**Distributions and Conservation Status of Carrot Wild Relatives in Tunisia: A Case Study in the Western Mediterranean Basin**

Najla Mezghani, Colin K. Khoury,* Daniel Carver, Harold A. Achicanoy, Philipp Simon, Fernando Martínez Flores, and David Spooner

**ABSTRACT**
Crop wild relatives, the wild progenitors and closely related cousins of cultivated plant species, are sources of valuable genetic resources for crop improvement. Persisting gaps in knowledge of taxonomy, distributions, and characterization for traits of interest constrain their expanded use in plant breeding and likewise negatively affect ex situ (in genebanks) and in situ (in natural habitats) conservation planning. We compile the state of knowledge on the taxonomy and distributions of the wild relatives of carrot (genus *Daucus* L.) natively occurring within Tunisia—a hotspot of diversity of the genus, containing 13 taxa (27% of species worldwide). We use ecogeographic information to characterize their potential adaptations to abiotic stresses of interest in crop breeding and assess their ex situ and in situ conservation status. We find substantial ecogeographic variation both across taxa and between populations within taxa, with regard to adaptation to high temperatures, low precipitation, and other traits of potential interest. We categorize three of the taxa high priority for further conservation both ex situ and in situ, five medium priority, and five low priority, with none currently considered sufficiently conserved. Geographic hotspots for species diversity, especially in the northern coastal areas, represent particularly high value regions for efficient further collecting for ex situ conservation and for in situ protection in Tunisia.

**Crop wild relatives**—the wild progenitors and close relatives of cultivated plant species—have been the sources of a wide range of beneficial agronomic and quality traits used in...
crop improvement (Dempewolf et al., 2017; Hajjar and Hodgkin, 2007). As some populations of these wild plants are adapted to extreme climates, challenging soils, and significant pests and diseases, they have substantial potential to contribute to crop breeding for emerging and projected future agricultural challenges (Dempewolf et al., 2014).

Persisting gaps in knowledge of these plant genetic resources constrain their use in plant breeding, including information on taxonomy, crossability, and gene pool classifications, distributions, and characterization for traits of interest (Dempewolf et al., 2017; Miller and Khoury, 2018). Such foundational knowledge gaps also affect conservation efforts for the taxa, which are necessary to protect vulnerable populations from habitat destruction, invasive species, climate change, and overharvesting in situ (Bellon et al., 2017; Brummitt et al., 2015; Jarvis et al., 2008a), as well as to ensure that genetic resources are safeguarded for the long term and are available for research in ex situ repositories (Castañeda–Álvarez et al., 2016; Gepts, 2006).

Cultivated carrot [Daucus carota L. subsp. sativus (Hoffm.) Schübl. & G. Marten] is the most widely grown member of the Apiaceae and among the 10 most important vegetable crops, with global production on 1.2 million ha, and a crop value of US$13.9 billion (2011–2015 average; FAO, 2019). Beyond its economic value, carrot is the richest source of provitamin A carotenoids in the US diet, accounting for 45 to 50% of total intake (Simon et al., 2009). Micronutrient intake values for individual vegetables are not available for the human population as a whole, but per capita consumption of carrots in the United States is similar to global per capita consumption (Simon, 2019) indicating the important role of carrots as a source of dietary vitamin A globally.

Wild carrot (D. carota L. subsp. carota) has played an important role in carrot breeding, as efficient hybrid carrot seed production depends on cytoplasmic male sterility, a maternally inherited trait that prevents the production of viable pollen without affecting female fertility (Banga et al., 1964). The predominant source of sterility used today is petaloid cytoplasm discovered in two wild carrot sources, first in New England by Henry Munger in 1953, then in Wisconsin by Teddy Morelock in 1970 (reviewed by Simon, 2000). Hybrid cultivars today account for nearly all of the large-scale carrot production industry in North America, Europe, and Japan, so most of the carrots consumed in these markets have wild carrot cytoplasm. Additional sources of male sterile cytoplasm have been found in European wild carrot (Nothnagel et al., 2000).

Other significant traits found in wild carrot include resistance to leaf blight [caused by Alternaria dauci (J.G. Kühn) J.W. Groves & Skolko] (Simon, 2000; Tas, 2016) and powdery mildew [caused by Erysiphe heradei DC.] (Bonnet, 1983), and high micronutrient content (Luby et al., 2014). Salinity tolerance during germination was found in wild carrot seed from Pakistan and Turkey (Bolton and Simon, 2019). Heat tolerance during germination was found in wild carrot seed from Pakistan, Portugal, Tunisia, and Turkey (Bolton et al., 2019).

Traits of interest found in other carrot wild relative taxa include carrot fly resistance in D. carota subsp. capitillifolius (Gilli) Arbizu and D. carota subsp. gummifer (Syme) Hook. f. (Ellis, 1999; Ellis et al., 1993; Grzebelus et al., 2011; Simon 2000), cytoplasmic male sterility in D. carota subsp. gummifer, and salinity tolerance in D. carota subsp. halophilus (Brot.) A. Pujadas (Tavares et al., 2010).

