

Genomic, spatial and morphometric data for discrimination of four species in the Mediterranean *Tamus* clade of yams (*Dioscorea*, Dioscoreaceae)

Miguel Campos¹, Emma Kelley¹, Barbara GRAVENDEEL¹, Frédéric Médail¹, J. M. Maarten Christenhusz¹, Michael F. Fay¹, Pilar Catalán¹, Ilia J. Leitch¹, Félix Forest¹, Paul Wilkin¹, and Juan Viruel¹

¹Affiliation not available

October 24, 2022

Abstract

Background and aims. Among the numerous pantropical species of the yam genus, *Dioscorea*, only a small group occurs in the Mediterranean basin, including two narrow Pyrenean endemics (Borderea clade), and two Mediterranean-wide species (*D. communis* and *D. orientalis*, *Tamus* clade). However, several currently unrecognized species and infraspecific taxa have been described in the *Tamus* clade due to significant morphological variation associated with *D. communis*. Our overarching aim was to investigate taxon delimitation in the *Tamus* clade using an integrative approach combining phylogenomic, spatial and morphological data.

Methods. We analysed 76 herbarium samples using Hyb-Seq genomic capture to sequence 260 low-copy nuclear genes and plastomes, together with morphometric and environmental modelling approaches.

Key results. Phylogenomic reconstructions confirmed that the two previously accepted species of the *Tamus* clade, *D. communis* and *D. orientalis*, are monophyletic and form sister clades. Three subclades showing distinctive geographic patterns were identified within *D. communis*. These subclades were also identifiable from morphometric and climatic data, and introgression patterns were inferred between subclades in the eastern part of the distribution of *D. communis*.

Conclusions. We propose a taxonomy that maintains *D. orientalis*, endemic to the eastern Mediterranean region, and splits *D. communis sensu lato* into three species: *D. edulis*, endemic to Macaronesia (Canary Islands and Madeira); *D. cretica*, endemic to the eastern Mediterranean region; and *D. communis sensu stricto*, widespread across western and central Europe. Introgression inferred between *D. communis s.s.* and *D. cretica* is likely to be explained by their relatively recent speciation at the end of the Miocene, disjunct isolation in eastern and western Mediterranean glacial refugia and a subsequent westward recolonization of *D. communis s.s.* Our study shows that the use of integrated genomic, spatial and morphological approaches allows a more robust definition of species boundaries and the identification of species that previous systematic studies failed to uncover.

Introduction

The yam genus, *Dioscorea* L. (Dioscoreaceae) is a diverse group currently containing 631 accepted species (POWO, 2022) possessing underground storage organs and, in most, a climbing habit. Species with starchy tubers constitute a food staple for millions of people, resulting in seven to ten species being cultivated on a large scale (Asiedu and Sartie, 2010), including two (*D. alata* L. and *D. cayenensis* Lam.) which together are the most widely cultivated crops (Price *et al.* , 2016). More than 40 wild species are harvested as food sources (Martin and Degras, 1978). In addition, some yams have been used in traditional medicine and as a source of steroidal precursors (De Luca *et al.*, 2012; Hua *et al.*, 2017; Price *et al.* , 2016). While most wild yam species are found in tropical regions (Caddick *et al.*, 2002), a few species are distributed in

temperate regions and exhibit unique morphological traits (Viruel *et al.*, 2010). For example, only six species occur in the Mediterranean-Macaronesian region: two species of the Stenophora clade (*D. balcanica* Košanin native to Montenegro and Albania, and *D. caucasica* Lipsky, found in Georgia and Caucasian Russia), the Borderea clade, which contains two well-defined and narrow endemic species from the Pyrenean mountains (*D. chouardii* Gaussen and *D. pyrenaica* Bubani & Bordère ex Gren.), and the Tamus clade, which is defined by having berries rather than winged capsules and is more widely distributed across the Mediterranean Basin, Macaronesia and Atlantic Europe (Viruel *et al.*, 2016).

The Tamus clade currently comprises two species (Wilkin *et al.*, 2005): *D. communis* (L.) Caddick & Wilkin, distributed throughout the Mediterranean Basin and the Macaronesian Islands (Canary Islands and Madeira), and with infraspecific variation in ploidy (Viruel *et al.*, 2019); and *D. orientalis* (J. Thiébaud) Caddick & Wilkin, restricted to Lebanon and Israel. However, like in many *Dioscorea* clades (Viruel *et al.*, 2010), the Tamus clade has had multiple previous taxonomic circumscriptions. The Tamus clade was considered as a separate genus, *Tamus* L., distinct from *Dioscorea*, until 2002 (Caddick *et al.*, 2002). Linnaeus (1753) recognized two species: *T. communis* L. with cordate leaves and a Mediterranean distribution, and *T. cretica* L. with trilobed leaves and typified with material from the Greek island of Crete. In the 19th and early 20th centuries, four Macaronesian endemic species were described (*T. edulis* Lowe, *T. parviflora* Kunth, *T. norsa* Lowe and *T. canariensis* Willd. ex Kunth), while *T. cirrhosa* Hausskn. ex Bornm., *T. cordifolia* Stokes and *T. racemosa* Gouan were treated as distinct Mediterranean species. In the late 20th Century, *T. cretica* was placed as a subspecies in *T. communis* (*T. communis* subsp. *cretica* (L.) Nyman), and *T. communis* f. *subtriloba* (Guss.) O. Bolòs & Vigo was described as a variety with trilobed leaves found in the Balearic Islands and northeastern Spain (Catalonia). All these names were subsequently united under the currently accepted *D. communis* (Caddick *et al.*, 2002). The second species currently recognized in the Tamus clade, *D. orientalis*, was originally described as *T. orientalis* J. Thiébaud named after its eastern Mediterranean distribution.

From the above, it is clear that species concepts have undergone many changes since Linnaeus described two *Tamus* species using morphology, especially reproductive traits (De Queiroz, 2007). As for many other species in other plant genera and families, integrative taxonomic and systematic approaches combining genetic data, morphometrics and climatic envelope data have successfully helped to delimit species in challenging groups of plants (e.g., Frajman *et al.*, 2019). The emergence of high-throughput sequencing (HTS) techniques and the production of thousands of molecular markers have massively increased our ability to resolve relationships between and within species, and subsequently redefine species boundaries (e.g., Fayet *et al.*, 2019; Escudero *et al.*, 2020). Among these HTS methods, Hyb-Seq has become widely adopted across plant phylogenomic studies due to its ability to generate data from degraded herbarium materials (e.g., Brewer *et al.*, 2019; Viruel *et al.*, 2019) and to resolve relationships at different taxonomic scales (e.g., Villaverde *et al.*, 2018). Hyb-Seq techniques rely on genome skim data and target capture probes either designed specifically for some genera or families (e.g., Soto Gomez *et al.*, 2019) or more widely across larger groups, including all angiosperms (e.g., Johnson *et al.*, 2019). In this study, we use a multidisciplinary approach combining genomic, morphometrics, and environmental niche modelling data generated from herbarium specimens to identify taxon boundaries in the challenging Tamus clade of *Dioscorea*, and to explore their phylogeographic patterns across the Mediterranean.

Material and methods

Plant material

Seventy-six herbarium specimens identified as *D. communis* or *D. orientalis* were used to obtain genomic, spatial, and morphometric data. They were selected as being representative of the macromorphological diversity and geographical distribution ranges of the two species as currently circumscribed (Supplementary Data Table S1). The genomic sampling also included material that were used as outgroup taxa, comprising the two species from the Borderea clade (*D. chouardii* and *D. pyrenaica*), which is sister to the Tamus clade (Viruel *et al.*, 2016), and two members of the more distantly related African clade (*D. elephantipes* (L'Hér.) Engl. and *D. sylvatica* Eckl.).

Phylogenomics

Total genomic DNA was extracted from herbarium specimens using a modified CTAB protocol (Doyle and Doyle, 1987). Nuclear target enrichment was used to capture 260 low- to single-copy nuclear (LSCN) genes using RNA baits design for *Dioscorea* (Soto Gomez *et al.*, 2019). Genomic libraries were prepared using NEBNext® Ultra II DNA Library Prep Kit for Illumina(r) (New England Biolabs, Ipswich, Massachusetts, USA) with AMPure XP magnetic beads and NEBNext(r) Multiplex Oligos for Illumina(r) (Dual Index Primer Sets I and II) as tags for simultaneous sequencing. Subsequently, the enriched libraries were multiplexed and sequenced on a HiSeq X platform (Illumina, Inc.) lane.

We filtered raw paired-end reads by removing adapter sequences and low-quality reads using Trimmomatic v.0.36 (Bolger *et al.*, 2014). We used HybPiper v.1.3.1 (Johnson *et al.*, 2016) to recover 260 nuclear genes and associated introns (Soto Gomez *et al.*, 2019) using sequence data from the transcriptome of *D. communis* (SRA SAMN11290810) as a reference file following Viruel *et al.* (2019). We used nQuire (Weiss *et al.*, 2018) to calculate the number of read counts for each allele per single nucleotide polymorphism (SNP), and estimated the median value of allelic ratios per sample to classify each individual as diploid (<2) or polyploid (>2) as described and optimized for *Dioscorea* in Viruel *et al.* (2019). The percentage of polymorphic sites was calculated as the percentage of SNP positions compared to the total number of base pairs retrieved for each sample. Plastome data were recovered using HybPiper and the plastome of *D. elephantipes* as reference (GenBank NC_009601).

Sequences were aligned with MAFFT v.7 (Katoh, *et al.*, 2002) using the *-auto* parameter, and debugged with trimAl v.1.4.1 (Capella-Gutierrez *et al.*, 2009) using the *-automated1* command. Phylogenomic trees were reconstructed using the concatenated nuclear DNA (nDNA) and plastid DNA (pDNA) datasets independently, and for each nuclear gene independently, using RAxML-NG (Katoh *et al.*, 2019) and IQ-TREE (Nguyen *et al.*, 2015), with a GTR+GAMMA substitution model and 1,000 bootstrap replicates. We used ASTRAL-III (Zhang *et al.*, 2018) to construct a species tree based on the independent nuclear gene trees, and SVDquartets to evaluate 10,000,000 random quartets -or all possible quartets if lower- and 10,000 bootstrap replicates, as implemented in PAUP* 4.0a146 (Swofford, 2002). Haplotype networks were reconstructed with plastid data using the TCS method as implemented in Popart v1.7 (Clement *et al.*, 2002; Leigh and Bryant, 2015).

We used Structure (Pritchard *et al.*, 2000) to further investigate the genetic clusters within and between taxa based on filtered SNP data from the concatenated nDNA dataset. We tested one to six genetic groups ($K = 1-6$) allowing admixture at individual level, and correlated allele frequencies, by running five replicates with a 100,000 burn-in and a chain length of 1,000,000 simulations each. Structure Harvester (Earl and vonHoldt, 2012) was used to obtain likelihood values for the multiple values of K and to apply the delta K criterion to select the optimal K . We plotted Structure results for each K value using StructuRly (Criscuolo and Angelini, 2020).

