



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^{1,2} | Nicolas Le Galliot¹ | Samuel Pironon² | Gonzalo Nieto Feliner³ | Jean-Pierre Suc⁴ | Fatma Lakhal-Mirleau¹ | Marianick Juin¹ | Marjorie Selva¹ | Magda Bou Dagher Kharrat⁵ | Lahcen Ouahmane⁶ | Stefano La Malfa⁷ | Katia Diadema⁸ | Hervé Sanguin^{9,10} | Frédéric Médail¹ | Alex Baumel¹

¹Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), Aix Marseille Univ, Avignon Université, CNRS, IRD, Faculté des Sciences et Techniques St-Jérôme, Marseille, France

²Royal Botanic Gardens, Richmond, Kew, UK

³Real Jardín Botánico (CSIC), Madrid, Spain

⁴Institut des Sciences de la Terre Paris, IStEP, UMR 7193, Sorbonne Université, CNRS-INSU, Paris, France

⁵Laboratoire Biodiversité et Génomique Fonctionnelle, Faculté des Sciences, Université Saint-Joseph, Beirut, Lebanon

⁶Laboratoire d'Ecologie et Environnement, Faculté des Sciences Semlalia, Université Cadi Ayyad, Marrakech, Morocco

⁷Dipartimento di Agricoltura, Alimentazione e Ambiente (Di3A), Università degli Studi di Catania, Catania, Italy

⁸Conservatoire Botanique National Méditerranéen de Porquerolles (CBNMed), Hyères, France

⁹BGPI, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

¹⁰LSTM, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France

Correspondence

Alex Baumel, 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

Abstract

Aim: Phylogeography of fruit trees is challenging due to 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 Mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial (LIG) using species distribution modelling.

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 continuous presence since the Oligocene, *Ceratonia* was rarely found in the fossil record during the Pleistocene but present in the western and the eastern Mediterranean. Plastid and nuclear markers, characterized by low allelic richness, revealed a strong west-east genetic structuring. 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.

Marseille Cedex 20, France.
Email: alex.baumel@imbe.fr

Funding information

Marie Skłodowska-Curie Individual Fellowship, Grant/Award Number: 704464 - YAMNOMICS - MSCA-IF-EF-ST; Agence Nationale de la Recherche, Grant/Award Number: ANR-14-CE02-0016

Handling Editor: Hans-Peter Comes

KEYWORDS

Approximate Bayesian computation (ABC), coalescence, fossil, origin, palaeobotany, phylogeography, species distribution modelling, wild genetic resources

1 | INTRODUCTION

Phylogeography and palaeobotany have greatly contributed to our knowledge on species evolutionary histories and their ecological responses to environmental changes (Hu, Hampe, & Petit, 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 (Migliore, Baumel, Leriche, Juin, & Médail, 2018; Nieto-Feliner, 2014). 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 c. 6,000 years ago in the eastern Mediterranean and entailed recurrent exchanges between crops and their wild relatives (Meyer, Duval, & Jensen, 2012; Zohary & Hopf, 2012).

Among the Mediterranean fruit trees, the carob (*Ceratonia siliqua* L., Leguminosae) is an evergreen dioecious tree characteristic of thermophilous vegetation (Baumel et al., 2018; Quézel & Médail, 2003; Zohary, 2002). 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 (Baumel et al., 2018; Ramon-Laca & Mabberley, 2004) 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, *Ceratonia oreoethauma* Hillc., G.P.Lewis & Verdc. (Hillcoat, Lewis, & Verdcourt, 1980), suggests a centre 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; Servera-Vives et al., 2018; Zapata et al., 2013). 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; Ramon-Laca & Mabberley, 2004; Zohary, 2002). 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 c. 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. oreoethauma*. 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 analysed 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).

2 | MATERIALS AND METHODS

2.1 | Plant material

Leaves of 1,067 carob trees from 78 populations were sampled across the Mediterranean between 2015 and 2017. Localities were selected to cover the geographic and ecological ranges of *C. siliqua* as described in Baumel et al. (2018; Table S1.1 in Appendix S1). 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 Appendix S1). *Ceratonia oreoethauma* was sampled from Edinburgh herbarium and Oman botanic garden materials. DNA extraction method is described in Appendix S2.

2.2 | 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 (*matK*) region (see Appendix S2). Divergence times were estimated using a Bayesian relaxed-clock approach implemented in BEAST 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, Herendeen, and Wojciechowski (2005) were applied to deeper nodes of the tree using normal prior distributions: (a) 54 ± 3.4 Ma assigned to the Umtiza stem node, (b) 58.6 ± 0.25 Ma applied to the papilionoid stem node and (c) 59 ± 0.2 Ma set to the legume stem node.

