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

Miguel Campos, Emma Kelley, Barbara Gravendeel, Frédéric Médail, J. M. Maarten Christenhusz, Michael F. Fay, Pilar Catalán, Ilia J. Leitch, Félix Forest, Paul Wilkin and Juan Viruel

DOI:
https://doi.org/10.1093/aob/mcad018

Annals of Botany **131**: 635–654, 2023
https://doi.org/10.1093/aob/mcad018, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

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

Miguel Campos[1,2,3], Emma Kelley[1], Barbara Gravendeel[4,5], Frédéric Médail[6], J. M. Maarten Christenhusz[1], Michael F. Fay[1,7], Pilar Catalán[3,8], Ilia J. Leitch[1], Félix Forest[1], Paul Wilkin[1] and Juan Viruel[1,*]

[1]Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK, [2]Department of Plant Biology and Ecology, University of Seville, 41012, Spain, [3]Universidad de Zaragoza-Escuela Politécnica Superior de Huesca, 22071, Huesca, Spain, [4]Naturalis Biodiversity Center, Leiden 2333 CR, The Netherlands, [5]Radboud Institute for Biological and Environmental Sciences, RIBES 6500 GL, Nijmegen, The Netherlands, [6]Institut Méditerranéen de Biodiversité et d'Écologie marine et continentale (IMBE), Aix Marseille University, Avignon University, CNRS, IRD, Campus Aix, Technopôle de l'Environnement Arbois-Méditerranée, F-13545 Aix-en-Provence cedex 4, France, [7]School of Biological Sciences, University of Western Australia, Crawley, WA 6009, Australia and [8]Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza 50018, Spain

**For correspondence. E-mail j.viruel@kew.org, juanviruel@gmail.com*

Received: 22 October 2022 Returned for revision: 12 January 2023 Editorial decision: 17 January 2023 Accepted: 23 January 2023
Electronically published: 22 January 2023

• **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.

Key words: Hyb-Seq, target capture, yams, phylogeography, polyploidy, phylogenomics, *Dioscorea communis*, *Tamus edulis*, *Dioscorea orientalis*, *Dioscorea cretica*.

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* and *D. cayenensis*) that 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; Price *et al.*, 2016; Hua *et al.*, 2017). 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*, native to Montenegro and Albania, and *D. caucasica*, found in Georgia and Caucasian Russia), the Borderea clade, which contains two well-defined and narrow endemic species from the Pyrenean mountains (*D. chouardii* and *D. pyrenaica*), 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*, 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*, restricted to Lebanon and Israel. However, like in many *Dioscorea* clades (Viruel *et al.*, 2010), the Tamus clade has had multiple previous taxonomic circumscriptions. Based on their berry fruits, the Tamus clade was considered as a separate genus, *Tamus*, distinct from *Dioscorea*, until 2002 (Caddick *et al.*, 2002). The latter study unified several previously recognized genera (*Epipetrum*, *Nanarepenta*, *Rajania*, *Testudinaria* and *Borderea*) to maintain the monophyly of *Dioscorea*. Moreover, fruits with different degrees of fleshiness have been observed in other *Dioscorea* species (e.g. *D. ovinala*, *D. antaly*; Caddick *et al.*, 2002).

Linnaeus (1753) recognized two species: *Tamus communis*, with cordate leaves and a Mediterranean distribution, and *T. cretica*, 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*, *T. parviflora*, *T. norsa* and *T. canariensis*), while *T. cirrhosa*, *T. cordifolia* and *T. racemosa* 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*), and *T. communis* f. *subtriloba* was described as a variety with trilobed leaves found in the Balearic Islands and north-eastern Spain (Catalonia). All these names were subsequently united under the currently accepted *D. communis* (Caddick *et al.*, 2002); however, this decision was not supported by phylogenetic or morphological data. The second species currently recognized in the Tamus clade, *D. orientalis*, was originally described as *T. orientalis*, 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. Fay *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 designed either 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, morphometric 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.

MATERIALS AND METHODS

Plant material

Seventy-six herbarium specimens identified as *Dioscorea communis* or *D. orientalis* were used to obtain genomic, spatial and morphometric data. They were selected as being representative of the macromorphological diversity and geographic distribution ranges of the two species as currently circumscribed (Supplementary Data Table S1). The genomic sampling also included material that was 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* and *D. sylvatica*).

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 designed for *Dioscorea* (Soto Gomez *et al.*, 2019). Genomic libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) with AMPure XP magnetic beads and NEBNext® Multiplex Oligos for Illumina® (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) 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 with 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-Gutiérrez *et al.*, 2009) using the *-automated1* command. Phylogenomic trees were reconstructed using the concatenated and partitioned nuclear DNA (nDNA) and plastid DNA (pDNA) datasets independently, and for each nuclear gene independently, using maximum likelihood analysis as implemented in RAxML-NG (Katoh *et al.*, 2019) and IQ-TREE (Nguyen *et al.*, 2015), with a GTR+GAMMA substitution

model and 1000 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 v.1.7 (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 ΔK 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 million years (MY) and a maximum age of 40 MY, based on the age estimates from Viruel *et al.* (2016; 95% highest posterior density interval of 24.3469–39.2223 MY). In BEAST analysis, 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. Convergence and mixing of the Markov chain Monte-Carlo in BEAST analysis was assessed using the effective sampling size (ESS > 200) criterion in TRACER v.1.7.1, and all parameters showed ESS values >200. The treePL analysis was conducted in two consecutive runs: (1) applying the 'prime' option to select the most optimal parameter values; and (2) 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 the Results section) under current and past climatic conditions using the maximum entropy algorithm implemented in the R package 'Maxent' (Phillips *et al.*, 2017). 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 Fig. S1). We selected a different set of uncorrelated variables for each geographic region with a high percentage 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; and bio3 (isothermality), bio4 (temperature seasonality), bio16 (precipitation of wettest quarter) and bio18 (precipitation of warmest quarter) for the Macaronesian region. The ENM analyses were carried out under current climatic conditions, and projected to climatic conditions of the Mid Holocene (MH, ~6000 years ago), the Last Glacial Maximum (LGM, ~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 section) to maximize the reliability of the results and discard false occurrences using the package 'raster' (v.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 1000 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 subsequent visualization and selection of variables using a dendrogram cluster. The following ten uncorrelated variables with the highest contribution to the PCA were selected: 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 at 2 kV (Hitachi High-Tech Corporation, Tokyo, Japan).

A Pearson correlation of >0.7 was used as a threshold to exclude correlated variables from the analysis. The normality Kolmogorov–Smirnov test with Lilliefors correction and

Levene's test for homoscedasticity were applied to the variables following a normal distribution. Subsequently, comparisons among taxonomic units identified in phylogenomic analyses (see the Results section) were conducted using Kruskal–Wallis and Bonferroni *post hoc* tests. Statistical analyses were performed using the 'nortest', 'Hmisc', 'corrplot', 'PerformanceAnalytics' and 'car' packages in 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 in the SRA repository: PRJNA525269 and PRJNA895370. 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* (Fig. 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 s.l.* clade in the nuclear tree (Fig. 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. Clade DC3 was subsequently further subdivided into three subclades: a first divergence of an eastern Mediterranean subclade, and a subsequent split of central Mediterranean and western Europe subclades. Overall bootstrap support was >90 % for most of the nodes (Fig. 1). This phylogenetic topology was congruent with the reconstructions obtained with ASTRAL-III (Supplementary Data Fig. S2A) and SVDquartets (Supplementary Data Fig. S2B).