Divergence times were estimated using a Bayesian relaxed-clock approach implemented in BEAST 1.10.4 (Drummond and Rambaut, 2007) and a penalized likelihood approach as implemented in treePL (Smith and O’Meara, 2012) using the concatenated nDNA dataset containing one representative per taxon. In both analyses, the crown node of the African/Mediterranean clade was used as calibration by applying a minimum age of 24 MY and a maximum age of 40 MY, based on the age estimates from Viruel *et al.* (2016; 95% HPD of 24.3469–39.2223 MY). We applied the GTR+I+G substitution model, Yule tree prior, and an uncorrelated lognormal molecular clock and ran the analysis for 1 billion generations, sampling every 100,000 generations. The treePL analysis was conducted in two consecutive runs: i) applying the “prime” option to select the most optimal parameter values, and ii) a “thorough” analysis by setting *opt* = 1, *optad* = 2 and *optcvad* = 5.

Spatial analysis

Occurrence records were obtained from 287 observations from the Global Biodiversity Information Facility (211 occurrences validated by morphology, <http://www.gbif.org/>) and data from herbarium specimens (76 occurrences). For modelling purposes, the dataset was reduced to keep only georeferenced data.

We used environmental niche modelling (ENM) approaches to reconstruct the potential distribution of the four main clades uncovered in the phylogenomic analyses (see Results) under current and past climatic conditions using the maximum entropy algorithm implemented in the R package ‘Maxent’ (Phillips, 2021). Nineteen bioclimatic variables were extracted from the Bioclim dataset, provided by WorldClim 1.4 in a GIS-based raster format (2.5 min resolution). The correlations between environmental variables were determined with a Pearson’s correlation matrix and subsequent realization of a dendrogram cluster for its visualization (Supplementary Data Figure S1). We selected a different set of uncorrelated variables for each geographical region with a high percentage of contribution (PC): bio4 (temperature seasonality), bio8 (mean temperature of wettest quarter), bio9 (mean temperature of driest quarter), bio15 (precipitation seasonality) and bio16 (precipitation of wettest quarter) for the circum-Mediterranean region; bio3 (isothermality), bio6 (min temperature of coldest month), bio8 (mean temperature of wettest quarter), bio14 (precipitation of driest month), bio15 (precipitation seasonality) and bio16 (precipitation of wettest quarter) for the Eastern Mediterranean region; bio3 (isothermality), bio4 (temperature seasonality), bio16 (precipitation of wettest quarter) and bio18 (precipitation of warmest quarter) for the Macaronesian region. ENM analyses were carried out under current climatic conditions, and projected to climatic conditions of the Mid Holocene (MH, about 6,000 years ago), the Last Glacial Maximum (LGM, about 22,000 years ago; Braconnot *et al.*, 2007) and the Last Interglacial (LIG, ~120,000 - 140,000 years BP; Otto-bliesner *et al.*, 2006) using the palaeoclimatic ‘Community Climate System Model’ (CCSM; Gent *et al.*, 2011). Layers were cropped to represent the distribution range of each phylogenetic group (i.e., DC1, DC2 and DC3 for *D. communis*, and *D. orientalis*; see Results) to maximize the reliability of the results and discard false occurrences using the package ‘raster’ (version 3.5-15; R v.4.0.5). We used the Schoener’s D and Hellinger’s I indices as implemented in ENMtools v.1.0.4 to evaluate niche overlap (Warren *et al.*, 2008, 2010). Equivalence and similarity tests with 1,000 replicates were carried out to assess if the overlap between ENMs is higher than expected under randomized ENMs.

A multivariate ordination analysis (principal component analysis, PCA) was carried out using uncorrelated bioclimatic variables obtained from WorldClim using the packages ‘ade4’, ‘factoextra’, ‘magrittr’, ‘dismo’ and ‘HH’ in R v.4.0.5. The correlation analysis was performed with a Pearson correlation matrix and the subsequent visualization and selection of variables using a dendrogram cluster. The selection of uncorrelated variables with the highest contribution to the PCA were 10: bio1 (annual mean temperature), bio2 (mean diurnal range), bio3 (isothermality), bio7 (temperature annual range), bio8 (mean temperature of wettest quarter), bio9 (mean temperature of driest quarter), bio10 (mean temperature of warmest quarter), bio12 (annual precipitation), bio15 (precipitation seasonality) and bio19 (precipitation of coldest quarter).

Morphometrics

We studied vegetative and reproductive traits of 76 herbarium specimens (60 males and 16 females), previously used to delimit taxa boundaries in other *Dioscorea* species (e.g., Viruel *et al.*, 2010), using a 150 mm calliper, a stereomicroscope and ImageJ 1.52a (Schneider *et al.*, 2012). Traits were measured and treated independently for male and female individuals (Supplementary Data Table S2). Pollen grains were sputter coated with platinum and examined using a Hitachi S4700 cold field emission scanning electron microscope (SEM) at 2kV (Hitachi High-Technologies Corporation, Tokyo, Japan).

A value of >0.7 Pearson correlation was used as a threshold to exclude correlated variables from the analysis. The normality Kolmogorov-Smirnov test with Lilliefors correction and the homoscedasticity Levene’s test were applied to the variables following a normal distribution. Subsequently, comparisons among taxonomic units identified in phylogenomic analyses (see Results) were conducted using Kruskal-Wallis and Bonferroni post-hoc tests. Statistical analyses were performed using ‘nortest’, ‘Hmisc’, ‘corrplot’, ‘PerformanceAnalytics’ and ‘car’ packages in R v.4.0.5 (R Core Team, 2022).

Results

Data recovery and phylogenomic results

Target capture data recovery for 76 samples of the *Tamus* clade of *Dioscorea* and the four outgroup samples

included in this study are summarized in Table 1. Sequence data are available at SRA repository, PRJNA525269. An average of 2,240,287 quality filtered paired-end reads were retrieved per sample, ranging between 43,612 and 16,196,619 reads. While the samples from herbarium specimens dated from 1788 to recently collected material, the differences in number of retrieved reads were not related to the age of the specimens. On average, the proportion of reads on target (enrichment efficiency) was 0.33 (0.09–0.60), and although sequences were assigned on average to 258 genes per sample, assemblies at 50% of the expected size of each gene were retrieved on average for 215 genes per sample.

Our target capture approach allowed us to recover an average of 326,149 bp (45,171–394,977) of nuclear data per sample, which corresponds to a recovery rate of 76.9% (10.6–93.1%), while the off-target reads contained plastid data that permitted the assembly of 131,543 bp on average (30,666–151,239) of the plastome per sample. No differences were observed in recovery rates between the clades reconstructed in our analysis (see below); the overall sequencing and target capture data obtained for the four outgroup samples were in the range of the remaining samples.

Both nuclear- and plastid-based phylogenomic reconstructions support the monophyly of the *Tamus* clade in *Dioscorea* (Figure 1), with two highly supported clades that corresponded to the two currently recognized species (*D. communis* and *D. orientalis*). Three highly supported subclades were reconstructed in the *D. communis* clade in the nuclear tree (Figure 1). A first split separated the samples of *D. communis* from Macaronesia (clade DC1). The remaining samples of *D. communis* fell into two sister subclades corresponding to samples of *D. communis* from the eastern Mediterranean (clade DC2) and Mediterranean and Europe (clade DC3), respectively. The latter clade DC3 was subsequently further subdivided into (eastern Mediterranean, (central Mediterranean, western Europe)) subclades. Overall bootstrap support was >90% for most of the nodes (Figure 1).

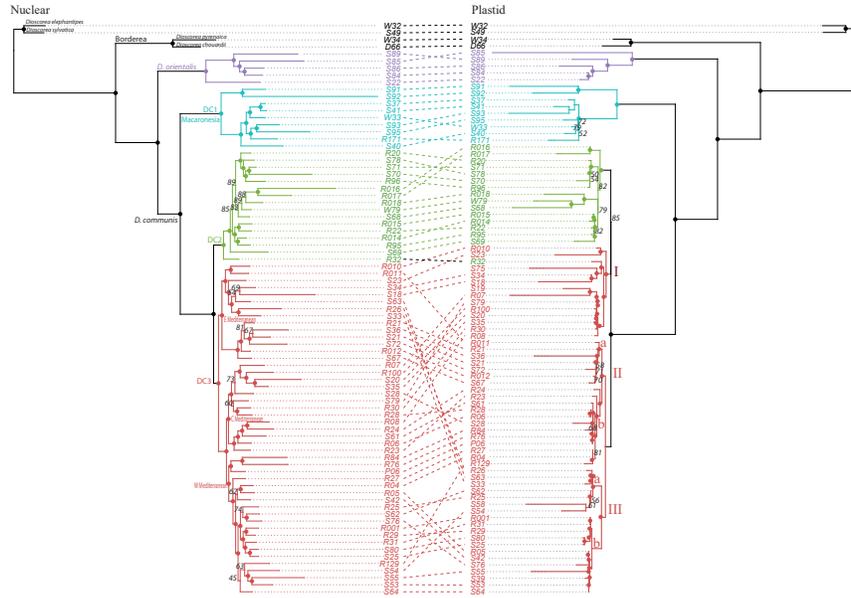


Figure 1: Phylogenomic trees reconstructed using the concatenated nuclear (left) and plastid data (right) obtained using Hyb-Seq for 76 samples of the Tamus clade of *Dioscorea*. Samples included are representative of its distribution range across the Mediterranean region. Filled circles represent branches with support values over 90%, while lower values are shown on branches. Colours represent the main clades found in the nuclear tree, see Results (*D. orientalis* violet, DC1 blue, DC2 green, DC3 red), and roman numbers identifying the subclades in *D. communis* in the plastid tree (I, II, III; subclades within each clade were indicated with lowercase letters a and b). Dashed lines connect the same samples in the nuclear and plastid trees.

The plastid tree was generally congruent with the nuclear tree. However, the plastid topology differed in the resolution of the most recent Mediterranean subclades. In addition, one DC2 sample (R32) was placed in a different subclade (see Figure 1). The remaining Mediterranean *D. communis* samples (DC3) were intermingled in three plastid subclades (I, II and III in Figure 1). In the nuclear tree, the subclade DC3 from the eastern Mediterranean contained samples with plastid haplotypes from all plastid subclades I, II and III, and the subclades of DC3 from central Mediterranean and western Europe contained samples with those from plastid subclades I and II, and II and III, respectively (Figure 1).

To better understand the source of topological incongruence between the nuclear and plastid phylogenetic trees of the Mediterranean samples of *D. communis*, we used STRUCTURE to investigate whether introgression between samples from different subclades may have occurred. The highest deltaK value (15821.9) was obtained for $K = 3$ genetic groups (Supplementary Data Figure S2). All samples in clade DC2 of *D. communis* showed genetic profiles corresponding to one genetic group (Cluster 1), and five samples also showed a percentage of membership $<20\%$ to Cluster 2 (i.e., samples R16, R17, S69 and S70) or Cluster 3 (sample R32, 26%). The samples of *D. communis* in clade DC3 showed genetic profiles corresponding predominantly to Cluster 3 and multiple patterns of admixture with Clusters 1 and 2 (Figure 2). The genetic profile of one DC3 sample from the eastern Mediterranean subclade showed a percentage of membership $>20\%$ to Cluster 1, and nine DC3 samples had membership percentages of $>20\%$ to Cluster 2 (five from the eastern Mediterranean, one from the central Mediterranean and three from the western Europe subclades). Other alternative K groups did not increase the number of clusters in DC2 and showed higher admixture in the DC3 subclades (Supplementary Data Figure S3). In all cases, the sample R32 was always resolved with ca. 20% admixture with the main genetic cluster of the DC3.