2.3 | Plastid DNA haplotype variation within *C. siliqua*

Four plastid markers (*matK* gene, *psbD-trnT* spacer, *rpl32-trnL* spacer and *ccSA-ndhD* spacer) were sequenced for 424 *C. siliqua* and 10 *C. oreoethauma* individuals. A haplotype network was constructed with the concatenated dataset using TCS (Clement, Posada, & Crandall, 2000). Nucleotide diversity (π) and Tajima's *D* (*D*) were computed with DNASP (Rozas et al., 2017).

2.4 | 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 paralogous 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 1,037 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), Shannon index, observed (H_O) and unbiased expected heterozygosity (H_E), and private allele rate (number of alleles by locus unique to a single genetic group) were computed with GENEAL software (Peakall & Smouse, 2012). A map showing N_A and H_E per population was elaborated to document the geographic patterns of genetic diversity.

STRUCTURE software (Pritchard, Stephens, & Donnelly, 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 geographic 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 in Appendix S3). 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, Mayzel, Jakobsson, Rosenberg, & Mayrose, 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 analysed 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.

2.5 | 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 Section 3). 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 (Migliore et al., 2018; Nieto Feliner, 2014).

All coalescent simulations were done with FASTSIMCOAL 2.6 (Excoffier & Foll, 2011). For model selection, 1×10^5 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 S2. 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 1 million simulations, with the neural network method (R package ABC; Csilléry, François, & Blum, 2012) and a logit transformation of parameter values, considering the closest 1,000 simulations to the observed data (tolerance = 0.1%). The 95% highest posterior densities (HPD) were obtained with the HPDinterval function (R coda package, Plummer, Best, Cowles, & Vines, 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 1,000 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.

2.6 | Environment Niche Modelling

We included 650 occurrence points to model the distribution range of the carob tree using a consensus approach modelling procedure (Marmion, Parviainen, Luoto, Heikkinen, & Thuiller, 2009) based on climatic variables at a 2.5 arc-min resolution from version 1.4 of the Worldclim database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). We projected the species distribution under current, Mid-Holocene (c. 6 ka), Last Glacial Maximum (LGM; c. 22 ka), and Last Interglacial (LIG; c. 130 ka) climatic conditions. More details about carob distribution modelling are given in Appendix S2.

2.7 | 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 in Appendix S2). A map was elaborated using GeoMapApp (Ryan et al., 2009).

3 | RESULTS

3.1 | Divergence time analyses

Divergence time estimates (Figure S3.1 in Appendix S3) support a Late Miocene divergence between *C. siliqua* and its sister species *C. oreoethauma* 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).

3.2 | Geographic 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 1,472 bp. The presence of three substitutions and five indels identified six haplotypes in *C. siliqua*. Nucleotide diversity π (0.00106), number of segregating sites (3) and Tajima's D (3.34, $p < 0.01$) indicated low nucleotide variation and a lack of singletons. Three haplotypes were identified in *C. oreoethauma* due to the existence of two indels. The haplotypes of *C. siliqua* and *C. oreoethauma* differ by fifteen substitutions and four indels.

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. oreoethauma* 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 South Morocco, in which the samples collected north of Agadir

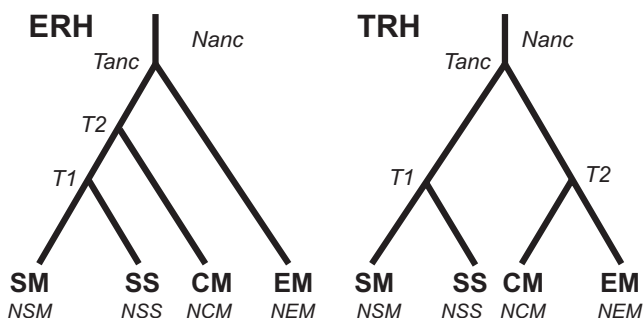


FIGURE 1 Candidate models for the phylogeography of *Ceratonia siliqua*. ERH is the eastern refugium hypothesis and TRH, the two refugia hypothesis. SM, SS, CM and EM correspond to the South Morocco, South Spain, central Mediterranean and eastern Mediterranean genetic groups, respectively