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 (Fig. 1). The remaining Mediterranean *D. communis* samples (DC3) were intermingled in three plastid subclades (I, II and III in Fig. 1). In the nuclear tree, 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 (Fig. 1). These subclades were also represented in the haplotype network (Supplementary Data Fig. S2C).

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 ΔK value (15821.9) was obtained for $K = 3$ genetic groups (Supplementary Data Fig. S3). 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 in 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 (Fig. 2). The genetic profile of one DC3 sample from the eastern Mediterranean subclade showed a percentage of membership >20 in Cluster 1, and nine DC3 samples had membership percentages of >20 in 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 Fig. S4). In all cases, sample R32 was always resolved with ~20 % admixture with the main genetic cluster of DC3.

Based on our phylogenomic and genetic structure analyses, we performed divergence time analysis using treePL and BEAST (Supplementary Data Fig. S5) 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 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).

TABLE 1. *Target capture efficiency for the 76 herbarium samples of the Tamus clade of* Dioscorea *and the four outgroup taxa studied. *DNA extracted from fresh material. The column headed 'Paired-end reads' shows 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 base pairs retrieved for nuclear genes; 'bp pDNA' is the total number of base pairs retrieved for the plastome; 'P(%)' is the percentage of polymorphic sites (P) calculated as the coefficient between the SNPs and the total number of base pairs recovered for each sample; AR, allelic ratio; '%>2', percentage of SNPs with an allelic ratio <2*

Species	Clade	Sample	Year collected	Paired-end reads	Enrich. eff.	Assembled target genes	Genes at 50%	Length %	bp nDNA	bp pDNA	SNPs	P(%)	Mean AR	Median AR	%<2
D. orientalis	*D. orientalis*	S23	1909	2 980 360	0.526	260	252	34.5	146 415	146 223	1652	1.13	2.79	2.96	18.71
		R34	1958	43 612	0.516	208	34	34.1	144 729	76 269	1474	1.02	2.28	2.00	45.9
		S85	2008	1 020 080	0.448	260	134	50.3	213 231	66 768	835	0.39	2.71	2.80	31.9
		S84	1957	2 068 305	0.528	260	244	82.1	348 456	149 412	779	0.22	2.42	2.05	47.0
		S86	1957	795 799	0.513	260	245	77.9	330 582	110 763	697	0.21	2.57	2.40	34.7
D. communis	Clade DC1 (Macaronesia)	S92	1960	299 677	0.311	255	0	10.6	45 171	34 242	427	0.95	2.73	2.67	22.0
		S91	1960	1 637 342	0.455	260	239	79.2	335 889	141 522	388	0.12	2.13	2.01	41.8
		S40	2008	10 339 082	0.525	260	256	93.1	394 977	145 812	991	0.25	2.16	2.05	41.9
		W33	2003	1 370 314	0.429	260	253	91.1	386 241	145 767	1025	0.27	2.83	3.05	19.5
		S37	2008	10 212 202	0.502	260	257	92.6	392 901	145 089	1299	0.33	2.07	2.04	44.8
		S41	2008	16 196 619	0.492	260	256	92.6	392 970	141 618	1459	0.37	2.04	2.02	46.8
		R171	1788	5 772 951	0.381	260	240	80.4	341 028	133 494	1129	0.33	2.80	2.75	14.3
		S95	1975	663 825	0.454	260	142	51.1	216 750	143 307	705	0.33	2.49	2.27	34.9
		S93	1867	2 521 444	0.460	260	230	72.5	30 8190	50 046	1327	0.43	2.87	2.81	17.8
	Clade DC2 (Trilobed leaves, Eastern Mediterranean)	R32	1956	1 238 692	0.426	260	252	88.7	37 6359	14 4702	2304	0.61	2.62	2.74	22.6
		S69	1905	782 006	0.114	257	169	56.9	241 176	119 904	833	0.35	2.63	2.50	27.6
		R15	1979	2 481 398	0.248	260	254	90.9	38 5665	14 1975	1466	0.38	1.46	1.22	85.3
		R22	1955	1 206 365	0.297	260	255	91.1	386 331	141 084	0606	0.16	1.63	1.36	78.5
		R95	1951	1 483 426	0.382	260	251	89.4	379 077	143 058	803	0.21	2.45	2.18	47.4
		R14	1967	3 918 262	0.334	260	253	91.3	387 393	141 879	2230	0.58	1.86	1.62	67.9
		S68	2007	782 006	0.114	257	169	90.8	385 149	77 787	1404	0.36	1.64	1.29	81.5
		W79	2010	1 637 458	0.485	260	254	91.5	387 951	140 592	1065	0.27	1.48	1.21	77.56
		R18	1942	578 726	0.604	260	246	84.3	357 756	59 802	1186	0.33	3.03	3.28	15.1
		R16	1943	164 691	0.330	256	172	76.6	324 924	70 617	3367	1.04	2.56	2.50	30.7
		R17	1943	281 364	0.595	260	231	58.3	247 470	136 824	3805	1.54	2.95	3.00	8.86
		R20	2004	1 925 906	0.265	260	253	90.5	383 769	141 294	871	0.23	1.63	1.25	74.3
		S78	1955	2 705 587	0.233	260	249	85.4	362 205	144 948	525	0.14	1.94	1.34	61.7
		R96	1970	1 514 421	0.410	260	246	86.1	365 292	143 595	772	0.21	2.83	3.18	30.3
		S71	1963	3 234 283	0.220	260	251	87.1	369 312	145 470	808	0.22	1.94	1.53	65.6
		S70	1971	120 618	0.245	243	34	29.4	124 917	144 003	259	0.21	2.59	2.50	39.0