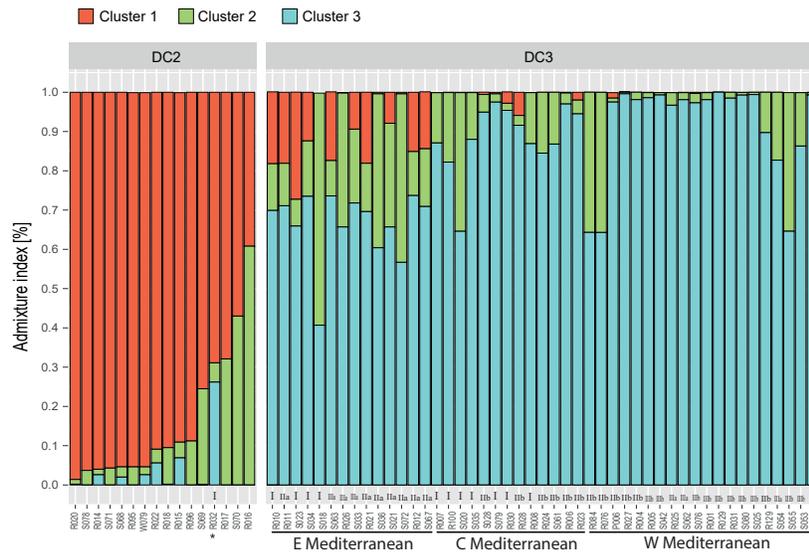


Figure 2: Admixture proportions for $K = 3$ genetic groups obtained from genetic structure analysis of nuclear data of Mediterranean *Dioscorea communis* samples through ten replicates in STRUCTURE (see text for details). Each sample is shown by a vertical bar partitioned according to its membership to one of the K clusters, represented in red, green and blue. The samples follow the same order as in the nuclear phylogenetic tree (Figure 1) and organized by clades (DC2 and DC3). The plastid clades (I, II and III, Figure 1) were labelled next to the sample codes. An asterisk (*) was used to represent the only sample of *D. communis* (R32) that was resolved as part of clade DC1 based on nuclear data, and in a different clade with plastid data (see Figure 1).

Based on our phylogenomic and genetic structure analyses, we performed divergence time analysis using treePL and BEAST (Supplementary Data Figure S4) using one representative sample of each defined clade and genetic group: *D. orientalis*, DC1, DC2 and DC3. Both approaches reconstructed an early Oligocene origin for the crown node of the Tamus and Borderea clades (BEAST 28.5 MY and treePL 28.2 MY), and a late Miocene split for the two species of the Borderea clade (BEAST 8.1 MY and treePL 10.6 MY). The split of *D. orientalis* from *D. communis* was inferred to have occurred during the early Miocene; BEAST 18.2 MY and treePL 20.6 MY), and the Macaronesian clade (DC1) likely diverged from the Mediterranean lineage of *D. communis* during the mid-Miocene; BEAST 13.5 MY and treePL 16.0 MY). The most recent split between the clades DC2 and DC3 was estimated to have taken place during the late Miocene (BEAST 5.6 MY and treePL 6.6 MY).

The mean and median values of allelic ratios were >2 in all cases for *D. orientalis* and clade DC1, and only ten samples showed allelic ratio values <2 : eight samples from clade DC2 and two samples from the eastern Mediterranean subclade of DC3 (Table 1). The lowest incidence of estimated polyploidy based on allelic ratio estimates was found in DC2, with 50% of the samples classified as diploids (Table 1).

Current and past overlaps in species distribution ranges between the Tamus lineages

ENM analysis was performed for each of the four lineages in the Tamus clade (Figure 3) to explore whether differences exist in (former) distribution between the four clades identified within the Tamus clade of *Dioscorea*. Predicted distribution models were estimated using available occurrence data points (DC1: 27, DC2: 26, DC3: 218, *D. orientalis*: 17) and selected bioclimatic variables (DC1: bio16, bio9, bio15, bio3 and bio8; DC2: bio16, bio14, bio6, bio15, bio3, bio8; DC3: bio4, bio8, bio16, bio15 and bio9; *D. orientalis*: bio16, bio14, bio6, bio15, bio3 and bio8; in order of importance); and showed high AUC values (DC1: 0.9989 ± 0.0003 , DC2: 0.980 ± 0.005 , DC3: 0.937 ± 0.002 , *D. orientalis*: 0.9994 ± 0.0001) and a 10% threshold was applied (DC1: 0.24, DC2: 0.29, DC3: 0.20, *D. orientalis*: 0.50 probability).

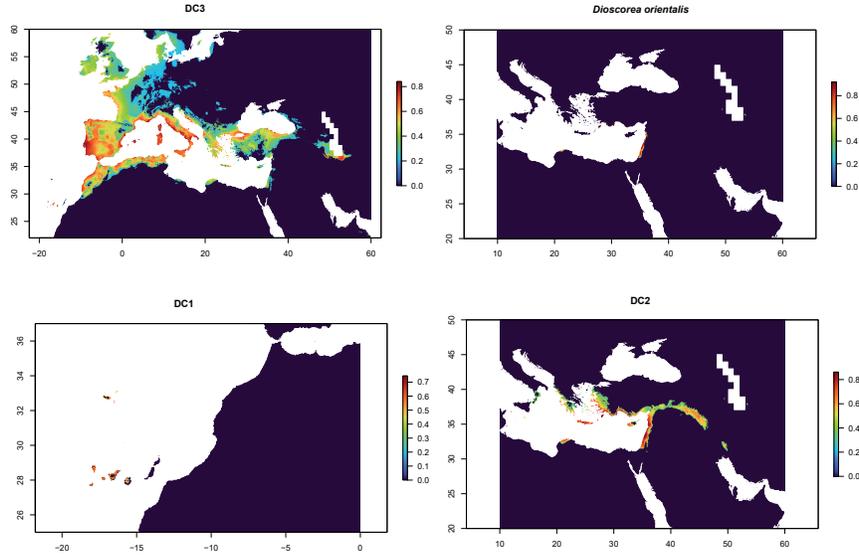


Figure 3: Environmental niche models (ENM) constructed using a current climate envelop and occurrence data for the samples of the Tamus clade of *Dioscorea* analysed in this study. ENMs were estimated for each of the four lineages identified in the Tamus clade with background maps adjusted to reflect their estimated potential distributions: DC3 clade of *D. communis*; DC2 clade of *D. communis*; the Macaronesian DC1 lineage of *D. communis*; and *D. orientalis*. A 10% threshold cropping was applied for each distribution range model (see Results). The legend represents the prediction of probability of occurrence.

The group formed by DC3 samples had its highest distribution probability in several Mediterranean areas, central and southern Europe (including England and Belgium to Crimea), north-western Africa and western Asia (Turkey, Syria, Caucasus, Caspian shores). The most optimal distribution for DC2 was in the eastern Mediterranean, specifically in the eastern Aegean islands and the Mediterranean coastal zone of Turkey, Lebanon and Israel. The distribution model of the latter group overlapped with the distribution range inferred for *D. orientalis*, which also presented its optimum in the eastern Mediterranean area, along the coasts of Lebanon, Syria, Palestine and Israel. The highest probabilities of potential distribution ranges for the DC1 group are restricted to the humid areas of the western Canary Isles (Figure 3), with Madeira and western Morocco showing a lower probability of occurrence. Overlap was observed between DC2, DC3 and *D. orientalis* in the eastern Mediterranean region, and between DC1 and DC3 in the Canary Islands.

The highest niche breadth obtained corresponded to the clade with the largest modelled distribution projection (i.e., 0.875 for DC3 clade), followed by DC2 (0.731), *D. orientalis* (0.516) and DC1 (0.499). We calculated Schoener's D and Hellinger's I indexes as metrics of niche overlap between pairs of distribution models. The highest overlaps between current niches were found between the DC3 clade with DC1 ($D = 0.52, I = 0.64$) and DC2 ($D = 0.48, I = 0.67$), whereas *D. orientalis* showed lower overlap values with DC2 ($D = 0.32, I = 0.60$) and DC3 ($D = 0.36, I = 0.48$). Although niche overlap between DC2 and *D. orientalis* clades showed the lowest values, a PCA using the raw bioclimatic data obtained from the studied samples did not differentiate between them (Figure 4); a better separation was, however, found between DC3 and DC1. These results are supported by observed differences between the four clades for each of the bioclimatic variables (Supplementary Data Figure S5).

S3). Regarding reproductive traits (Supplementary Data Table S2), we found significant differences in male individuals for the total length of the inflorescence, the length of pedicels and the number of fascicles, but the number of flowers did not show significant differences (Supplementary Data Table S3B). The following morphological traits showed Pearson correlations >0.7 (Supplementary Data Figure S7): leaf length, leaf width, leaf area, leaf perimeter, petiole, leaf coverage, main nerve length, leaf length ratio, inflorescence length and number of flowers. We therefore selected leaf area, leaf coverage and male inflorescence length as potential diagnostic variables. Post-hoc analyses were performed (Supplementary Data Table S3), and we used these statistical differences to describe the morphological variability of each genetic group in Table 2.

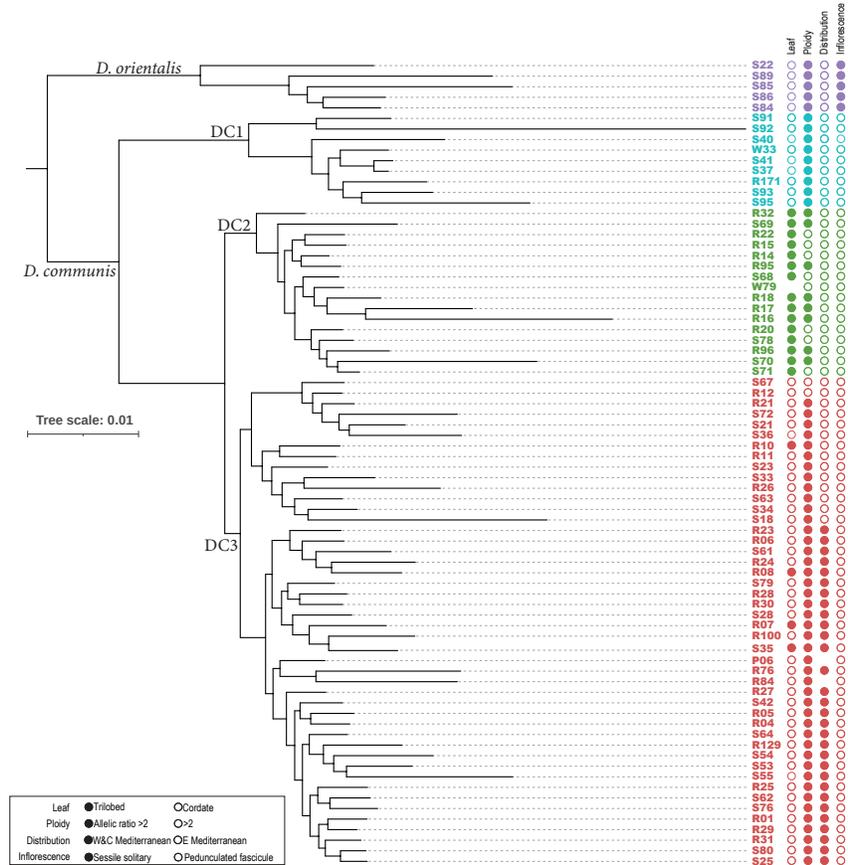


Figure 5: Phylogenetic tree reconstructed using the concatenated nuclear data obtained using Hyb-Seq for 76 samples of the Tamus clade of *Dioscorea*. Morphological, ploidy and geographic distribution trait variation is represented with circles for each sample. Leaf: filled circle, trilobed; empty circle, cordate; ploidy level: filled circle, allelic ratio >2 (estimated as polyploid); empty circle, allelic ratio <2 (estimated as diploid; see Results); geographic distribution: filled circle, W and C Mediterranean; empty circle, E Mediterranean; inflorescence: filled circle, sessile solitary; empty circle, pedunculated fascicle.