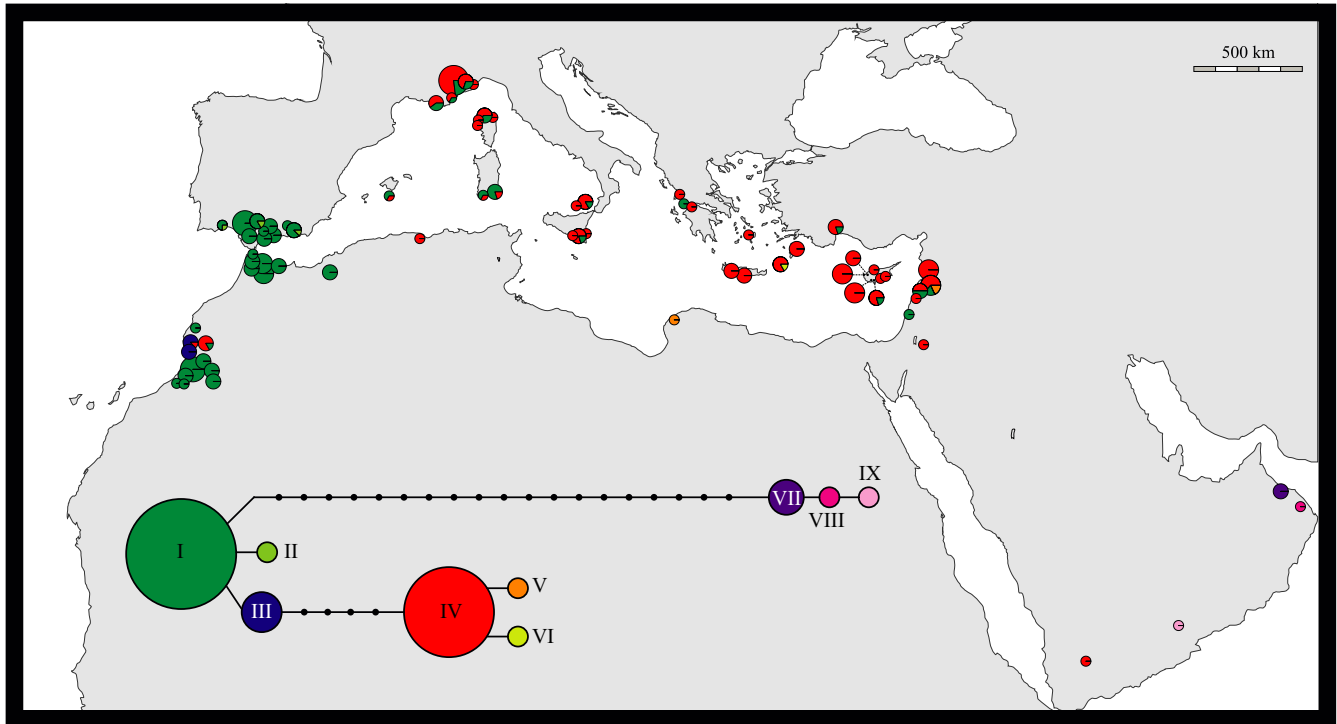


FIGURE 2 Geographic distribution of the plastid haplotypes identified in *Ceratonia siliqua* (I–VI) and *Ceratonia oreothauma* (VII–IX). In the haplotype network, the size of the circles is representative of sample size. Four markers were sequenced to obtain these haplotypes for a total of 1,472 bp

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 7 samples whereas haplotypes VIII and IX appeared in one and two samples, respectively.

3.3 | 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, 1,033 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 Figure S3.2 in Appendix S3. According to STRUCTURE HARVESTER results (Table S3.1 in Appendix S3), 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 in Appendix S3) 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 in Appendix S3) 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, South Morocco; SS, South Spain; CM, central Mediterranean including genotypes from Portugal, Algeria, France, Sardinia, Sicily and Balearics; EM, eastern Mediterranean (i.e. Greece, Cyprus and Lebanon).

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 in Appendix S1). 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.

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 in Appendix S3).