TABLE 1. *Continued*

Species	Clade	Sample	Year collected	Paired-end reads	Enrich. eff.	Assembled target genes	Genes at 50%	Length %	bp nDNA	bp pDNA	SNPs	P(%)	Mean AR	Median AR	%<2
	Clade DC3 (Eastern Mediterranean)	S67	1962	1 879 003	0.587	260	247	83.9	355 812	145 620	1064	0.30	1.47	1.27	83.7
		R12	1957	2 716 763	0.355	260	251	89.6	380 058	145 545	1411	0.37	1.65	1.29	81.7
		R21	1937	1 624 504	0.226	260	250	86.9	368 490	146 073	472	0.13	2.25	2.29	40.5
		S72	2004	198 920	0.266	256	161	55.0	233 424	143 145	878	0.38	2.67	2.67	32.9
		S21	2004	991 796	0.301	260	181	61.4	260 340	147 450	641	0.25	2.64	2.83	32.8
		S36	1957	298 300	0.537	259	160	54.6	231 513	93 522	512	0.22	2.44	2.40	30.7
		R11	1936	1 955 793	0.256	260	251	87.5	370 965	146 694	4029	1.09	2.30	2.27	41.1
		R10	1931	1 859 291	0.187	260	251	86.8	368 178	145 881	1079	0.29	2.90	2.94	13.7
		S22	1969	1 300 587	0.233	259	55	89.3	378 783	112 770	859	0.23	2.58	2.66	23.6
		R26	1858	759 663	0.095	258	173	60.4	256 383	147 990	2769	1.08	2.81	2.80	19.5
		S33	1932	1 895 199	0.103	260	218	71.9	304 818	150 111	972	0.32	2.62	2.58	23.9
		S63	1999	1 742 636	0.226	260	252	89.5	379 584	143 289	1452	0.38	2.59	2.74	24.4
		S18	1900	1 014 934	0.181	256	2	11.2	47 622	31 500	332	0.70	2.33	2.00	38.0
		S34	1959	1 441 193	0.191	260	250	83.8	355 620	144 501	775	0.22	2.84	2.90	12.1
	Clade DC3 (Central Mediterranean)	R23	1978	1 319 618	0.238	260	253	87.4	370 599	143 055	814	0.22	2.76	2.89	23.6
		R06	1965	2 526 479	0.280	260	251	88.0	373 401	148 890	3306	0.89	2.35	2.42	35.8
		S61	1990	1 760 430	0.090	259	233	73.6	312 126	151 014	812	0.26	2.32	2.19	36.1
		R08	1931	1 169 102	0.192	260	242	79.4	337 011	146 019	2148	0.64	3.03	3.17	9.73
		R24	1930	895 451	0.121	259	214	71.1	301 512	144 822	1340	0.44	2.77	2.75	20.7
		S79	2016	3 929 348	0.283	260	256	92.5	392 268	151 239	3010	0.77	2.48	2.67	27.3
		R30	2016	1 677 314	0.256	260	253	91.3	387 120	144 900	2464	0.64	2.59	2.73	24.8
		R28	2016	1 999 986	0.342	260	253	91.3	387 408	142 863	3335	0.86	2.51	2.67	26.7
		S28	1967	2 114 903	0.288	260	250	85.6	363 141	148 317	1107	0.30	2.66	2.74	25.6
		R07	1847	1 912 677	0.136	260	244	79.9	339 057	104 967	1365	0.40	2.81	3.00	21.5
		R100	1906	1 536 342	0.114	260	202	66.0	279 993	146 673	903	0.32	2.51	2.43	30.6
		S35	1929	1 846 641	0.125	260	226	73.7	312 624	148 557	902	0.29	2.57	2.67	33.0
		S20	1929	776 828	0.172	260	9	21.1	89 439	144 906	520	0.58	2.37	2.00	36.9

TABLE 1. *Continued*

Species	Clade	Sample	Year collected	Paired-end reads	Enrich. eff.	Assembled target genes	Genes at 50%	Length %	bp nDNA	bp pDNA	SNPs	P(%)	Mean AR	Median AR	%<2
	Clade DC3 (Western European)	P06*	2018	793 772	0.576	260	253	91.6	388 401	142 761	3371	0.87	2.51	2.72	25.8
		R84*	2018	1 525 356	0.226	260	240	86.2	365 586	148 377	5476	1.50	2.89	2.90	12.9
		R76	2004	1 653 906	0.448	260	239	86.3	365 946	144 525	2086	0.57	3.15	3.38	13.8
		R27	2017	2 060 161	0.340	260	253	91.7	388 935	143 38	2810	0.72	2.46	2.66	33.0
		S42	2017	1 432 342	0.336	260	254	92.3	391 608	143 334	2576	0.66	2.48	2.57	28.7
		R04	1956	2 921 528	0.304	260	253	89.1	377 823	149 163	2165	0.57	2.60	2.73	23.2
		R05	1948	2 418 558	0.322	260	251	87.4	370 722	144 150	3443	0.93	2.22	2.19	44.2
		S64	2015	4 389 479	0.307	260	255	92.3	391 437	142 944	2165	0.55	2.34	2.54	37.7
		R129	1932	1 893 572	0.262	260	243	78.7	333 720	146 958	1221	0.37	2.43	2.13	42.6
		S54	1976	1 744 816	0.164	260	214	68.7	291 213	112 173	662	0.23	2.50	2.40	33.8
		S55	1980	124 177	0.576	254	26	28.6	121 518	30 666	294	0.24	2.37	2.00	42.2
		S53	1971	1 811 439	0.130	260	221	72.3	306 672	145 788	714	0.23	2.38	2.30	33.5
		R25	1934	1 189 706	0.151	260	251	83.5	354 219	141 843	758	0.21	2.49	2.46	31.5
		S76	1972	1 882 846	0.589	260	247	85.2	361 521	145 740	1424	0.39	2.72	2.82	16.2
		S62	2003	2 401 716	0.200	260	249	88.4	375 072	150 357	1452	0.39	2.24	2.18	43.6
		R29	2016	2 093 906	0.341	260	252	91.5	388 347	142 563	2415	0.62	2.57	2.75	24.14
		R01	2016	5 236 808	0.415	260	255	92.0	390 150	142 731	3043	0.78	2.41	2.68	32.3
		R31	2016	8 920 132	0.566	260	255	92.1	390 780	143 364	2972	0.76	2.41	2.72	31.1
		S25	2016	3 546 012	0.358	260	252	90.6	384 462	147 363	1004	0.26	2.43	2.55	42.5
		S80	2016	1 341 416	0.231	260	252	88.6	375 798	142 617	764	0.20	2.46	2.50	34.0
D. elephantipes	Outgroup	W32	2001	1 020 073	0.467	260	254	95.2	403 968						
D. sylvatica	Outgroup	S49*	2018	2 658 330	0.541	260	255	94.4	400 443	150 822					
D. chouardii	Borderea	D66	2003	5 285 968	0.355	260	257	92.5	392 553	146 505					
D. pyrenaica	Borderea	W34	1956	1 648 853	0.298	260	237	76.8	325 572	146 466					

