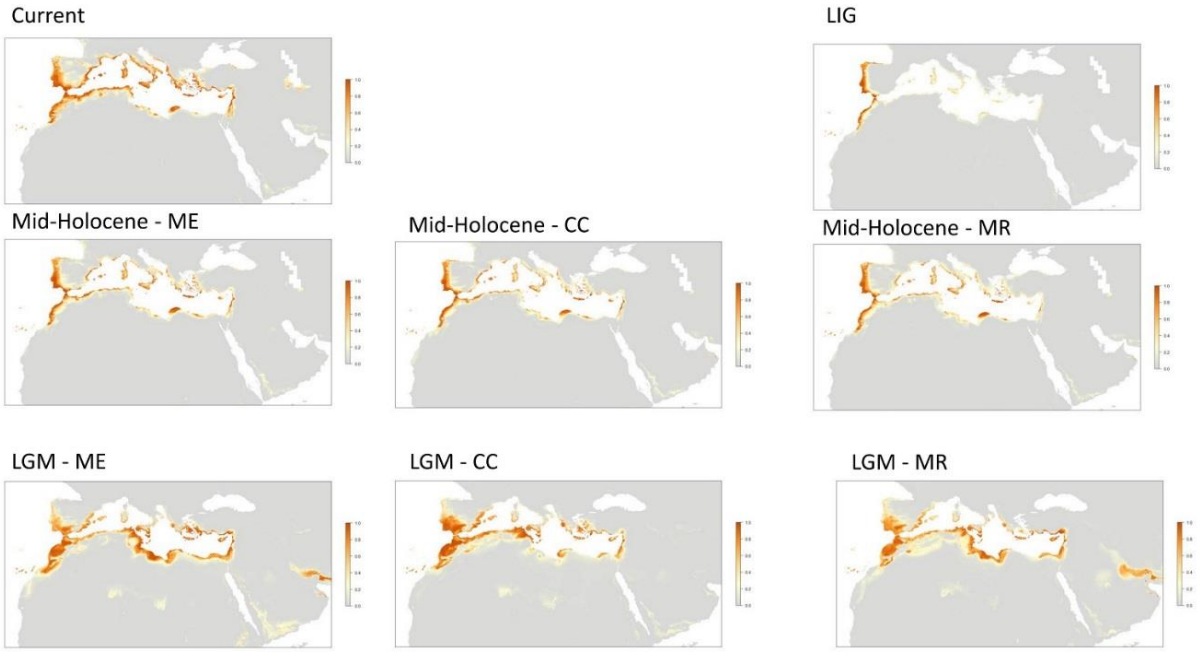


Figure S3.7: Species distribution modelling of the carob in four periods, for Mid-Holocene and LGM the results are presented for to three Global Circulation Models (GCMs); CCSM4 (CC), MIROC-ESM (MR), and MPI-ESM-P (ME).