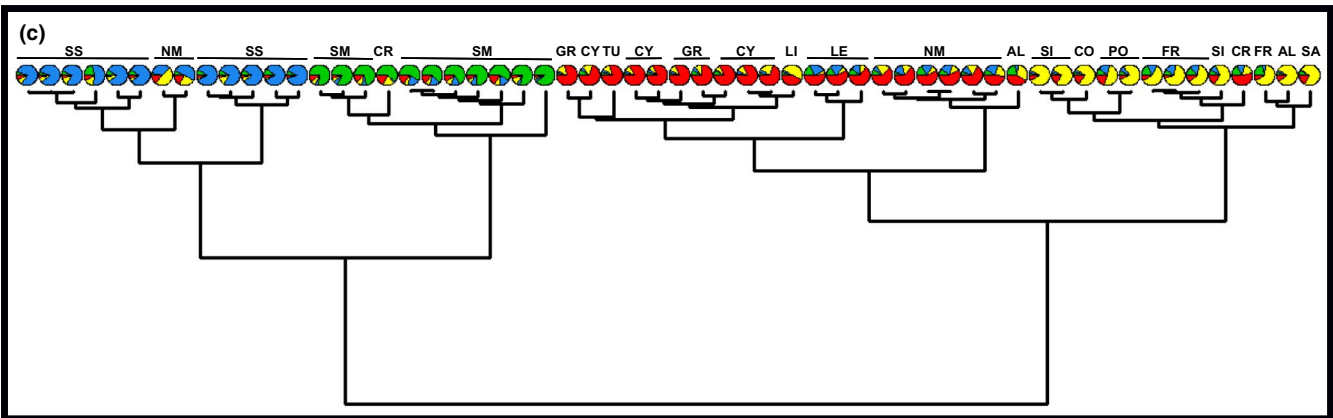
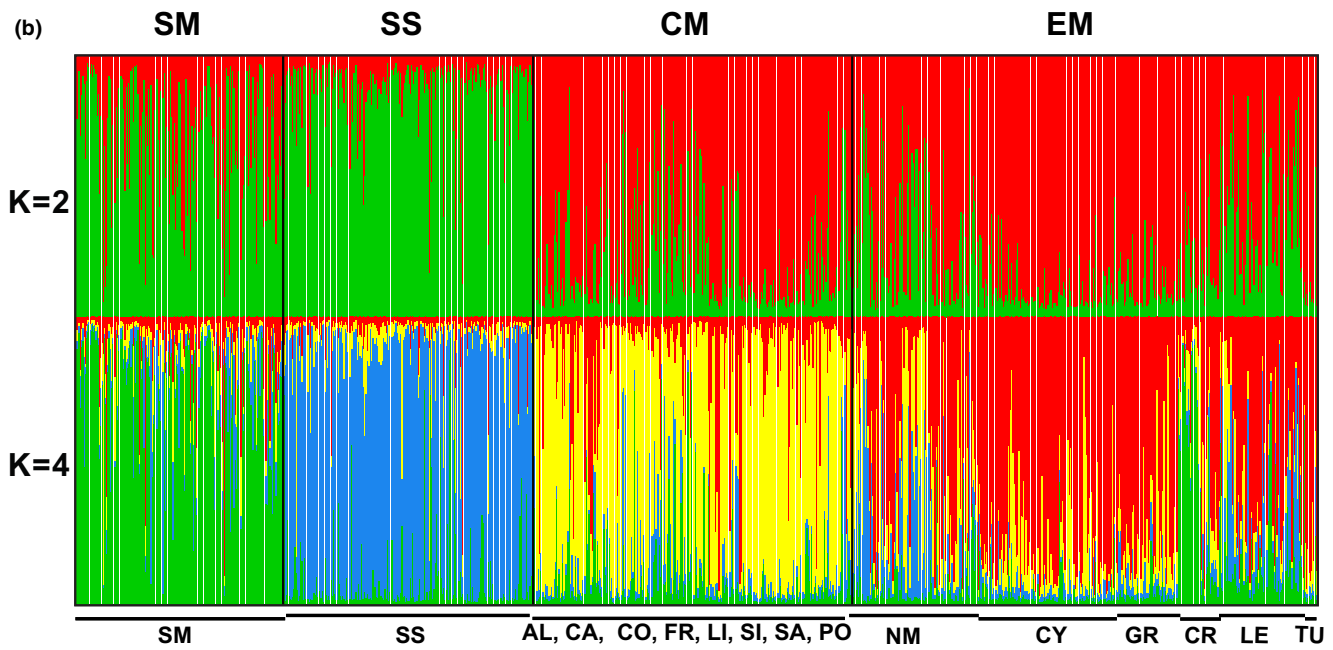
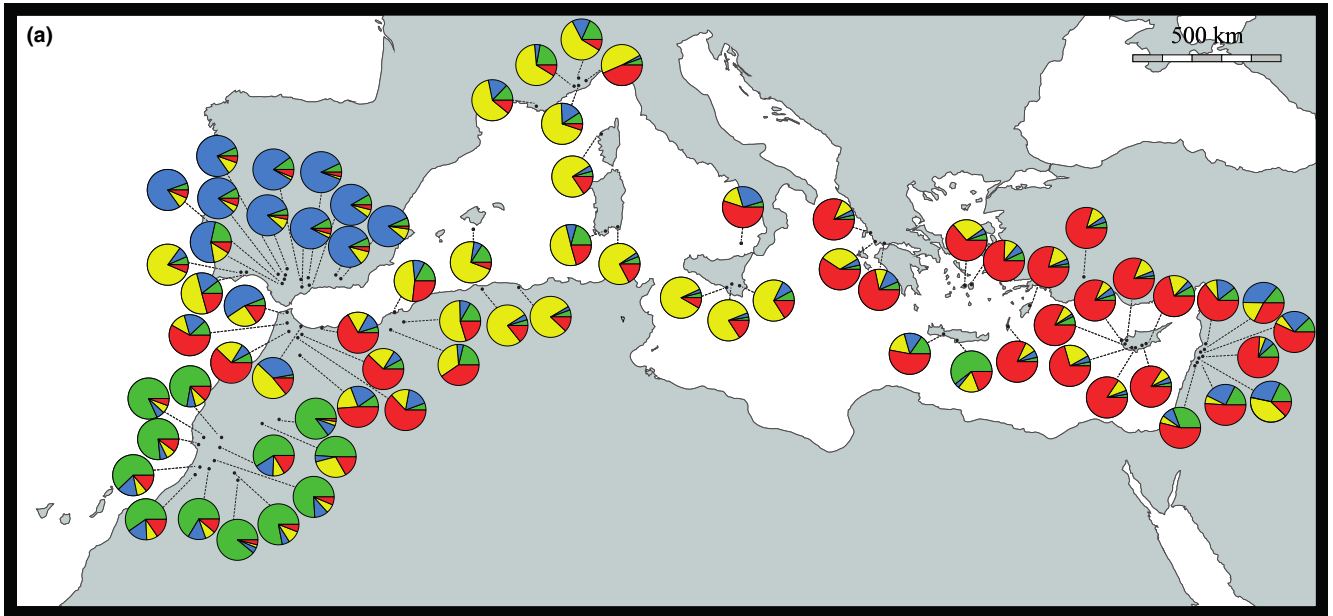