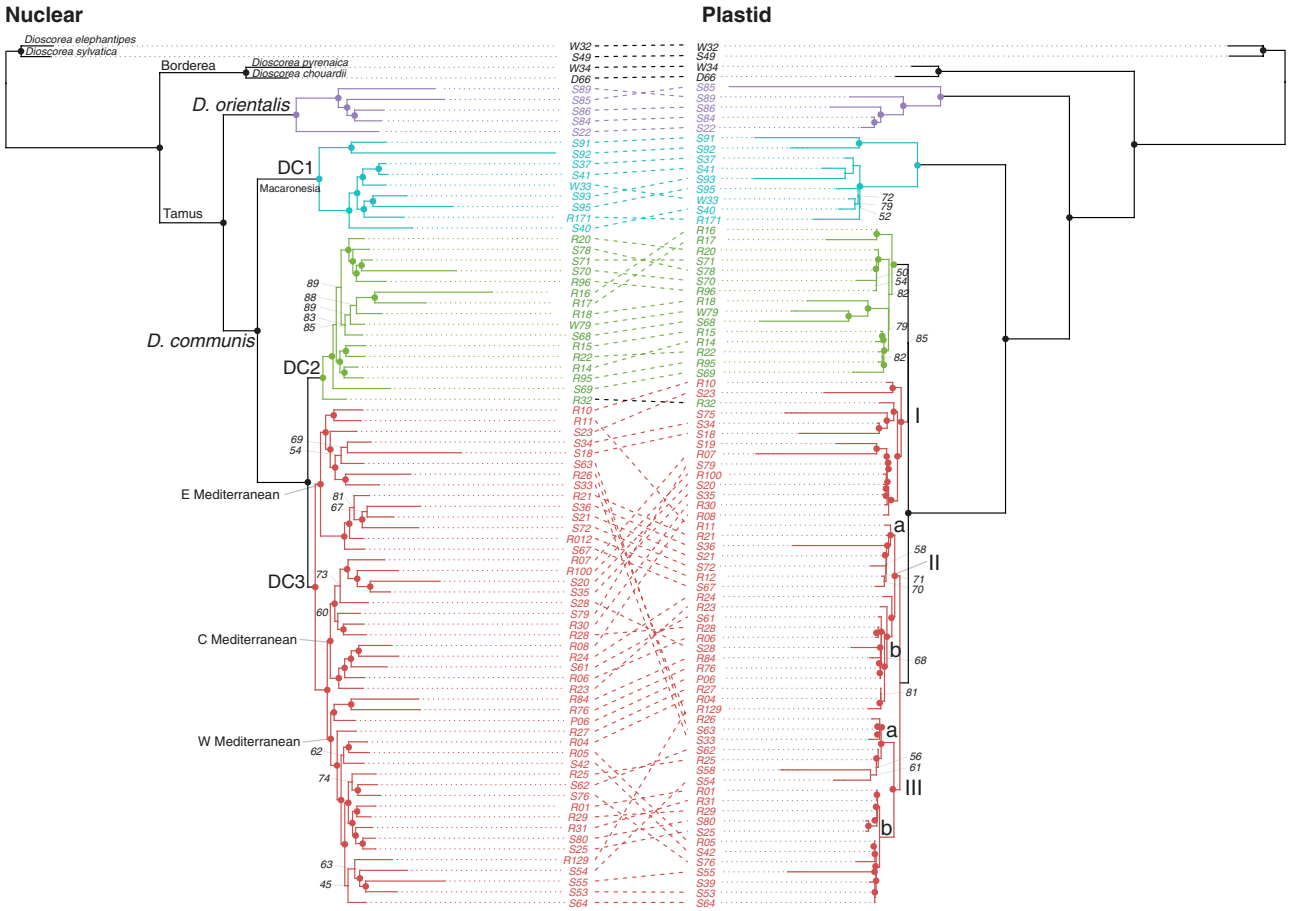

FIG. 1. Phylogenomic trees reconstructed using the concatenated and partitioned nuclear (left) and plastid (right) data obtained using Hyb-Seq for 76 samples of the Tamus clade of *Dioscorea*. Samples included are representative of its distribution range across the Mediterranean region. Filled circles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

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

An ENM analysis was performed for each of the four lineages in the Tamus clade (Fig. 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 and bio8; DC3, bio4, bio8, bio16, bio15 and bio9; *D. orientalis*, bio16, bio14, bio6, bio15, bio3 and bio8; in order of importance), and showed high values of area under the curve (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).

The group formed by DC3 samples had its highest distribution probability in several Mediterranean areas, central and southern Europe (including England and Belgium to the 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 last 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 Islands (Fig. 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 the 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,

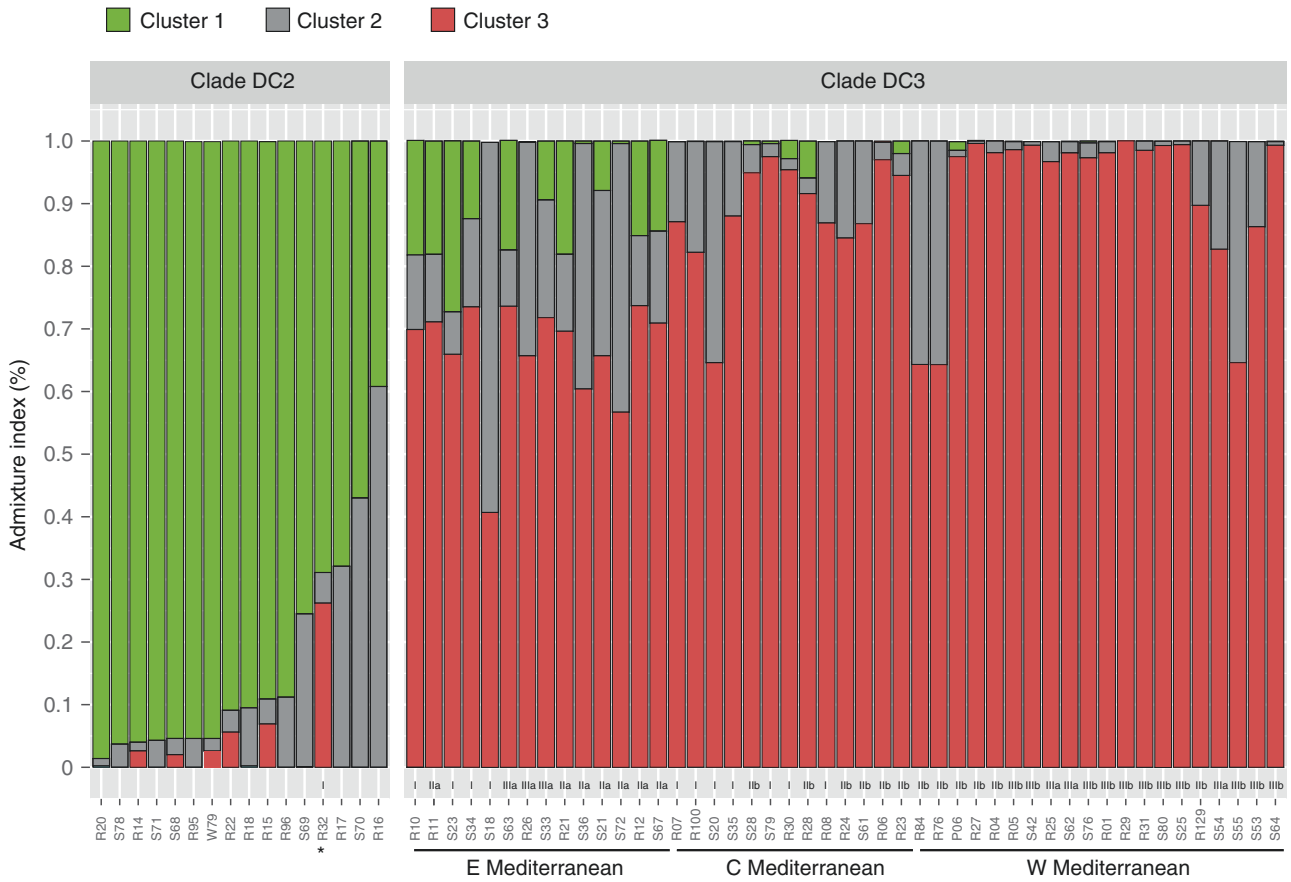

FIG. 2. Admixture proportions for $K = 3$ genetic groups obtained from genetic structure analysis of nuclear data of Mediterranean *D. communis* samples through ten replicates in Structure (see text for details). Each sample is shown by a vertical bar partitioned according to its membership in one of the K clusters, represented in red, green and blue. The samples follow the same order as in the nuclear phylogenetic tree (Fig. 1) and are organized by clades (DC2 and DC3). The plastid clades (I, II and III; Fig. 1) are labelled next to the sample codes. An asterisk (*) is 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 (Fig. 1).

$I = 0.64$) 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 the 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 (Fig. 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 Fig. S6).