Morphological differences were found between the four clades and served as diagnostic characters in an identification key to distinguish the four distinct species (see Identification key below); these morphological features matched the genetic and niche data of these taxa. Our multidisciplinary approach combining genetic, spatial and morphometric data has delimited four taxa at species level, three species in the previously

recognized *Dioscorea communis sensu lato* , namely *D. communis sensu stricto* , *D. edulis* and *D. cretica* , plus *D. orientalis* .

Identification key for *Dioscorea communis*, *D. cretica*, *D. edulis* and *D. orientalis*

1a. Flowers sessile, solitary; perigonium white or violet in male inflorescences; leaf pedicel up to 1.6 cm
D. orientalis

1b. Flowers pedicellate in fascicles (2–4 flowers each); perigonium whitish-greenish or violet; leaf pedicel usually longer than leaf 2

2a. Leaves cordate-sagittate; perigonium violet; Macaronesia

D. edulis

2b. Leaves cordate-trilobed; perigonium whitish-greenish; not in Macaronesia. . . . 3

3a. Leaves cordate (rarely trilobed in Balearic Islands); flowers turbinate infundibuliform *D. communis*

3b. Leaves trilobed; flowers urceolate *D. cretica*

Taxonomic treatment and descriptions

1. *Dioscorea orientalis* (Thiébaud) Caddick & Wilkin

Basionym = *Tamus orientalis* Thiébaud., Bull. Soc. Bot. France 81: 119. 1934.

– Holotype: Lebanon. Between Batroun and Saïda, January 1933, *Thiébaud s.n.* (P00301666).

Perennial herb, glabrous. Stems simple or little branched. Leaves ovate, acuminate, cordate at the base of up to 56 × 48 mm, petiole with glandular base. Flowers sessile, in axillary spikes, hanging 1–23 (male), 1–4 (female) flowers. Whitish purple (i.e., including a range of shades of purple) perigonium; six lobes, recurved ovals. Six stamens, three stigmas and six naked filaments (female). Bracteoles 1–2 widely ovate, with final peak, adpressed perigonium, with a hull in the outer surface. Fleshy fruit, oblong, 7–11 × 6–10 mm, red at maturity. Pollen grains (Supplementary Data Figure S8) with two apertures with longer axis up to 40 µm and shorter axis up to 30 µm. Perforated with 1.9 perforations per µm and perforation size up to 0.85 µm, with spines. Eastern Mediterranean region. $2n$ = unknown.

2. *Dioscorea edulis* (Lowe) Campos, Wilkin & Viruel, *comb. nov.* (= Clade DC1)

=Basionym: *Tamus edulis* Lowe, Trans. Cambridge Philos. Soc. 4: 1. 1833.

–Type: Portugal, Madeira, Pta Moniz, 7 May 1828, Herb. Lowe 504 (K; K000099334!, K001081657!).

Heterotypic synonyms:

= *Tamus canariensis* Willd. ex Kunth, Enum. Pl. 5: 455. 1850, nom. illeg. pro syn. Type: Spain, Canary Islands, *Herbarium Willdenow no 18374* (B-W).

= *Tamus parviflora* Kunth, Enum. Pl. 5: 454. 1850. Holotype: Spain, Canary Islands, Teneriffa, ex Museo Paris, 1821 (B 10 0160963!), male plant.

Perennial herb, glabrous. Leaves varying from cordiform to sagittate, petiole 1.60–3.88 cm, leaves up to 125 × 110 mm, slightly wavy, coarse, chartaceous and secondary veins visible, but not prominent. Male inflorescence compound is arranged in fascicles of three or four flowers, compound inflorescence up to 12 cm in length and bearing up to 65 flowers in total. Female flowers composed of six violaceous tepals with erect filaments. Fruit a globose berry, 6–12 × 1–8 mm, reddish-orange. Pollen grains (Supplementary Data Figure S8) with two apertures with longer axis up to 36 µm and shorter axis up to 18 µm. Perforated with 3.9 perforations per µm and perforation size up to 0.37 µm, without spines. Macaronesia (Madeira, Gran Canaria, Tenerife, Gomera, Hierro, La Palma). $2n$ = 36, 48.

3. *Dioscorea cretica* (L.) Campos, Wilkin & Viruel, *comb. nov.* (= Clade DC2)

Basionym: *Tamus cretica* L., Sp. Pl., 1028. 1753.

= *Tamus communis* L. subsp. *cretica* (L.) Kit Tan, Notes Roy. Bot. Gard. Edinburgh 41(1): 47. 1983.

– Neotype (designated by Kit Tan in Davis, 1984, as isotype, corrected by Jarvis, 2007): "Habitat in Creta", Herb. Tournefort no. 283 (P-00665856!).

Perennial herb, glabrous. Leaves deeply trilobate, or apex or basal lobes elongated from a cordate leaf, up to 70 × 70 mm, with 3-9 prominent main nerves and secondary reticulated nerves not visible. Male inflorescence branched, to 25 cm long with up to 50 flowers in total, short female inflorescence up to 7 cm with 1–6 urceolate, greenish-yellow flowers. Fruit a globose berry, 7–11 × 6–10 mm, reddish-orange. Pollen grains (Supplementary Data Figure S8) with two apertures with longer axis up to 36 μm and shorter axis up to 29 μm. Perforated with 0.5 perforations per μm and perforation size up to 1.2 μm, without spines. Eastern Mediterranean. $2n$ = unknown.

4. *Dioscorea communis* (L.) Caddick & Wilkin (= Clade DC3)

Basionym: *Tamus communis* L., Sp. Pl.: 1028. 1753.

= *Tamus communis* (L.) f. *subtriloba* (Guss.) O.Bolòs & Vigo – Fl. Països Catalans 4: 171 (2001).

- Lectotype (designated by Ferrer-Gallego & Boisset, 2016): "habitat in Europa australi", *anon.*, Herb. Linn. No. 1181.2 (LINN!).

Heterotypic synonym: *Tamus communis* L. var. *subtriloba* Guss., *Fl. Siculae Syn.* 2(2): 880 (1884). *Tamus communis* L. f. *subtriloba* (Guss.) O.Bolòs & Vigo, *Fl. Països Catalans* 4: 171 (2001). Lectotype (designated by Ferrer-Gallego & Boisset, 2016): Italy, Sicily, Palermo, Vergine Maria, April–May, *anon.* (NAP).

Perennial herb, glabrous. Stem, striated, branched. Leaves cordate, rarely trilobed, up to 180 × 150 mm, with 3–9 prominent main nerves and secondary reticulated nerves not visible. Male inflorescence branched, up to 35 cm with up to 73 flowers in total, short female inflorescence up to 7 cm with 1–14 flowers, turbinate-infundibuliform, greenish-yellow. Fruit a somewhat tapered, globose berry, 7 to 12 mm, reddish-orange. Pollen grains (Supplementary Data Figure S8) with two apertures with longer axis up to 36 μm and shorter axis up to 30 μm. Perforated with 0.5 perforations per μm and perforation size up to 1.2 μm, without spines. Western and central Mediterranean, western Europe. $2n$ = 96.

Discussion

Species discovery based on integrative approaches

The biological species concept (Mayr, 1942) defines a species as a group of populations reproductively isolated from others. This concept is difficult to apply to species delimitation in flowering plants due to the high incidence of hybridization and introgression (Mitchell *et al.*, 2019). Although plant taxonomy has relied on morphological traits to differentiate and discover new taxa for centuries using a typological species concept (Haider, 2018), biological processes such as hybridization can obscure the morphological attributes used to differentiate species. Combined morphological and molecular approaches have been used to identify cryptic species in angiosperms (Maguilla and Escudero, 2016). Alternative species concepts have been proposed in plants to accommodate a broad spectrum of approaches, such as cytology, phytochemistry, anatomy, embryology or phylogenetics (De Queiroz, 2007; Aldhebiani, 2018). These are method-based concepts, such as the evolutionary species concept using phylogenetic inference, or the ecological species concept based on niche differentiation.

Here, we use an integrative approach combining morphological, phylogenetic, and ecological niche data to decipher species delimitation in the *Tamus* clade of *Dioscorea* and uncovered the existence of introgression in some individuals (Figure 2) that could at least partly explain the overlap in some of the morphological characteristics between taxa. The discovery of cryptic species in this group shapes our current understanding of it, specifically for what has been to date accepted as *D. communis* (Caddick *et al.*, 2002). Based on our

results, we propose the maintenance of *D. orientalis* as a species and divide *D. communis sensu lato* into three distinct species: *D. communis sensu stricto*, *D. edulis* and *D. cretica* (see Results; Figure 6).

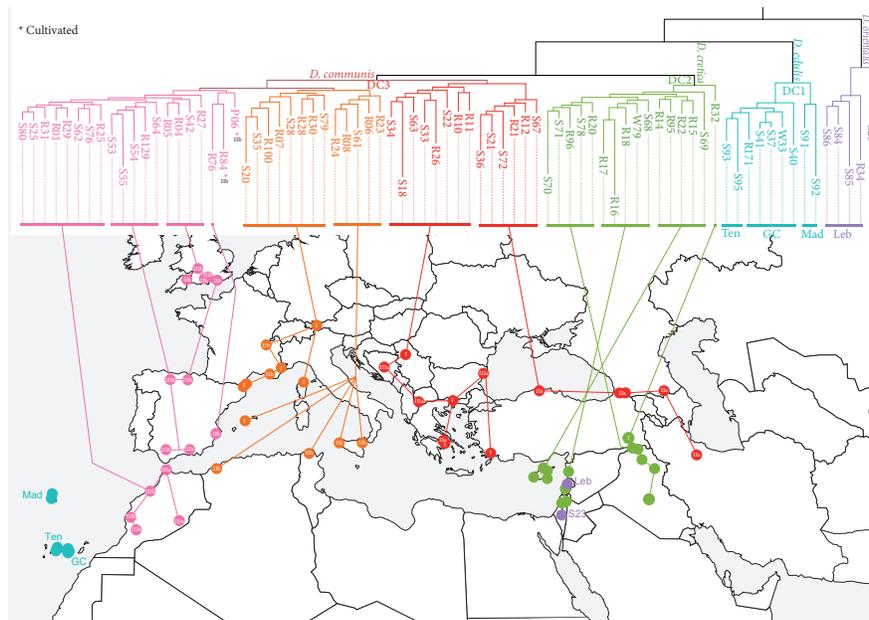


Figure 6: Geographic distribution of the 76 samples of the Tamus clade of *Dioscorea* coded according to their respective lineage in the phylogenetic tree based on concatenated nuclear data (Figure 1), *D. cretica* (DC2) in green, and subclade; and pink, orange, and red for western European, central Mediterranean and eastern Mediterranean clades in *D. communis s.s.* (DC3), respectively. For *D. orientalis*, in purple, codes in the map indicate their phylogenetic position named “Leb” (Lebanon) and “S23”. For *D. edulis* (DC1), in light blue, codes in the map indicate their phylogenetic position named “Mad” (Madeira), “Ten” (Tenerife) and “GC” (Gran Canaria). Roman numbers in the mapped dot samples represent their respective plastid tree lineage (Figure 1). Asterisks indicate cultivated samples in Botanic Gardens of unknown geographic origin.