FIGURE 3 Genetic groups inferred for carob tree using 17 microsatellite markers. SM, South Morocco; SS South Spain; CM, central Mediterranean; EM, eastern Mediterranean. (a) Map of the membership frequencies of 1,037 carob genotypes from 67 populations according to the $K = 4$ clustering solution of STRUCTURE. (b) Admixture proportions for $K = 2$ and $K = 4$ obtained from 10 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 10 individuals (the tree with population names is shown in Figure S3.3 in Appendix S3). The pie charts correspond 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

3.4 | Phylogeographic inferences by ABC

The model selected by the random forest in the 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 1,000 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 in Appendix S3. The mode was used as a point estimate for all parameters after graphical inspection of posterior distributions (Figure S3.5 in Appendix S3). Comparisons of distribution of simulated summary statistics (1,000 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 in Appendix S3).

ABC analyses estimated 1,815 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 in Appendix S3 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 3–76 ka). These nodes ages, despite their large 95% HPD, have well peaked distribution modes (Figure S3.5 in Appendix S3) supporting that the main genetic groups of the carob tree diverged well before the onset of crop tree domestication in the Mediterranean.

3.5 | Distribution hindcasting and past occurrences of the carob

Distribution modelling for past periods infers a potential contraction towards the south during the 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 (Figure S3.7 in Appendix S3) although only the results from the CCSM4 model are shown in Figure 5. The strongest change is inferred during the LIG at c. 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.

Forty-four references allowed compiling 55 past occurrences of *Ceratonia* mostly based on pollen records (Table S2.5 in Appendix S2). 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 c. 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 c. 11 ka, localities 17 and 15), indicate the probable presence of the carob tree before domestication in the western Mediterranean.

4 | 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

TABLE 1 Genetic diversity for the four main genetic groups of *Ceratonia siliqua*

Genetic group	N	MLG	N_a	$N_a (\geq 5\%)$	I	H_o	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	1,067	1,037	6.2	3.8	1.05	0.47	0.55	0.14	

N is the number of trees screened, MLG the number of multilocus genotype obtained, N_a is the allelic richness, $N_a (\geq 5\%)$ is the allelic richness restricted to allele having a minimum frequency of 5%, I is the Shannon index, H_o is the observed heterozygosity, H_E the unbiased expected heterozygosity, F is the fixation index and *Priv* is the private allelic rate.