Hindcast species distribution models projected to the past (Supplementary Data Fig. S7) support a long-term presence of *D. orientalis* and DC2 clades in the eastern Mediterranean region, and a prevalence of the DC3 clade in the western Mediterranean since the LIG (~120 000–140 000 years ago). The overlap between current and MH (~6000 years ago) distribution models was high for DC2 (0.907), *D. orientalis* (0.798) and DC3 (0.859), slightly lower between the MH and the LGM (~22 000 years ago) for DC2 (0.835), *D. orientalis* (0.564) and DC3 (0.719), and moderately to drastically lower between the LGM and the LIG for DC2 (0.216), *D. orientalis* (0.419) and DC3 (0.558).

Identification of morphological traits defining taxa based on lineage divergences in the Tamus clade

We explored whether differences may exist between the four clades identified in the Tamus clade of *Dioscorea* in macro- and micromorphological characteristics (Fig. 5). Only three traits (male flower pedicel length, female inflorescence length and leaf coverage) were found to have a normal distribution (Supplementary Data Table S3). An assumption of homoscedasticity was corroborated using a Levene test for these variables. In the vegetative traits analysed (Supplementary Data Table S2), we found significant differences between clades in leaf area, including leaf length, leaf width, leaf perimeter and petiole length, while leaf coverage and the main nerve length and leaf length ratio did not show significant differences between groups (Supplementary Data Table 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

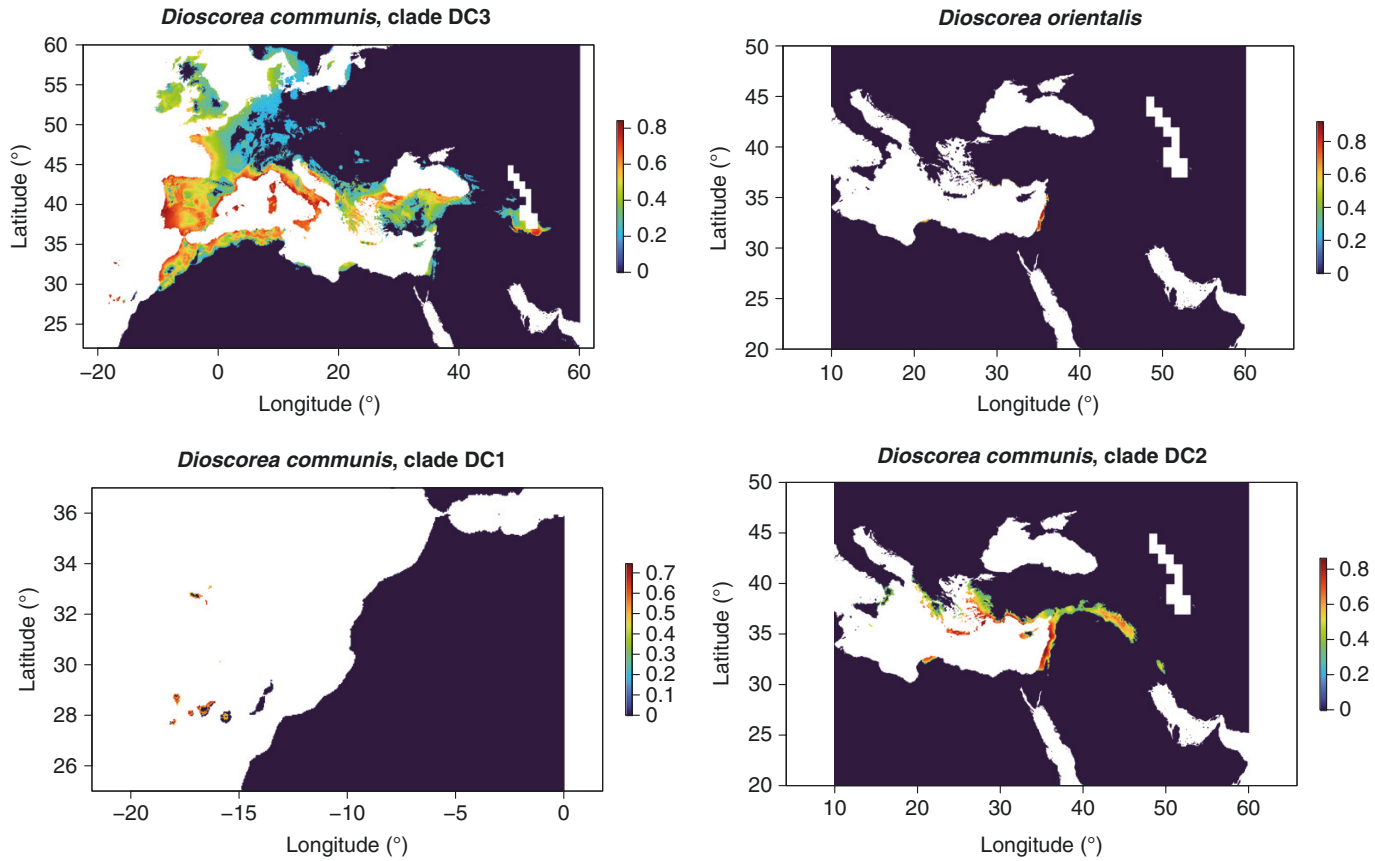

FIG. 3. Environmental niche models constructed using a current climate envelope and occurrence data for the samples of the Tamus clade of *Dioscorea* analysed in this study. The ENMs were estimated for each of the four lineages identified in the Tamus clade with background maps adjusted to reflect their 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 section). The scale bars represent the prediction of probability of occurrence.

(Supplementary Data Fig. S8): leaf length, leaf width, leaf area, leaf perimeter, petiole length, 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.

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 *D. 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; leaf pedicel up to 1.6 cm, shorter than leaf …. *D. orientalis*
1b. Flowers pedicellate in fascicles (2–4 flowers each); leaf pedicel leaf pedicel up to 10.1 cm, 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* (Thiébaut) Caddick & Wilkin = Basionym: *Tamus orientalis* Thiébaut., Bull. Soc. Bot. France 81: 119. 1934.

– Holotype: Lebanon. Between Batroun and Saïda, January 1933, *Thiébaut 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 Fig. S9) with two apertures with the long