The emergence of HTS methodologies has allowed the detection of cryptic species (Carstens and Satler, 2013), the resolution of complex phylogenetic trees (Bogarín *et al.*, 2018; Frajman *et al.*, 2019; Hassemmer *et al.*, 2019; Yang *et al.*, 2019) and the reconstruction of evolutionary patterns in extinct species (Moreno-Aguilar *et al.*, 2020). Thus, HTS methods provide new sources of useful data to clarify phylogenetic enigmas that classical molecular methods could not decipher (e.g., Urtubey *et al.*, 2018). Among HTS methods, target capture is currently being used in a broad number of plant systematics and evolutionary studies due to its versatility to successfully sequence hundreds of loci from highly degraded DNA samples (Brewer *et al.*, 2019, Viruel *et al.*, 2019). Herbarium samples constitute a valuable and vast source of information for morphological and niche modelling approaches, and recently proved to be equally important for phylogenetic studies based on DNA sequence data obtained using HTS methods. In our study, we used herbarium material, with the oldest specimen sequenced collected in 1788, and a custom bait capture kit targeting 260 low copy nuclear genes (Soto Gomez *et al.*, 2019), to reveal the evolutionary patterns and relationships between taxa belonging to the Tamus clade of *Dioscorea* (Figure 1). By sequencing 76 samples of the Tamus clade, the phylogenomic and genetic clustering approaches revealed extensive infraspecific variability in *D. communis sensu lato*, clearly dividing it into three genetic groups, each showing a distinct geographic distribution across the Mediterranean and western Europe (Figure 6). Two of these genetic groups are congruent with the previously recognized *Tamus edulis* and *T. cretica*, which were recently placed within the large morphological variability and wide

distribution of *D. communis s.l.*. HTS methodologies applied to herbarium material allowed us to recognize the species rank for these genetic groups and to support the split of *D. communis s.l.* into *D. edulis*, *D. cretica* and *D. communis*, and to maintain *D. orientalis* as a species.

Whole-genome duplication events (i.e., polyploidy) have been commonly reported across flowering plants and have been correlated with diversification of gene functions and new genetic architecture, which could be linked with adaptative traits (Wendel *et al.*, 2018). Increased speciation events have been observed in some angiosperm lineages reported to have a high incidence of whole-genome duplication events (Wood *et al.*, 2009; Zhan *et al.*, 2016). Polyploidy is a common phenomenon, which has been frequently reported in several *Dioscorea* species (Viruel *et al.*, 2008), although defining ploidy of the Tamus and Borderea clades has been challenging. The two *Dioscorea* species belonging to the Borderea clade, *D. chouardii* and *D. pyrenaica*, have chromosome counts of $2n=24$. Based on the discovery of allotetraploidy using microsatellite markers (Segarra-Moragues *et al.*, 2003), it was proposed that the chromosome base number for the Borderea clade was $x=6$ (see also Viruel *et al.*, 2008). Extrapolating this find to the sister Tamus clade, the known chromosome counts reported for *D. communis s.s.* of $2n=36$ and 48 (Al-Shehbaz and Schubert, 1989; Viruel *et al.*, 2019) would therefore represent hexaploid and octoploid forms, respectively. Similarly, the Macaronesian *D. edulis*, with $2n=96$, would be 16-ploid assuming a base chromosome number of $x=6$. Using flow cytometry to estimate ploidy in *D. communis s.s.*, multiple ploidies were observed (1C-values ranging from 0.41 to 1.36 pg; Viruel *et al.*, 2019). The chromosome number and genome size of *D. orientalis* and *D. cretica* remain unknown, but allelic ratios estimated for each SNP per sample using HTS data can be used as a proxy to distinguish between diploid and polyploid forms when multiple ploidies are expected in a group of plants (Viruel *et al.*, 2019). Median and mean values of allelic ratios based on the number of reads supporting each SNP were recently proposed to classify *Dioscorea* samples as diploid forms when the allelic ratio is <2 , and polyploids when >2 (Viruel *et al.*, 2019). For example, all samples of *D. edulis* had mean and median allelic ratios >2 (Table 1), confirming the polyploid nature of this species based on chromosome data. In all cases, *D. orientalis* samples studied here showed allelic ratios >2 and would therefore be estimated to be a polyploid species (Table 1). For *D. communis s.s.*, all samples were estimated to be polyploids except for two samples of the clade DC3 from the eastern Mediterranean with mean and median allelic ratio values <2 (samples S67 and R12, Table 1). Samples estimated to be diploid based on allelic ratios were also observed in *D. cretica*, with half of the samples (eight) having average and median allelic ratios <2 (Table 1). The incidence of diploid forms, as estimated using allelic ratio values, in the eastern Mediterranean will require further investigation applying cytological and flow cytometry methodologies.

Evolutionary patterns of the Tamus clade of *Dioscorea* in the Mediterranean

Overall evolutionary patterns of *Dioscorea* lineages have been thoroughly studied using plastid markers (e.g., Wilkin *et al.*, 2005; Hsu *et al.*, 2013; Maurin *et al.*, 2016; Viruel *et al.*, 2016) and low copy nuclear genes (e.g. Viruel *et al.*, 2018, Soto Gomez *et al.*, 2019). Previous studies determined that yams diverged and expanded since the Late Cretaceous, probably from Laurasia, and that a split of the African-Mediterranean lineage, which includes the Tamus clade, likely occurred in the Oligocene, following a westward migration ca. 33 MY (Viruel *et al.*, 2016). Fossil records indicate that *Dioscorea* ancestors persisted in Europe during the Oligocene (Andreànzky, 1959). Based on data from four plastid markers, the split between the two Mediterranean clades, Borderea and Tamus, was estimated to have occurred during the late Oligocene (ca. 25.7 MY) (Viruel *et al.*, 2016), a similar divergence time to the one we obtained with our analyses based on 260 nuclear genes, which indicate that this divergence took place in the early Oligocene (28.2–28.4 MY); Supplementary Data Figure S4). Two narrowly endemic species of the Borderea clade survived in refugia in the Pyrenean mountains, *D. chouardii*, a critically endangered species with only one known locality (Martínez & Otano, 2011), and *D. pyrenaica*, with a slightly wider distribution in the central Pyrenees and Pre-Pyrenees (Segarra & Catalán, 2005; Catalan *et al.*, 2006; Segarra-Moragues & Catalan, 2008; García *et al.*, 2012). A previous study (Viruel *et al.*, 2016) estimated a split between these two species during the early Pliocene (ca. 4.3 MY), whereas our results indicated that this divergence likely occurred during the late Miocene (8.1–10.6 MY).

The differences observed in divergence times for the *Tamus* clade in comparison with previous studies are a consequence of the newly recognized species (*D. cretica*). In Viruel *et al.* (2016), the crown node of the *Tamus* clade was estimated to be 15.3 MY (*D. edulis* (*D. communis*, *D. orientalis*)), and the subsequent split between *D. orientalis* and *D. communis* at 10.4 MY. However, the samples of *D. orientalis* included in Viruel *et al.* (2016) have now been reidentified as *D. cretica*, and thus the older age estimates herein for the crown node of the *Tamus* clade (20.6–18.2 MY) demonstrate an early split of *D. orientalis* in the eastern Mediterranean, followed by a split of *D. edulis* ca. 16.0–13.5 MY, and the divergence of *D. communis s.s.* and *D. cretica* ca. 6.6–5.6 MY. Given the findings presented here, with a sampling representative of the whole distribution range of the *Tamus* clade across the circum-Mediterranean region, we conclude that the divergence times estimated here are more robust and taxonomically more representative, which allowed us to reassess the species delimitation in this group.

The Mediterranean region is considered one of the major biodiversity hotspots of the world (Médail and Quézel, 1997). Fossil records and evolutionary studies have confirmed that the ancestors of several plant lineages were part of a tropical flora that occupied the Mediterranean region during the Miocene and early Pliocene (Suc *et al.*, 2018). The drastic subsequent climatic changes that came during the Pliocene (3.5–2.4 MY), with a significant drop in temperature and a marked seasonality in thermal and rainfall regimes, impacted the diversification patterns of plant lineages and resulted in narrow endemics in the margins of the distribution range of their sister species (e.g., *Ceratonia oreothauma* Hillc., G.P.Lewis & Verdc.; Viruel *et al.*, 2020). The diversification of species in the *Tamus* clade likely occurred during the Miocene when subtropical climatic conditions were present across the Mediterranean (Suc *et al.*, 2018). The most recent common ancestor of all *Tamus* clade taxa likely diversified during the early Miocene (20.6–18.2 MY), when the lineages that gave rise to the current *D. orientalis* and the clade comprising the three lineages of *D. communis s.l.* likely split. This was followed by a subsequent split of the Macaronesian *D. edulis* that would have taken place in the mid-Miocene (16.0–13.5 MY), after the formation of some of the Canary Islands, which has been estimated to start around 23 MY (Sanmartín *et al.*, 2008; Florencio *et al.*, 2021). The most recent split between *D. communis s.s.* and *D. cretica* is estimated to have occurred during the Messinian (Miocene, 6.6–5.6 MY). During this period, the significant and rapid lowering of the sea level of the Mediterranean also resulted in new terrestrial biogeographical connections allowed by the formation of land-bridges.

Several phylogeographic studies have attempted to explain the biodiversity patterns and processes that shaped the Mediterranean region and its development into one of the world’s biodiversity hotspot (e.g., Nieto Feliner, 2014; Thompson, 2021). Two main areas of high plant endemism were identified in the western (Iberian Peninsula and Morocco) and eastern Mediterranean (including Turkey and Greece) (Médail and Quézel, 1997). In both these areas, Quaternary glaciations likely played a major role shaping the distribution of species and left a footprint in the genetic structure of many Mediterranean species, particularly in refugia (Médail and Diadema, 2009). Western and eastern genetic groups have been identified in phylogeographic patterns of several Mediterranean plants, leading to disjunct distributions in some cases, such as in *Microcne-mum* Ung.-Sternb (Amaranthaceae) and *Mandragora* L. (Solanaceae) (Kadereit and Yaprak, 2008; Volis *et al.*, 2018), or by differentiating morphotypes that later hybridized in intermediate zones (e.g., *Quercus ilex* L., Lumaret *et al.*, 2002). The strong geographic influence in the genetic structuring of *D. communis* across the Mediterranean may have also been slightly influenced by bird dispersal. Bird dispersals have contributed to shaping the postglacial recolonization of the Mediterranean, such as seen in *Frangula alnus* Mill. (Hampe *et al.*, 2003). The birds that consume berries produced by species of the *Tamus* clade, mainly blackbirds (*Turdus merula* L.), robins (*Erithacus rubecula* L.) and blackcaps (*Sylvia communis* Latham) (Chiscano, 1983; Herrera, 1984), are predominantly sedentary birds or have modern migratory routes that do not strictly coincide with the past and current distribution patterns estimated in this study (Adriaensen, 1988; Burfield and Van Bommel, 2004). It would thus be useful to analyse the patterns of genetic structure at the population level of *D. communis s.s.* in more detail and in connection with the possible magnitude of ornithochory, which has never been studied in detail to our knowledge.