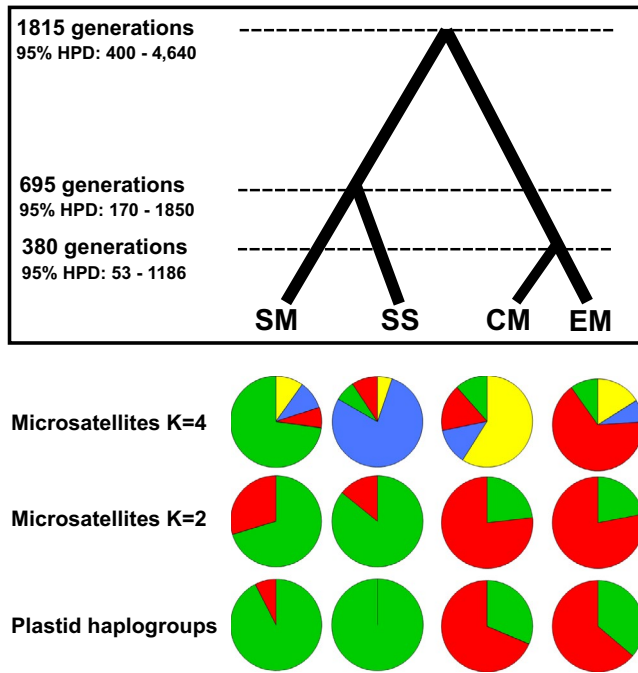


FIGURE 4 The best fitting model inferred by ABC and admixture proportions from microsatellite data (STRUCTURE models for $K = 2$ and $K = 4$, full data set of 1,037 carob genotypes from 67 provenances) and plastid markers data. SM = South Morocco; SS = South Spain; CM = central Mediterranean; EM = eastern Mediterranean. Plastid haplogroups, in green haplotypes I, II and III, in red haplotypes IV, V and VI

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. oreo-thauma* 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. oreo-thauma*, by the number of apertures and surface ornamentation (Figure S2.3 in Appendix S2). The geographical 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, Guinet, & Suc, 1984), corresponding to the pre-Mediterranean sclerophyllous vegetation (Axelrod, 1975; Jiménez-Moreno & 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 (Mijarra, Barrón, Manzanque, & Morla, 2009; Suc, 1984; 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., *O. europaea* or *Myrtus communis* L. (Besnard, Terral, & Cornille, 2017; Désamoré et al., 2011; Migliore et al., 2018); but are congruent with the low plastid variation observed in *Laurus nobilis* L. (Rodríguez-Sánchez, Guzmán, Valido, Vargas, & Arroyo, 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 is 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 South Morocco and South Spain form a distinct cluster and exhibit higher genetic diversity (Figure S3.4 in Appendix S3). 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 (Besnard et al., 2017; Breton, Tersac, & Bervillé, 2006).

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, Gonzalez-Rodriguez, Pahlich, Koehler, & Deacon, 2011; Tsuda, Nakao, Ide, & Tsumura, 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 c. 3 ka (Meyer et al., 2012),