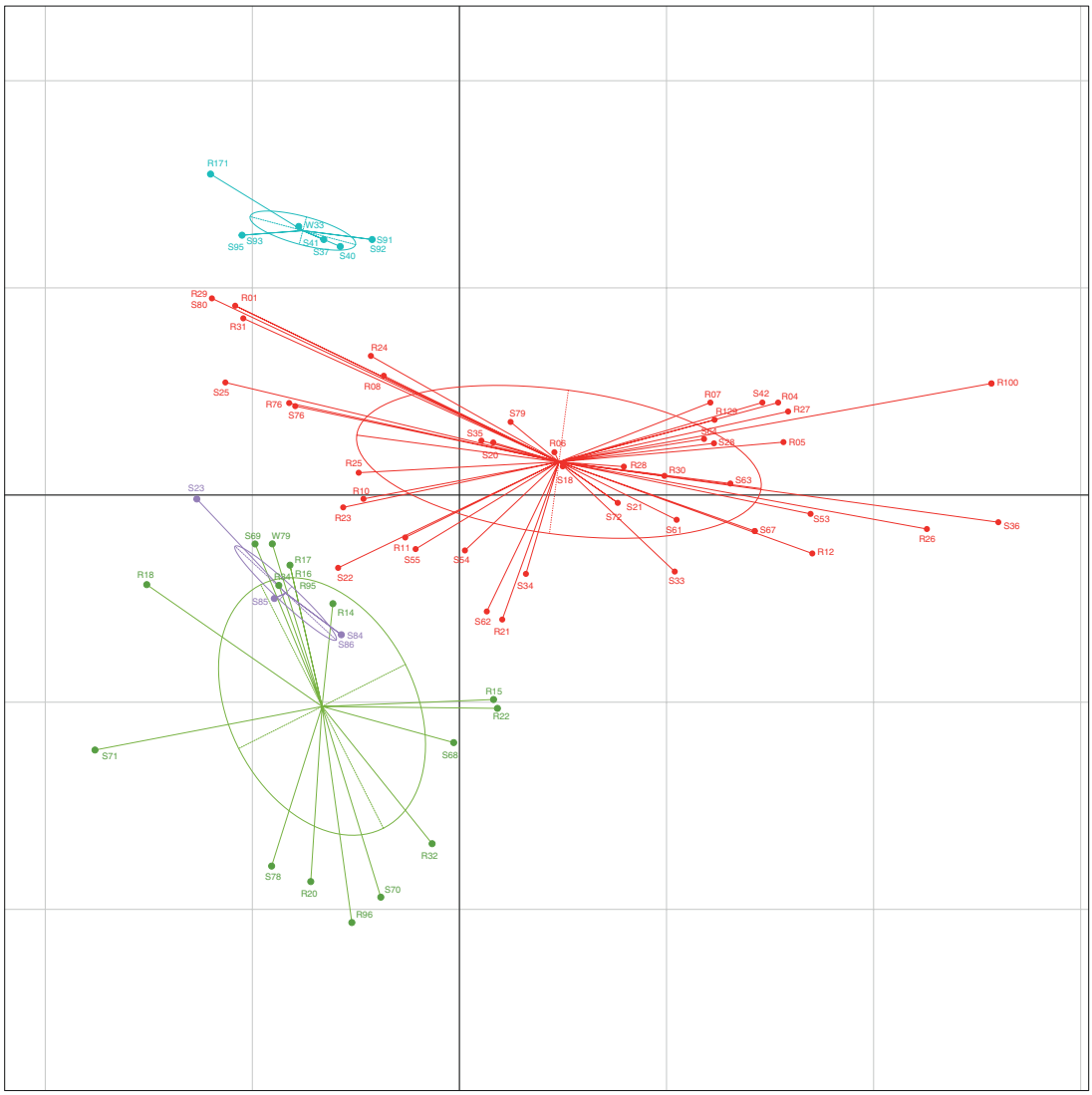

FIG. 4. Principal component analysis conducted with 19 bioclimatic variables for samples of the Tamus clade of *Dioscorea* across the Mediterranean region separated into the four main lineages (see Results section): Macaronesian DC1 clade of *D. communis* in light blue; *D. communis* clade DC2 in green;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

axis up to 40 µm and the short axis up to 30 µm. Perforated, with 1.9 perforations per micrometre and perforation size up to 0.85 µm, with spines. Eastern Mediterranean region. 2*n* = 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 Fig. S9) with two apertures with the long axis up to 36 µm and the short axis up to 18 µm. Perforated, with 3.9 perforations per micrometre and perforation size up to 0.37 µm, without spines. Macaronesia (Madeira, Gran Canaria, Tenerife, Gomera, Hierro, La Palma). 2*n* = 36, 48.

3. *Dioscorea cretica* (L.) Campos, Wilkin & Viruel, *comb. nov.* (= Clade DC2) = Basionym: *Tamus cretica* L., Sp. Pl., 1028. 1753.

FIG. 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 (as indicated in the key).

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

– Neotype (designated by Kit Tan in Mill and Tan (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 Fig. S9) with two apertures with the long axis up to 36 µm and the short axis up to 29 µm. Perforated, with 0.5 perforations per micrometre 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.

TABLE 2. *Morphological traits and variability in the four morphological groups identified in the Tamus clade of* Dioscorea*, one corresponding to* D. orientalis *(five samples) and three to the currently recognized* D. communis s.l.*: Macaronesian clade DC1 (*= D. edulis*, 13 samples), trilobed leaf Eastern Mediterranean clade (DC2) (*= D. cretica*, 12 samples) and the DC3 cordate leaf clade of* D. communis *(*= D. communis s.s.*, 46 samples) (see Fig. 1)*

		***D. communis* DC3**	***D. communis* DC2**	***D. communis* DC1**	***D. orientalis***
Leaf	Petiole length (cm)	0.24–10.1	1.7–8.8	1.6–3.88	0.6–1.6
	Shape	Cordate	Trilobed	Cordate to sagittate	Cordate
	Size length × width (mm)	Up to 180 × 150	Up to 70 × 70	Up to 125 × 110	Up to 56 × 48
	Margin	Thickened, not undulate	Thickened, not undulate	Thickened, slightly undulate	Thickened and undulate
	Consistency	Chartaceous	Chartaceous	Thinly chartaceous	Thinly coriaceous
	Secondary veins	Reticulated, usually not evident	Reticulated, usually not evident	Dark, visible but not prominent	Concolorous and prominent
Male inflorescence	Disposition	In a fascicle, up to four flowers	In a fascicle, up to four flowers	In a fascicle, up to four flowers	Solitary
	Branching	Compound	Compound	Compound	Simple
	Size (cm)	Up to 35	Up to 25	Up to 12	6.7
	Number of flowers	3–4 per fascicle, up to 73 per inflorescence	1–4 per fascicle, up to 50 per inflorescence	3–4 per fascicle, 65 per inflorescence	1–23
	Pedicel (mm)	0.5–5.4	1.04–5.17	1.4–5.4	Sessile or subsessile
Female inflorescence	Number of flowers	1–14 per inflorescence	1–6 per inflorescence	1–8 per inflorescence	1–4
	Pedicel (mm)	4.7–11.3	2.5–9.3	3.8–7.3	Sessile or subsessile
Fruit	Size length × width (mm)	7–11, 6–10	7–11, 6–10	6–12, 1–8	5.5–11, 1–4
	Shape	Globose to ellipsoid	Globose to ellipsoid	Globose to ellipsoid	Ellipsoid
Seed	Number	1–6	1–6	1–6	1–6
	Colour	Dark brown	Dark brown	Dark brown	Dark brown
Pollen	Apertures	2	2	2	2
	Long axis (μm)	31–36	31–36	28–36	35–40
	Short axis (μm)	27–30	26–29	14–18	27–30
	Perforation size (μm)	0.6–1.2	0.6–1.2	0.18–0.37	0.31–0.85
	Perforations/μm^2	0.5	0.5	3.9	1.9
	Spines	No	No	No	Yes

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

– Lectotype (designated by Ferrer-Gallego and 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 and 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, 7–12 mm, reddish-orange. Pollen grains (Supplementary Data Fig. S9) with two apertures with the long axis up to 36 μm and the short axis up to 30 μm. Perforated, with 0.5 perforations per micrometre 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.