Changes in ploidy, morphological differences and introgression between the central Mediterranean and western European populations have been shown to have occurred between the *D. communis s.s.* and *D. cretica*

lineages (Figures 2, 5). The central-eastern Mediterranean area constitutes the contact region between these two species and is congruent with the introgression patterns found in our study (Figure 2). Five out of 16 samples studied of *D. cretica* exhibited <20% of admixture index with *D. communis* s.s. , and all individuals of *D. communis* belonging to the eastern clade of *D. communis* s.s. showed <20% of admixture index with *D. cretica* (Figure 2). However, only four individuals from the central Mediterranean and western European subclades of *D. communis* were detected as introgressing with a sister species, and one sample of *D. cretica* was placed in a clade of *D. communis* s.s. in the phylogenetic tree based on plastid data (R32, Figure 1). These results are congruent with their potential distribution in disjunct refugia followed by secondary contact through recolonization, and by maintaining some capacity of interspecific gene flow between closely related species (see Viruel *et al.*, 2021). The topological incongruencies found between the nuclear and plastid phylogenetic trees, indicative of plastid capture events (Figure 1), are congruent with these hypothesized introgression patterns: plastid clades I and II are found in the central and eastern Mediterranean lineages without a clear geographic separation (Figure 6), whereas clade III is uniquely found in the western part of the Mediterranean where lower introgression events have been inferred.

Conclusions

The identification of new plant species usually requires a broad understanding of taxon boundaries applying multidisciplinary methodologies. Our study exemplifies the complexity of identifying new species by integrating different types of data: target capture sequencing data from herbarium specimens to reveal phylogenomic patterns and introgression between clades, differences in allelic ratios to estimate ploidy, spatial analysis estimating current and past distribution ranges and niche overlaps, and macro- and micromorphometric comparisons. By integrating all these results, we have newly corroborated the existence of four species in the Mediterranean *Tamus* clade of *Dioscorea* , maintaining *D. orientalis* as a distinct species, and demonstrating that *D. edulis* and *D. cretica* are species discrete from the synonymy of the morphologically variable *D. communis* .

Acknowledgements

We are thankful to the herbaria of Kew (K) and Seville (SEV) for providing access to their collections for this study and to the colleagues who collected samples in the field. This work was supported by a Marie Skłodowska-Curie Individual Fellowship (704464 – YAMNOMICS – MSCA-IF-EF-ST) and the RBG Kew-funded pilot study to Juan Viruel; and Erasmus+ Programme for Miguel Campos. Emma Kelley received funding from the Muriel Kohn Pokross '34 Travel/Internship Fund, Smith College, Massachusetts. Pilar Catalán was supported by a European Social Fund and Spanish Aragón Government Bioflora grant.

Literature cited

- Adriaensen F. 1988. An analysis of recoveries of robins (*Erithacus rubecula*) ringed or recovered in Belgium: winter distributions. *Le Gerfaut* 781: 25–42.
- Aldhebiani AY. 2018. Species concept and speciation. *Saudi Journal of Biological Sciences* 25: 437–440. doi: 10.1016/j.sjbs.2017.04.013
- Andreàndzky G. 1959. Contributions à la connaissance de la flore de l'Oligocène inférieur de la Hongrie et un essai sur la reconstruction de la flore contemporaine. *Acta Botanica Academiae Scientiarum Hungaricae* 5: 1–37.
- Asiedu R, Sartie A. 2010. Crops that feed the world 1. Yams. *Food Security* 2: 305–315.
- Bogarín D, Pérez-Escobar OA, Groenenberg D, *et al.*, 2018. Anchored hybrid enrichment generated nuclear, plastid and mitochondrial markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex. *Molecular Phylogenetics and Evolution* 129: 27–47. doi: 10.1016/j.ympev.2018.07.014
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. doi: 10.1093/bioinformatics/btu170

- Braconnot P, Otto-Bliesner B, Harrison S, *et al.* 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. doi: 10.5194/cp-3-261-2007
- Brewer GE, Clarkson JJ, Maurin O, *et al.* 2019. Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science* 10: 1102. doi: 10.3389/fpls.2019.01102
- Burfield I, van Bommel F. 2004. Birds in Europe: population estimates, trends and conservation status: 12 (BirdLife Conservation Series). Cambridge, GB: BirdLife International.
- Caddick LR, Wilkin P, Rudall PJ, Hedderson TAJ, Chase MW. 2002b. Yams reclassified: a recircumscription of Dioscoreaceae and Dioscoreales. *Taxon* 51: 103–114. doi: doi.org/10.2307/1554967
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. doi: 10.1093/bioinformatics/btp348
- Carstens BC, Satler JD. 2013. The carnivorous plant described as *Sarracenia alata* contains two cryptic species. *Biological Journal of the Linnean Society* 109: 737–746. doi: 10.1111/bij.12093
- Catalán P, Segarra-Moragues JG, Palop-Esteban M, Moreno C, Gonzalez-Candelas F. 2006. A Bayesian approach for discriminating among alternative inheritance hypotheses in plant polyploids: the allotetraploid origin of genus *Borderea* (Dioscoreaceae). *Genetics* 172: 1939–1953.
- Chiscano JLP. 1983. La ornitocoria en la vegetación de Extremadura. *Studia Botanica* 2: 155–168.
- Clement M, Snell Q, Walker P, Posada D, Crandall K. 2002. TCS: estimating gene genealogies. *16th International Parallel and Distributed Processing Symposium (IPDPS 2002)*, 15–19 April 2002, Fort Lauderdale, FL, USA, CD-ROM/Abstracts Proceedings.
- Criscuolo NG, Angelini C. 2020. StructuRly: a novel shiny app to produce comprehensive, detailed and interactive plots for population genetic analysis. *PLoS One* 15: e0229330. doi: 10.1371/journal.pone.0229330
- Davis PH. (ed.) 1984. *Flora of Turkey and the East Aegean Islands*, vol. 8. Edinburgh: Edinburgh University Press, pp. 552–554.
- De Luca V, Salim V, Atsumi SM, Yu F. 2012. Mining the biodiversity of plants: a revolution in the making. *Science* 336: 1658–1661. doi: 10.1126/science.1217410
- De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886. doi: 10.1080/10635150701701083
- Doyle J, Doyle JL. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochemical Bulletin* 19: 11–15.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 1–8. doi: 10.1186/1471-2148-7-214
- Earl DA, vonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* 4: 359–361. doi: 10.1007/s12686-011-9548-7
- Escudero M, Nieto Feliner G, Pokorny L, Spalink D, Viruel J. 2020. Phylogenomic approaches to deal with particularly challenging plant lineages. *Frontiers in Plant Science* 11: 591762. doi: 10.3389/fpls.2020.591762
- Fay MF, Gargiulo R, Viruel J. 2019. The present and future for population genetics, species boundaries, biogeography and conservation. *Botanical Journal of the Linnean Society* 191: 299–304. doi: 10.1093/botlinnean/boz076

- Ferrer-Gallego PP, Boisset F. 2016. Typification of *Dioscorea communis* and its synonym *Tamus communis* var. *subtriloba*(Dioscoreaceae). *Phytotaxa* 260(3): 258–266.
- Florencio M, Patiño J, Nogué S, *et al* . 2021. Macaronesia as a Fruitful Arena for Ecology, Evolution, and Conservation Biology. *Frontiers in Ecology and Evolution* 9: 718169. doi: 10.3389/fevo.2021.718169
- Frajman B, Závieská E, Gamisch A, Moser T, The STEPPE Consortium, Schönswetter P. 2019. Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbiaseguieriana* s.l. (Euphorbiaceae). *Molecular Phylogenetics and Evolution* 134: 238–252. doi: 10.1016/j.ympev.2018.10.046
- García MB, Espadaler X, Olesen JM. 2012. Extreme reproduction and survival of a true cliffhanger: the endangered plant *Borderea chouardii* (Dioscoreaceae). *PLoS One* 7: e44657. doi: 10.1371/journal.pone.0044657
- Gent PR, Danabasoglu G, Donner LJ, *et al*. 2011. The community climate system model version 4. *Journal of Climate* 24: 4973–4991. doi: 10.1175/2011JCLI4083.1
- Goñi Martínez D, Guzmán Otano D. 2011. *Borderea chouardii* . *The IUCN Red List of Threatened Species* 2011: e.T162110A5540643.
- Haider N. 2018. A Brief Review on Plant Taxonomy and its Components. *Journal of Plant Science Research* 34: 275–290. doi: 10.32381/JPSR.2018.34.02.17
- Hampe A, Arroyo J, Jordano P, Petit RJ. 2003. Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. *Molecular Ecology* 12: 3415–3426. doi: 10.1046/j.1365-294X.2003.02006.x
- Hassemer G, Bruun-Lund S, Shipunov AB, *et al* . 2019. The application of high-throughput sequencing for taxonomy: the case of *Plantago* subg. *Plantago* (Plantaginaceae). *Molecular Phylogenetics and Evolution* 138: 156–173. doi: 10.1016/j.ympev.2019.05.013
- Hasumi H, Emori S. 2004. *K-1 coupled model (MIROC) description (K-1 Technical Report 1)* . Tokyo: Center for Climate System Research, University of Tokyo.
- Herrera CM. 1984. A study of avian frugivores, bird-dispersed plants, and their interaction in Mediterranean scrublands. *Ecological Monographs* 54: 2–23. doi: 10.2307/1942454
- Hsu KM, Tsai JL, Chen MY, Ku HM, Liu SC. 2013. Molecular phylogeny of *Dioscorea* (Dioscoreaceae) in East and Southeast Asia. *Blumea: Journal of Plant Taxonomy and Plant Geography* 58: 21–27. doi: 10.3767/000651913X669022
- Hua W, Kong W, Cao X, *et al*. 2017. Transcriptome analysis of *Dioscorea zingiberensis* identifies genes involved in diosgenin biosynthesis. *Genes & Genomics* 39: 509–520. doi: 10.1007/s13258-017-0516-9
- Jarvis CE. 2007. *Order out of chaos . Linnaean plant names and their types* . London: Linnean Society of London with the Natural History Museum, 1016 pp.
- Johnson MG, Gardner EM, Liu Y, *et al*. 2016. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Applications in Plant Sciences* 4: 1600016. doi: 10.3732/apps.1600016
- Johnson MG, Pokorny L, Dodsworth S, *et al*. 2019. A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Systematic Biology* 68: 594–606. doi: 10.1093/sysbio/syy086
- Kadereit G, Yaprak AE. 2008. *Microcnemum coralloides*(Chenopodiaceae-Salicornioideae): An example of intraspecific East-West disjunctions in the Mediterranean region. *Anales del Jardín Botánico de Madrid* 68: 415–426. doi: 10.3989/ajbm.2008.v65.i2.303

- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066. doi: 10.1093/nar/gkf436
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35: 4453–4455. doi: 10.1093/bioinformatics/btz305
- Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110–1116. doi: 10.1111/2041-210X.12410
- Linnaeus C. 1753. *Species plantarum vol. 1* . Stockholm: L. Salvius.
- Lumaret R, Mir C, Michaud H, Raynal V. 2002. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Molecular Ecology* 11: 2327–2336. doi: 10.1046/j.1365-294X.2002.01611.x
- Maguilla E, Escudero M. 2016. Cryptic Species Due to Hybridization: A Combined Approach to Describe a New Species (*Carex* : Cyperaceae). *PLoS One* 11: e0166949. doi: 10.1371/journal.pone.0166949
- Martin FW, Degras L. 1978. *Tropical yams and their potential. Part 6. Minor cultivated Dioscorea species*. Washington, DC.: USDA, Science and Education Administration.
- Martínez GD, Otano GD. 2011. *Borderea chouardii* . The IUCN Red List of Threatened Species 2011: e.T162110A5540643.
- Maurin O, Muasya, AM, Catalán P, *et al.* 2016. Diversification into novel habitats in the Africa clade of *Dioscorea*(Dioscoreaceae): erect habit and elephant’s foot tubers. *BMC Evolutionary Biology* 16: 1–17. doi: 10.1186/s12862-016-0812-z
- Mayr E. 1942. *Systematic and the origin of species* . New York: Columbia University Press.
- Médail F, Quézel P. 1997. Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean. *Annals of the Missouri Botanical Garden* 84: 112–127. doi: 10.2307/2399957
- Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* 36: 1333–1345.
- Mitchell N, Campbell LG, Ahern JR, *et al* . 2019. Correlates of hybridization in plants. *Evolution Letters* 3: 570–585. doi: <https://doi.org/10.1002/evl3.146>
- Moreno-Aguilar MF, Arnelas I, Sánchez-Rodríguez A, Viruel J, Catalán P. 2020. Museomics Unveil the Phylogeny and Biogeography of the Neglected Juan Fernandez Archipelago *Megalachne* and *Podophorus* Endemic Grasses and Their Connection With Relict Pampean-Ventanian fescues. *Frontiers in Plant Science* 11: 819. doi: 10.3389/fpls.2020.00819
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274. doi: 10.1093/molbev/msu300
- 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. doi: 10.1016/j.ppees.2014.07.002
- Otto-Bliesner ABL, Marshall SJ, Overpeck JT, *et al.* 2006. Simulating Arctic Climate Warmth and Icefield Retreat in the Last Interglaciation. *Science* 311: 1751–1753. doi: 10.1126/science.1120808
- POWO. 2022. *Plants of the world online* . Richmond, UK: Royal Botanic Gardens, Kew. <http://www.plantsoftheworldonline.org/>
- Price EJ, Wilkin P, Sarasan V, Frase PD. 2016. Metabolite profiling of *Dioscorea* (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds. *Scientific Reports* 6: 29136. doi: 10.1038/srep29136

- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945–959. doi: 10.1093/genetics/155.2.945
- R Core Team. 2022. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Sanmartín I, Van der Mark P, Ronquist F. 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography* 35: 428–449. doi: 10.1111/j.1365-2699.2008.01885.x
- Schneider C, Rasband W, Eliceiri K. 2012. ImageJ. *Fundamentals of Digital Imaging in Medicine* 9: 185–188. doi: 10.1007/978-1-84882-087-6_9
- Segarra JG, Catalán P. 2005. *Borderea Miègeville* in Aedo C. and A. Herrero (eds.). *Flora Iberica* 21: 11–14. Real Jardín Botánico de Madrid, CSIC.
- Segarra-Moragues JG, Catalán P. 2008. Glacial survival, phylogeography, and a comparison of microsatellite evolution models for explaining population structure in two species of dwarf yams (*Borderea*, Dioscoreaceae) endemic to the central Pyrenees. *Plant Ecology and Diversity* 1: 229–243.
- Šmarda P. 2008. DNA ploidy level variability of some fescues (*Festuca* subg. *Festuca*, Poaceae) from Central and Southern Europe measured in fresh plants and herbarium specimens. *Biologia* 63: 349–367. doi: 10.2478/s11756-008-0052-9
- Smith SA, O’Meara BC. 2012. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690. doi: 10.1093/bioinformatics/bts492
- Soltis DE, Albert VA, Leebens-Mack J, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348. doi: 10.3732/ajb.0800079
- Soto Gomez M, Pokorny L, Kantar, MB, et al. 2019. A customized nuclear target enrichment approach for developing a phylogenomic baseline for *Dioscorea* yams (Dioscoreaceae). *Applications in Plant Sciences* 7: 1–13. doi: 10.1002/aps3.11254
- Suc J-P, Popescu S-M, Fauquette S, et al. 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. doi: 10.3406/ecmed.2018.2044
- Swofford DL. 2002. *PAUP**. *Phylogenetic analysis using parsimony (*and other methods)*. Version 4.0b10. Sunderland; Sinauer Associates.
- Thompson JD. 2021. *Plant evolution in the Mediterranean: insights for conservation (2nd edn)*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198835141.001.0001
- Urtubey E, Baeza CM, López-Sepúlveda P, et al. 2018. Systematics of *Hypochaeris* section *Phanoderis* (Asteraceae, Cichorieae). *Systematic Botany Monographs* 106: 1–204.
- Villaverde T, Pokorny L, Olsson S, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 36–650. doi: 10.1111/nph.15312
- Viruel J, Segarra-Moragues JG, Perez-Collazos E, et al. 2008. The diploid nature of the Chilean *Epipetrum* and a new base number in the Dioscoreaceae. *New Zealand Journal of Botany* 46: 327–339.
- Viruel J, Segarra-Moragues JG, Perez-Collazos E, et al. 2010. Systematic revision of the *Epipetrum* group of *Dioscorea* (Dioscoreaceae) endemic to Chile. *Systematic Botany* 35: 40–63. doi: 10.1600/036364410790862579

Viruel J, Segarra-Moragues JG, Raz L, *et al.* 2016. Late Cretaceous-Early Eocene origin of yams (*Dioscorea*, Dioscoreaceae) in the Laurasian Palaeartic and their subsequent Oligocene-Miocene diversification. *Journal of Biogeography* 43: 750–762. doi: 10.1111/jbi.12678

Viruel, J., Conejero, M., Hidalgo, O., *et al.* 2019. A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers in Plant Science* 10: 937. doi: 10.3389/fpls.2019.00937

Viruel J, Forest F, Paun O, *et al.* 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. <https://doi.org/10.1093/botlinnean/boy013>

Viruel J, Le Galliot N, Pironon S, *et al.* 2020. A strong east–west Mediterranean divergence supports a new phylogeographic history of the carob tree (*Ceratonia siliqua*, Leguminosae) and multiple domestications from native populations. *Journal of Biogeography* 47: 460–471. doi: 10.1111/jbi.13726

Viruel J, Kantar MB, Gargiulo R, *et al.* 2021. Crop wild phylorelatives (CWPs): phylogenetic distance, cytogenetic compatibility and breeding system data enable estimation of crop wild relative gene pool classification. *Botanical Journal of the Linnean Society* 195: 1–33. doi: 10.1093/botlinnean/boaa064

Volis S, Fogel K, Tu T, Sun H, Zaretsky M. 2018. Evolutionary history and biogeography of *Mandragora* L. (Solanaceae). *Molecular Phylogenetics and Evolution* 129: 85–95. doi: 10.1016/j.ympev.2018.08.015

Waltari E, Hijmans RJ, Peterson AT, *et al.* 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One* 2: e563. doi: 10.1371/journal.pone.0000563

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62: 2868–2883. doi: 10.1111/j.1558-5646.2008.00482.x

Warren DL, Glor RE, Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33: 607–611. doi: 10.1111/j.1600-0587.2009.06142.x

Weiss CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* 19: 122. doi: 10.1186/s12859-018-2128-z

Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* 49: 1–7. doi: 10.1016/j.gde.2018.01.004

Wilkin P, Schols P, Chase MW, *et al.* 2005. A plastid gene phylogeny of the yam genus, *Dioscorea*: Roots, Fruits and Madagascar. *Systematic Botany* 30: 736–749. doi: 10.1600/036364405775097879

Wood TE, Takebayashi N, Barker MS, *et al.* 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the U.S.A.* 106: 13875–13879. doi: 10.1073/pnas.0811575106

Yang L, Kong H, Huang JP, Kang M. 2019. Different species or genetically divergent populations? Integrative species delimitation of the *Primulina hochiensis* complex from isolated karst habitats. *Molecular Phylogenetics and Evolution* 132: 219–231. doi: 10.1016/j.ympev.2018.12.011

Zhan SH, Drori M, Goldberg EE, *et al.* 2016. Phylogenetic evidence for cladogenetic polyploidization in land plants. *American Journal of Botany* 103: 1252–1258. doi: 10.3732/ajb.1600108

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30. doi: 10.1186/s12859-018-2129-y

Figure 1. Phylogenomic trees reconstructed using the concatenated nuclear (left) and plastid data (right) obtained using Hyb-Seq for 76 samples of the *Tamus* clade of *Dioscorea*. Samples included are representative of its distribution range across the Mediterranean region. Filled circles represent branches with support values over 90%, while lower values are shown on branches. Colours represent the main clades found in the nuclear

tree, see Results (*D. orientalis* violet, DC1 blue, DC2 green, DC3 red), and roman numbers identifying the subclades in *D. communis* in the plastid tree (I, II, III; subclades within each clade were indicated with lowercase letters a and b). Dashed lines connect the same samples in the nuclear and plastid trees.

Figure 2. Admixture proportions for $K = 3$ genetic groups obtained from genetic structure analysis of nuclear data of Mediterranean *Dioscorea communis* samples through ten replicates in STRUCTURE (see text for details). Each sample is shown by a vertical bar partitioned according to its membership to one of the K clusters, represented in red, green and blue. The samples follow the same order as in the nuclear phylogenetic tree (Figure 1) and organized by clades (DC2 and DC3). The plastid clades (I, II and III, Figure 1) were labelled next to the sample codes. An asterisk (*) was used to represent the only sample of *D. communis* (R32) that was resolved as part of clade DC1 based on nuclear data, and in a different clade with plastid data (see Figure 1).

Figure 3. Environmental niche models (ENM) constructed using a current climate envelop and occurrence data for the samples of the Tamus clade of *Dioscorea* analysed in this study. ENMs were estimated for each of the four lineages identified in the Tamus clade with background maps adjusted to reflect their estimated potential distributions: DC3 clade of *D. communis* ; DC2 clade of *D. communis* ; the Macaronesian DC1 lineage of *D. communis* ; and *D. orientalis* . A 10% threshold cropping was applied for each distribution range model (see Results). The legend represents the prediction of probability of occurrence.