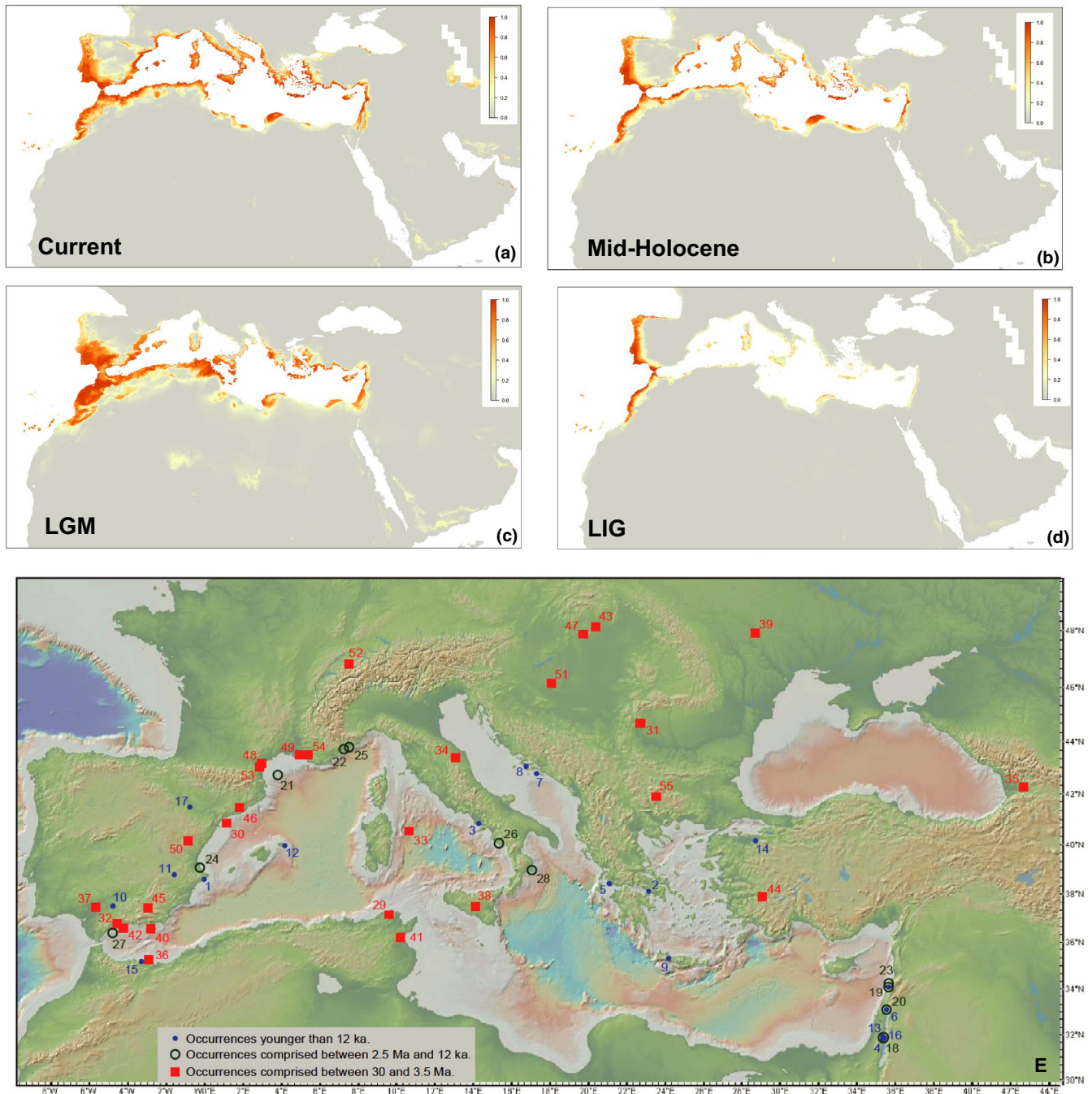


FIGURE 5 Species distribution modelling of the carob in four periods (A, Present; B, Mid-Holocene; C, last glacial maximum (LGM); D, last inter-glacial (LIG)), and E, Map of palaeobotanical data of *Ceratonia* from Oligocene to Middle Ages (macroremains and pollen grains). Numbers correspond to the input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

as well as the first tree crop domestication records c. 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 (Figure 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, Fernández de Castro, Fuertes-Aguilar, & Nieto Feliner, 2018).

The genetic diversity of the carob tree is also characterized by introgression events between the main four lineages. Carob trees from North 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, Ruedi, Petit, and Excoffier (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 North 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.

5 | CONCLUSION

Our multidisciplinary study based on phylogeography, palaeobotany and SDM 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 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), Philippe Ponel (IMBE), Omar Boudouma (ISTeP), Maria Panitsa (Patras University), and the services of Junta de Andalucía for providing information or their help to complete our sampling.

DATA AVAILABILITY STATEMENT

Sampling information, extended laboratory and statistical methods, paleobotanic data as well as extended results on phylogenetic divergence time analysis, genetic diversity structure, ABC inferences and SDM are available in supporting information (32 pages). 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; microsatellite genotypes; STRUCTURE results and occurrence points used for SDM.

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

DOI: doi:10.5061/dryad.k7m020r

Journal: Journal of Biogeography

Journal manuscript number: JBI-19-0177

ORCID

Juan Viruel  <https://orcid.org/0000-0001-5658-8411>

Nicolas Le Galliot  <https://orcid.org/0000-0001-7467-7768>

Samuel Pironon  <https://orcid.org/0000-0002-8937-7626>

Gonzalo Nieto Feliner  <https://orcid.org/0000-0002-7469-4733>

Jean-Pierre Suc  <https://orcid.org/0000-0002-5207-8622>

Magda Bou Dagher Kharat  <https://orcid.org/0000-0001-7969-1673>

Stefano La Malfa  <https://orcid.org/0000-0001-5869-0348>

Katia Diadema  <https://orcid.org/0000-0002-3682-8745>

Hervé Sanguin  <https://orcid.org/0000-0001-7160-2840>

Frédéric Médail  <https://orcid.org/0000-0002-1429-6661>

Alex Baumel  <https://orcid.org/0000-0003-4245-197X>

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. <https://doi.org/10.1038/s41437-018-0123-9>