The use of infraspecific ranks in plants has been discussed widely in the literature (e.g. Hamilton and Reichard 1992; Vogel Ely *et al.*, 2018; Huang *et al.*, 2020). Subspecies in plants have been defined as non-overlapping geographic populations with morphological differences (Patten and Remsen, 2017) or, considering additional sources, defined as taxonomic units supported by genetic markers, statistically distinguishable based on morphology and with geographic, ecological and/or reproductive isolation (Hardion *et al.*, 2017).

Here, we have used 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 (Fig. 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 s.s.*, *D. edulis* and *D. cretica* (see Results section; Fig. 6). The three taxonomic units identified in our study within *D. communis s.l.* are supported by morphological, cytological, ecological and evolutionary differentiation, and we therefore propose maintaining the species rank for them.

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; Hassemer *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 large number of plant systematics and evolutionary studies due to its versatility in successfully sequencing 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 (Brewer *et al.*, 2019; Viruel *et al.*, 2019). 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* (Fig. 1). By sequencing 76 samples of the Tamus clade, the phylogenomic and genetic clustering approaches revealed extensive infraspecific variability in *D. communis s.l.*, clearly dividing it into three genetic groups, each showing a distinct geographic distribution across the Mediterranean and western Europe (Fig. 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.* Application of HTS methodologies to herbarium material allowed us to recognize the species rank of 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 to 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, and has been frequently reported in several *Dioscorea* species (Viruel *et al.*, 2008), although defining the 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 the ratio is >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, the *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 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*,

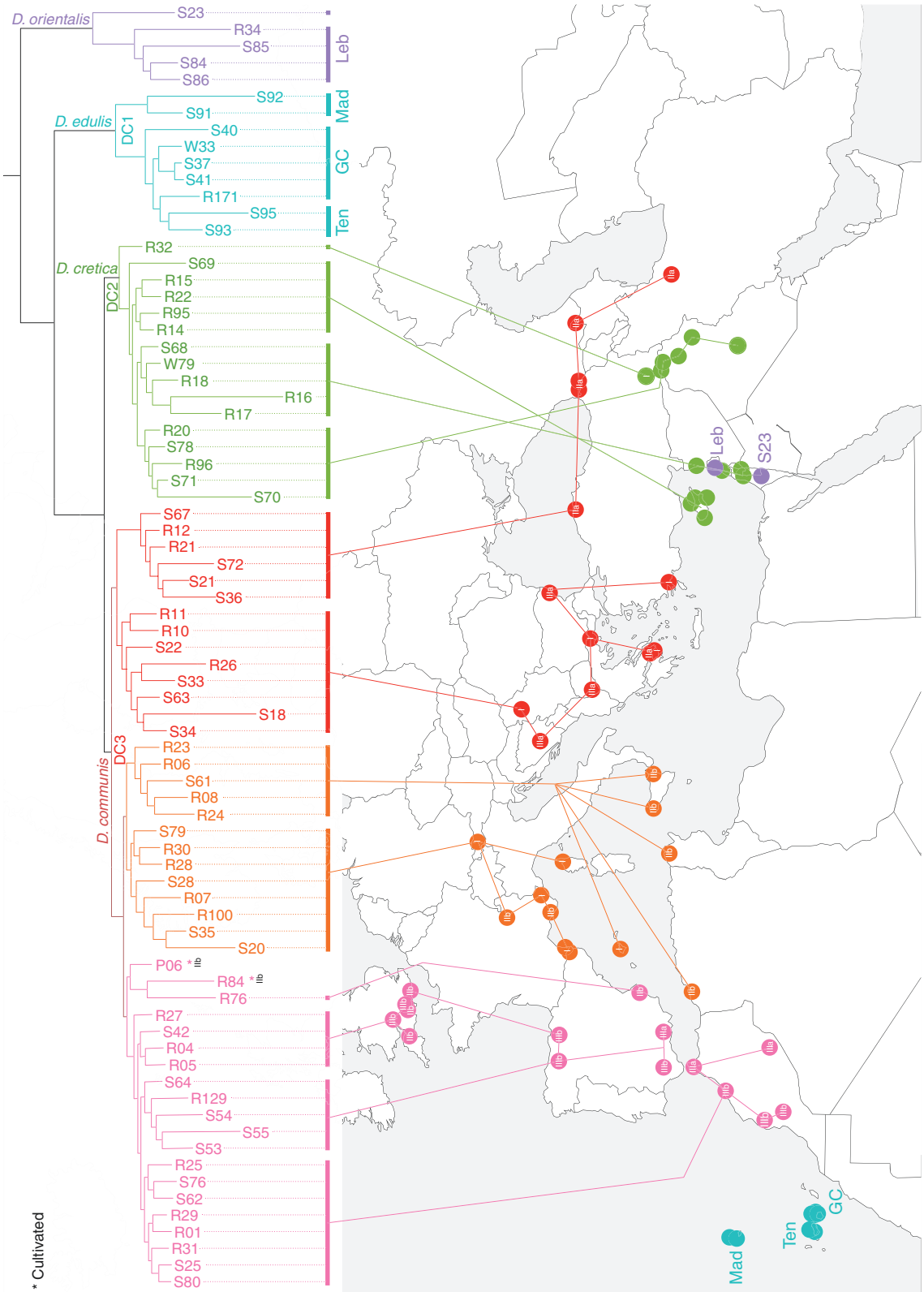

FIG. 6. Geographic distribution of the 76 samples of the Tamus clade of *Dioscorea* coded according to their respective lineage in the phylogenetic tree based on concatenated nuclear data (Fig. 1). *Dioscorea cretica* (DC2) in green, and pink, orange, and red for western European, central Mediterranean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

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 have 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 ~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 (~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 Fig. S5). 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 (Goñi Martínez and Guzmán Otano, 2013), and *D. pyrenaica*, with a slightly wider distribution in the central Pyrenees and Pre-Pyrenees (Segarra and Catalán, 2005; Catalán *et al.*, 2006; Segarra-Moragues and Catalán, 2008; García *et al.*, 2012). A previous study (Viruel *et al.*, 2016) estimated a split between these two species during the early Pliocene (~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* ~16.0 13.5 MY, and the divergence of *D. communis s.s.* and *D. cretica* ~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*; 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 have started 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 biogeographic 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 hotspots (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 in 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 the phylogeographic patterns of several Mediterranean plants, leading to disjunct distributions in some cases, such as in *Microcnemum* (Amaranthaceae) and *Mandragora* (Solanaceae) (Kadereit and Yaprak, 2008; Volis *et al.*, 2018), or by differentiating morphotypes that later hybridized in intermediate zones (e.g. *Quercus ilex*; 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* (Hampe *et al.*, 2003). The birds that consume berries produced by species of the Tamus clade, mainly blackbirds (*Turdus merula*), robins (*Erithacus rubecula*) and blackcaps (*Sylvia communis*) (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 the introgression between *D. communis* and *D. cretica*, 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 (Figs 2 and 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 (Fig. 2). Five out of 16 samples studied of *D. cretica* exhibited an admixture index <20 % with *D. communis s.s.*, and all individuals of *D. communis* belonging to the eastern clade of *D. communis s.s.* showed an admixture index <20 % with *D. cretica* (Fig. 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, Fig. 1). These results are congruent with their potential distribution in disjunct refugia followed by secondary contact through recolonization, and by maintaining some capacity for interspecific gene flow between closely related species (Viruel *et al.*, 2021). The topological incongruencies found between the nuclear and plastid phylogenetic trees, indicative of plastid capture events (Fig. 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 (Fig. 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*.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. 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. Table S2: traits used for morphometric analyses of the Tamus clade of *Dioscorea* samples. Table S3: results of the Lilliefors test, Levene test and the subsequent ANOVA or Kruskal–Wallis test for the selected variables analysed in the Tamus clade of *Dioscorea* samples. Figure S1: dendrograms constructed from correlation values between bioclimatic variables for the samples belonging to the *D. communis* and *D. orientalis* clades. Figure S2: phylogenomic trees reconstructed for 76 samples of the Tamus clade of *Dioscorea* using ASTRAL-III, SVDquartets and the haplotype network. Figure S3: results obtained in Structure Harvester based on genetic structure analysis of *D. communis* nuclear data though 10 replicates of Structure for each *K*. Figure S4: genetic structure analysis of nuclear SNP data of the studied Mediterranean *D. communis* samples with Structure. Figure S5: divergence time estimations and chronograms of the Tamus clade of *Dioscorea* and outgroup based on nuclear sequence data. Figure S6: boxplots representing the differences between the four main clades. Figure S7: past projections of the ENMs constructed for the four Tamus clades of *Dioscorea*. Figure S8. Pairwise Pearson correlation matrix calculated between the morphological variables measured to distinguish between Tamus clade taxa. Figure S9: scanning electron microscope images of the pollen grains of *D. communis* DC3 clade, *D. cretica*, *D. edulis* and *D. orientalis*.