Figure 4. Principal Component Analysis (PCA) conducted with 19 bioclimatic variables for samples of the Tamus clade of *Dioscorea* across the Mediterranean region separated into the four main lineages (see Results): Macaronesian DC1 clade of *D. communis* in light blue; *D. communis* clade DC2 in green; *D. communis* DC3 clade in red; and *D. orientalis* in purple.

Figure 5. Phylogenetic tree reconstructed using the concatenated nuclear data obtained using Hyb-Seq for 76 samples of the Tamus clade of *Dioscorea* . Morphological, ploidy and geographic distribution trait variation is represented with circles for each sample. Leaf: filled circle, trilobed; empty circle, cordate; ploidy level: filled circle, allelic ratio >2 (estimated as polyploid); empty circle, allelic ratio <2 (estimated as diploid; see Results); geographic distribution: filled circle, W and C Mediterranean; empty circle, E Mediterranean; inflorescence: filled circle, sessile solitary; empty circle, pedunculated fasciculate.

Figure 6. Geographic distribution of the 76 samples of the Tamus clade of *Dioscorea* coded according to their respective lineage in the phylogenetic tree based on concatenated nuclear data (Figure 1), *D. cretica* (DC2) in green, and subclade; and pink, orange, and red for western European, central Mediterranean and eastern Mediterranean clades in *D. communis* s.s. (DC3), respectively. For *D. orientalis* , in purple, codes in the map indicate their phylogenetic position named “Leb” (Lebanon) and “S23”. For *D. edulis* (DC1), in light blue, codes in the map indicate their phylogenetic position named “Mad” (Madeira), “Ten” (Tenerife) and “GC” (Gran Canaria). Roman numbers in the mapped dot samples represent their respective plastid tree lineage (Figure 1). Asterisks indicate cultivated samples in Botanic Gardens of unknown geographic origin.

Supplementary Data

Table S1. *Dioscorea* samples used with the name of the species, collector’s name and collection number, date of collection and location of the individuals. *DNA extracted from fresh material.

Table S2. Traits used for morphometric analyses of the Tamus clade of *Dioscorea* samples divided into macromorphological (vegetative and reproductive) and micromorphological (pollen) variables, with units and description of the trait measured.

Table S3. Table showing the results of Lilliefors test, Levene test and the subsequent ANOVA or Kruskal-Wallis test for the selected variables analysed in the Tamus clade of *Dioscorea* samples; asterisk indicates the level of signification ($p < 0.05$ = one asterisk; $p < 0.01$ = two asterisks). Bonferroni post-Hoc test showing the difference between groups, the level of significance is shown with asterisks ($p < 0.05$ = one asterisk; $p <$

0.01 = two asterisks). For a full description of the traits and abbreviations used, see Supplementary Data Table S2.

Figure S1. Dendrograms constructed from correlation values (Pearson’s correlation matrices) between bioclimatic variables for the samples belonging to the *D. communis* and *D. orientalis* clades. Three cluster dendrograms are shown corresponding to samples identified as belonging to the different phylogenetic clades identified (see Results). Clade DC1: *D. communis* samples from Macaronesia; DC2: samples of *D. communis* from the Eastern Mediterranean, DC3: samples of *D. communis* from western and central Europe.

Figure S2. Results obtained in Structure Haverster based on genetic structure analysis of *D. communis* nuclear data through 10 replicates of Structure for each *K* (see Materials and Methods). The best delta *K* value is highlighted in bold and represented in a graphic below.

Figure S3. Genetic structure analysis of nuclear SNP data of the studied Mediterranean *D. communis* samples with STRUCTURE imposing ancestral admixture and correlated allelic frequencies models for *K* =2 to *K* =6 hypothetical genetic groups.

Figure S4. Divergence time estimations and chronograms of the Tamus clade of *Dioscorea* and outgroup based on nuclear sequence data and compiled with treePL (a) and BEAST (b) showing mean node ages in MY. BEAST maximum credibility tree (b) includes bars showing the 95% highest posterior density intervals for nodes.

Figure S5. Boxplots representing the differences between the four main clades: DC1, DC2, DC3 (*D. communis*) and *D. orientalis* in the Tamus clade of *Dioscorea* for each of the 19 analysed WorldClim current climate bioclimatic variables.

Figure S6. Past projections of the environment niche model (ENMs) constructed for the four Tamus clades of *Dioscorea* with a 10% lower probability threshold cropping applied for each plot. Mid Holocene (MH), Last Glacial Maximum (LGM) and Last Interglacial (LIG) climate envelopes for the DC3 clade of *D. communis*; DC2 clade of *D. communis*; for *D. orientalis* clade, and for the Macaronesian DC1 clade.

Figure S7. Pairwise Pearson correlation matrix calculated between the morphological variables measured to distinguish between Tamus clade taxa.

Figure S8. SEM images of the pollen grains of *D. communis* DC3 clade (A, B), *D. cretica* (C, D), *D. edulis* (E, F) and *D. orientalis* (G, H).

Table 1. Target capture efficiency for the 76 herbarium samples of the Tamus clade of *Dioscorea* and four outgroup taxa studied. *DNA extracted from fresh material. ‘Paired-end reads’ is the number of reads obtained after quality trimming with Trimmomatic. ‘Length %’ is the recovery length percentage calculated based on the nucleotides retrieved per sample relative to the full length of the reference target gene set. ‘Enrich. eff.’ is the proportion of reads on target; ‘bp nDNA’ corresponds to the total number of bp retrieved for nuclear genes; ‘bp pDNA’ is the total number of bp retrieved for the plastome; ‘P(%)’ is the percentage of polymorphic sites (P) calculated as the coefficient between the SNPs and the total bp recovered for each sample; AR, allelic ratio; ‘%>2’ the percentage of SNPs with an allelic ratio <2.

Species	Clade	Sample	Year collected	Paired-end reads	Enrich. eff.
<i>D. orientalis</i>	<i>D. orientalis</i>	S23	1909	2980360	0.52
		R34	1958	43612	0.51
		S85	2008	1020080	0.44
		S84	1957	2068305	0.52
		S86	1957	795799	0.51
		S92	1960	299677	0.31
<i>D. communis</i>	Clade DC1 (Macaronesia)	S91	1960	1637342	0.45
		S40	2008	10339082	0.52
		W33	2003	1370314	0.42

	S37	2008	10212202	0.50
	S41	2008	16196619	0.49
	R171	1788	5772951	0.38
	S95	1975	663825	0.45
	S93	1867	2521444	0.46
Clade DC2 (Trilobed leaves, Eastern Mediterranean)	R32	1956	1238692	0.42
	S69	1905	782006	0.11
	R15	1979	2481398	0.24
	R22	1955	1206365	0.29
	R95	1951	1483426	0.38
	R14	1967	3918262	0.33
	S68	2007	782006	0.11
	W79	2010	1637458	0.48
	R18	1942	578726	0.60
	R16	1943	164691	0.33
	R17	1943	281364	0.59
	R20	2004	1925906	0.26
	S78	1955	2705587	0.23
	R96	1970	1514421	0.41
	S71	1963	3234283	0.22
	S70	1971	120618	0.24
Clade DC3 (Eastern Mediterranean)	S67	1962	1879003	0.58
	R12	1957	2716763	0.35
	R21	1937	1624504	0.22
	S72	2004	198920	0.26
	S21	2004	991796	0.30
	S36	1957	298300	0.53
	R11	1936	1955793	0.25
	R10	1931	1859291	0.18
	S22	1969	1300587	0.23
	R26	1858	759663	0.09
	S33	1932	1895199	0.10
	S63	1999	1742636	0.22
	S18	1900	1014934	0.18
	S34	1959	1441193	0.19
Clade DC3 (Central Mediterranean)	R23	1978	1319618	0.23
	R06	1965	2526479	0.28
	S61	1990	1760430	0.09
	R08	1931	1169102	0.19
	R24	1930	895451	0.12
	S79	2016	3929348	0.28
	R30	2016	1677314	0.25
	R28	2016	1999986	0.34
	S28	1967	2114903	0.28
	R07	1847	1912677	0.13
	R100	1906	1536342	0.11
	S35	1929	1846641	0.12
	S20	1929	776828	0.17
Clade DC3 (Western European)	P06*	2018	793772	0.57
	R84*	2018	1525356	0.22
	R76	2004	1653906	0.44

		R27	2017	2060161	0.34
		S42	2017	1432342	0.33
		R04	1956	2921528	0.30
		R05	1948	2418558	0.32
		S64	2015	4389479	0.30
		R129	1932	1893572	0.26
		S54	1976	1744816	0.16
		S55	1980	124177	0.57
		S53	1971	1811439	0.13
		R25	1934	1189706	0.15
		S76	1972	1882846	0.58
		S62	2003	2401716	0.20
		R29	2016	2093906	0.34
		R01	2016	5236808	0.41
		R31	2016	8920132	0.56
		S25	2016	3546012	0.35
		S80	2016	1341416	0.23
<i>D. elephantipes</i>	Outgroup	W32	2001	1020073	0.46
<i>D. sylvatica</i>	Outgroup	S49*	2018	2658330	0.54
<i>D. chouardii</i>	Borderea	D66	2003	5285968	0.35
<i>D. pyrenaica</i>	Borderea	W34	1956	1648853	0.29

Table 2. Morphological traits and variability in the four morphological groups identified in the *Tamus* clade of *Dioscorea*, one corresponding to *D. orientalis*, and three to the currently recognized *D. communis* s.l.: Macaronesian clade DC1 (= *D. edulis*), trilobed leaf Eastern Mediterranean clade (DC2) (= *D. cretica*) and the DC3 cordate leaf clade of *D. communis* (= *D. communis* s.s.) (see Figure 1).

		<i>D. communis</i> DC3	<i>D. communis</i> DC2
Leaf	Petiole (cm)	0.24–10.1	1.7–8.8
	Shape	Cordate	Trilobed
	Size (mm) length × width	Up to 180 × 150	Up to 70 × 70
	Margin	Thickened, not undulate	Thickened, not undulate
Male inflorescence	Consistency	Chartaceous	Chartaceous
	Secondary veins	Reticulated, usually not evident	Reticulated, usually evident
	Disposition	In a fascicle, up to four flowers	In a fascicle, up to four flowers
	Branching	Compound	Compound
Female inflorescence	Size (cm)	Up to 35	Up to 25
	Number of flowers	3–4 per fascicle, up to 73 per inflorescence	1–4 per fascicle, up to 73 per inflorescence
	Pedicel (mm)	0.5–5.4	1.04–5.17
	Number of flowers	1–14 per inflorescence	1–6 per inflorescence
Fruit	Pedicel (mm)	4.7–11.3	2.5–9.3
	Size (mm)	7–11, 6–10	7–11, 6–10
Seed	Shape	Globose to ellipsoid	Globose to ellipsoid
	Number	1–6	1–6
Pollen	Color	Dark brown	Dark brown
	Apertures	2	2
	Λονγεστ αξις (μμ)	31–36	31–36
	Σηορτερ αξις (μμ)	27–30	26–29
	Περφορατιον σιζε (μμ)	0.6–1.2	0.6–1.2
	Περφορατιονς/μμ²	0.5	0.5

Spines

No

No