- Axelrod, D. I. (1975). Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden*, 62, 280–334. <https://doi.org/10.2307/2395199>
- Baumel, A., Mirleau, P., Viruel, J., Bou Dagher 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. <https://doi.org/10.5091/plecevo.2018.1423>
- 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. <https://doi.org/10.1093/aob/mcx145>
- Bessedik, M., Guinet, P., & Suc, J.-P. (1984). Données paléofloristiques en Méditerranée nord-occidentale depuis l'Aquitainien. *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. <https://doi.org/10.1111/j.1365-2699.2006.01544.x>
- 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. <https://doi.org/10.1111/j.1365-2699.2010.02451.x>
- Clement, M., Posada, D., & Crandall, K. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1660. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- 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. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>
- Curat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, 62, 1908–1920. <https://doi.org/10.1111/j.1558-5646.2008.00413.x>
- de Candolle, A. (1883). *Origine des plantes cultivées*. Paris, France: Germer Baillière et Cie.
- 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. <https://doi.org/10.1111/j.1752-4571.2011.00223.x>
- 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. <https://doi.org/10.1111/j.1365-2699.2010.02387.x>
- Diez, C. M., Trujillo, I., Martínez-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. <https://doi.org/10.1111/nph.13181>
- Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214. <https://doi.org/10.1186/1471-2148-7-214>
- 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. <https://doi.org/10.1007/s12686-011-9548-7>
- Excoffier, L., & Foll, M. (2011). Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27, 1332–1334. <https://doi.org/10.1093/bioinformatics/btr124>
- 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.
- 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. <https://doi.org/10.1038/nature02546>
- 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. <https://doi.org/10.1111/nph.12929>
- 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. <https://doi.org/10.1086/285420>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hillcoat, D., Lewis, G., & Verdcourt, B. (1980). A New Species of *Ceratonia* (Leguminosae-Caesalpinioideae) from Arabia and the Somali Republic. *Kew Bulletin*, 35, 261. <https://doi.org/10.2307/4114570>
- Hu, F. S., Hampe, A., & Petit, R. J. (2009). Paleoeecology 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. <https://doi.org/10.1016/j.palaeo.2007.03.040>
- Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, 40, 593–612. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095702>
- 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.
- 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. <https://doi.org/10.1016/j.bse.2014.03.022>
- 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. <https://doi.org/10.1080/10635150590947131>
- 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. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- Mateu-Andrés, I., Cuirana, M. J., Aguilera, 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. <https://doi.org/10.1111/j.1469-8137.2012.04253.x>
- 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. <https://doi.org/10.1093/botlinnean/boy032>
- Mijarra, J. M. P., Barrón, E., Manzaneque, F. G., & Morla, C. (2009). Floristic changes in the Iberian Peninsula and Balearic Islands (south-west Europe) during the Cenozoic. *Journal of Biogeography*, 36, 2025–2043. <https://doi.org/10.1111/j.1365-2699.2009.02142.x>

- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, 31, 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- 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. <https://doi.org/10.1016/j.ppees.2014.07.002>
- Palamarev, E. (1989). Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendroflora. *Plant Systematics and Evolution*, 162, 93–107. <https://doi.org/10.1007/BF00936912>
- Peakall, R., & Smouse, P. E. (2012). GenAIEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics*, 28, 2537–2539.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- 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. <https://doi.org/10.1093/bioinformatics/btv684>
- Quézel, P., & Médail, F. (2003). *Écologie et biogéographie des forêts du bassin méditerranéen*. Paris, France: Elsevier Editions.
- Ramón-Laca, L., & Mabblerley, 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. <https://doi.org/10.1111/j.1095-8339.2003.00254.x>
- 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. <https://doi.org/10.1093/molbev/msx248>
- Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., ... Zerny, R. (2009). Global multi-resolution topography synthesis. *Geochemistry Geophysics Geosystems*, 10, Q03014. <https://doi.org/10.1029/2008GC002332>
- Servera-Vives, G., Riera, S., Picornell-Gelabert, L., Moffa-Sánchez, P., Llergo, Y., García, A., ... Calvo Trias, M. (2018). The onset of islandscapes in the Balearic Islands: A study-case of *Addaia* (northern Minorca, Spain). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 498, 9–23. <https://doi.org/10.1016/j.palaeo.2018.02.015>
- Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, 307, 429. <https://doi.org/10.1038/307429a0>
- Suc, J.-P., Popescu, S.-M., Fauquette, S., Bessedik, M., Jiménez-Moreno, G., Bachiri Taoufiq, N., ... 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. <https://doi.org/10.1093/botlinnean/boy025>
- Viruel, J., Haguenaue, A., Juin, M., Mirleau, F., Bouteiller, D., Boudagher-Kharrat, 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. <https://doi.org/10.1177/0959683613486944>
- 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, 141–215. <https://doi.org/10.1560/BW6B-4M9P-U2UA-C6NN>
- 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, UK: 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; J.V. and A.B. developed the method to sequence microsatellites; F.L.M., M.J., M.S., and J.V. conducted the laboratory work; A.B., N.L.G. and J.V. analysed the genetic data; S.P. and J.V. performed the SDM analyses; J.P.S. checked the paleobotanic data thoroughly; 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Viruel J, Le Galliot N, Pironon S, et al.

A strong east–west Mediterranean divergence supports a new phylogeographic history of the carob tree (*Ceratonia siliqua*, Leguminosae) and multiple domestications from native populations. *J Biogeogr.* 2019;00:1–12. <https://doi.org/10.1111/jbi.13726>