FUNDING

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.

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.

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.

Al-Shehbaz IA, **Schubert BG. 1989**. The Dioscoreaceae in the southeastern United States. *Journal of the Arnold Arboretum* **70**:57–95.

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 Scientarum Hungaricae* **5**: 1–37.

Asiedu R, **Sartie A. 2010**. Crops that feed the world 1. Yams. *Food Security* **2**: 305–315. doi:10.1007/s12571-010-0085-0.

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*. Cambridge: BirdLife International.

Caddick LR, **Wilkin P**, **Rudall PJ**, **Hedderson TAJ**, **Chase MW. 2002**. Yams reclassified: a recircumscription of Dioscoreaceae and Dioscoreales. *Taxon* **51**: 103–114. doi: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. In: Proceedings of the 16th International Parallel and Distributed Processing Symposium (IPDPS 2002), 15–19 April 2002, Fort Lauderdale, Florida. Los Alamitos, CA: Institute of Electrical and Electronics Engineers.

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.

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**: 2141–2148. doi:10.1186/1471-2148-7-214.

Earl DA, **vonHoldt BM. 2012**. 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**: 258–266. doi:10.11646/phytotaxa.260.3.5.

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, **Záveská E**, **Gamisch A**, **Moser T**, **Schönswetter P**; **The STEPPE Consortium**. **2019**. Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* 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. 2013**. *Borderea chouardii*. The IUCN Red List of Threatened Species 2013: e.T162110A5540643. doi:10.2305/IUCN.UK.2011-1.RLTS.T162110A5540643.en.

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.

Hamilton CW, **Reichard SH. 1992**. Current practice in the use of subspecies, variety, and forma in the classification of wild plants. *Taxon* **41**: 485–498. doi:10.2307/1222819.

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.

Hardion L, **Verlaque R**, **Vorontsova MS**, ***et al***. **2017**. Does infraspecific taxonomy match species evolutionary history? A phylogeographic study of *Arundo formosana* (Poaceae). *Botanical Journal of the Linnean Society* **183**: 236–249. doi:10.1093/botlinnean/bow006.

Hassemer G, **Bruun-Lund S**, **Shipunov AB**, **Briggs BG**, **Meudt HM**, **Rønsted N. 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.

Herrera CM. 1984. A study of avian frugivores, bird-dispersed plants, and their interaction in Mediterranean scrublands. *Ecological Monographs* **54**: 12–23. 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* **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.

Huang Y, **Morrison GR**, **Bresfold A**, ***et al***. **2020**. Subspecies differentiation in an enigmatic chaparral shrub species. *American Journal of Botany* **107**: 923–940. doi:10.1002/ajb2.1496.

Jarvis CE. 2007. *Order out of chaos. Linnaean plant names and their types*. London: Linnean Society of London with the Natural History Museum.

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* **65**: 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.

Katoh MA, **Darriba D**, **Flouri T**, ***et al***. **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.

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. *Systematics and the origin of species*. New York: Columbia University Press.

Médail F, **Diadema K. 2009**. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36**: 1333–1345. doi:10.1111/j.1365-2699.2008.02051.x.

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.

Mill RR, **Tan K. 1984**. DIOSCOREACEAE. In **Davis PH** (ed.) *Flora of Turkey and the East Aegean Islands*, Vol. **8**. Edinburgh: Edinburgh University Press, 552–554.

Mitchell N, **Campbell LG**, **Ahern JR**, **Paine KC**, **Giroldo AB**, **Whitney KD. 2019**. Correlates of hybridization in plants. *Evolution Letters* **3**: 570–585. doi: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.

Patten MA, **Remsen JV Jr**. **2017**. Complementary roles of phenotype and genotype in subspecies delimitation. *Journal of Heredity* **108**: 462–464. doi:10.1093/jhered/esx013.

Phillips SJ, **Anderson RP**, **Dudík M**, ***et al***. **2017**. Opening the black box: an open-source release of Maxent. *Ecography* **40**:887–893. doi:10.1111/ecog.03049.

POWO. 2022. *Plants of the world online*. Richmond: 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 CA, **Rasband WS**, **Eliceiri KW. 2012**. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**:671–675. doi:10.1038/nmeth.208

Segarra JG, **Catalán P. 2005**. Borderea Miègeville. In: **Aedo C**, **Herrero A.** eds. *Flora Iberica*, Vol. **21**. Madrid: Real Jardín Botánico de Madrid, CSIC, 11–14.

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. doi:10.1080/17550870802349757

Segarra-Moragues JG, **Palop-Esteban M**, **González-Candelas F**, **Catalán P. 2003**. Characterization of ten trinucleotide microsatellite loci in the critically endangered Pyrenean yam Borderea chouardii (Dioscoreaceae). *Molecular Ecology Notes* **3**:265–267. doi:10.1046/j.1471-8286.2003.00422.x

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.

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**: 63636–63650. doi:10.1111/nph.15312.

Viruel J, **Segarra-Moragues JG**, **Pérez-Collazos E**, **Villar L**, **Catalán P. 2008**. The diploid nature of the Chilean *Epipetrum* and a new base number in the Dioscoreaceae. *New Zealand Journal of Botany* **46**: 327–339. doi:10.1080/00288250809509771.

Viruel J, **Segarra-Moragues JG**, **Pérez-Collazos E**, **Villar L**, **Catalán P. 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 Palaearctic and their subsequent Oligocene-Miocene diversification. *Journal of Biogeography* **43**: 750–762. doi:10.1111/jbi.12678.

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. doi:10.1093/botlinnean/boy013.

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, **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.

Vogel Ely C, **Ott Andrade B**, **Vieira Iganci JR**, **Iob Boldrini I. 2018**. Integrative taxonomy improves delimitation in *Hypericum* subspecies. *Perspectives in Plant Ecology, Evolution and Systematics* **34**: 68–76. doi:10.1016/j.ppees.2018.08.005.

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.

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 USA* **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**, **Otto SP**, **Mayrose I. 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.