The Apiaceae (Umbelliferae) family, with 466 genera and 3820 species (Plunkett et al., 2019) is one of the largest families of seed plants. It is well supported as a monophyletic family and is well defined by a suite of morphological characters. However, generic circumscriptions within the family have been the subject of long dispute, including for the genus Daucus (Plunkett and Downie, 1999). A genus-level treatment of Daucus by Säenz Lain (1981) used morphological and anatomical data and recognized 20 species. Rubatzky et al. (1999) later estimated 25 species. The phylogenetic relationships among the species of Daucus and close relatives in the Apiioideae have been clarified by a series of molecular studies using DNA sequences of the plastid and nuclear genes as summarized in Spooner (2019). Of these recent studies, that by Banasiak et al. (2016) is taxonomically most significant in using a wide range of Daucus in-groups and out-groups analyzed with DNA sequences from nuclear ribosomal internal transcribed spacer (ITS) and three plastid markers. This study redefined and expanded the genus Daucus to include the following genera and species into its synonymy: Agrochartis Hochst. (four species), Melanosemium Hoffm. (one species), Monzia Lowe (one species), Pachyctenos Maire & Pamp. (one species), Pseudorlaya (Murb.) Murb. (two species), Rouya Coincy (one species), Tornabenea Parl. (six species), Athamanta della-cellae E.A. Durand & Barratte, and Cryptotaenia elegans Webb ex Bolle. Banasiak et al. (2016) made the relevant nomenclatural transfers into Daucus and following this classification, the genus contains ~40 species as well as an undetermined number of subspecies in D. carota (Spooner et al., 2014). Relative to the present study of Tunisia, Daucus now includes the former Pseudorlaya pumilia (L.) Grande [now Daucus pumilus (L.) Hoffmans & Link] and the former Rouya polygama (Desf.) Coincy (now Daucus rouyi Spalik & Reduron). Despite its small size, Tunisia contains 33% (13 of the 40) presently recognized species of Daucus (Table 1).

Recently, high-throughput, “next generation” DNA sequencing (NGS) has been used to infer phylogenetic relationships in Daucus. The utility of NGS is markedly improved when a high-quality whole genome “reference” sequence is available that serves as a heterologous template.
to guide mapping of sequences of related germplasm. Such whole-genome reference sequences are available in carrot for the plastid genome (Ruhlman et al., 2006) and for the plastid and nuclear genome (Iorizzo et al., 2016). Arbizu et al. (2014b) used these to identify 94 nuclear orthologs (single- to low-copy DNA regions) to produce a highly resolved phylogenetic tree. Spooner et al. (2017) used whole-plastid DNA sequences in Daucus and compared the results with prior phylogenetic results using plastid and nuclear DNA sequences, and like the nuclear tree of Arbizu et al. (2014b), the trees of the entire dataset were highly resolved, with 100% bootstrap support for most of the external and many of the internal clades.

All of these studies divide the genus Daucus into two well-supported clades. Relative to the Tunisian species they place D. aureus Desf., all subspecies of D. carota, D. crinitus Steinh., D. gracilis Steinh., D. muricatus (L.) L., D. pumilus, D. rouyi, D. setifolius Desf., D. syricus Murb., D. tenuisectus Coss. ex Batt., and D. virgatus (Poir.) Maire in one clade, with the members of the other clade not occurring in the country (D. reboudii Coss. was not examined). Tunisian species of Daucus have chromosome numbers ranging from 2n = 16 to 2n = 22 (Table 1). All taxa with diploid numbers of 2n = 18 (all subspecies of D. carota, D. syricus, and possibly D. gracilis and D. virgatus) form a subclade and would be expected to be easiest to use in carrot breeding programs.

Daucus carota presents an unresolved problem in the genus. It is highly polymorphic, and >60 infraspecific taxa have been described over the years, making it the most difficult species group in the Apiaceae (Small, 1978). Morphological studies (Arbizu et al., 2014a; Mezghani et al., 2014; Small, 1978; Spooner et al., 2014) do not distinguish most of the subspecies of D. carota. Iorizzo et al. (2013) used 3326 single nucleotide polymorphisms (SNPs) to study the genetic structure and domestication of carrot and found a clear separation between wild (subsp. carota) and cultivated (subsp. sativus) accessions of D. carota but did not separate the wild subspecies and varieties into traditionally named taxa. Taxonomic keys and descriptions in various floras lack consensus both about the number of infraspecific taxa of D. carota and the characters best distinguishing them. For instance, 11 wild subspecies were recognized by Heywood (1968a, 1968b), five were recognized by Sáenz Lain (1981), a different five were recognized by Arenas and García-Martín (1993), and Pujadas Salvà (2002) proposed nine subspecies for the Iberian Peninsula plus the Balearic Islands.

Molecular investigations are in progress to study the natural taxa in D. carota. For example, Arbizu et al. (2016a) used genotyping-by-sequencing (GBS), a NGS method, to examine the subspecies of D. carota. The results divided accessions of D. carota into geographic groups that did not consistently correspond to traditional taxonomic subdivisions. Mezghani et al. (2018) used GBS to examine the subspecies of D. carota, and found that the Tunisian members of D. carota subsp. gummifer had separate origins from other collections of this subspecies, with the Tunisian accessions grouping with those in Italy or its surrounding islands. Currently we are expanding our studies of the subspecies of D. carota by including many more accessions to help resolve the persisting taxonomic questions.

Despite the extensive historical taxonomic and breeding research, as well as recent prioritization in international conservation initiatives (Dempewolf et al., 2014), there remain questions of taxonomy, distributions, crossability, and trait characterization for the wild relatives of carrot, particularly with regard to populations occurring within the putative primary region of diversity—the Mediterranean basin. At the same time, preliminary analyses indicate that carrot wild relatives may be poorly represented in genebanks (Castañeda-Álvarez et al., 2016) and in protected areas (Khouri et al., 2019a), highlighting the urgency of addressing foundational information gaps so as to guide near-future conservation, as well as crop improvement efforts.

Here, we compile the state of knowledge on the taxonomy and distributions of the wild relatives of carrot occurring within Tunisia, a targeted area within the western Mediterranean basin where substantial recent collecting has been performed by the authors. We model the potential distributions across the western Mediterranean of all 13 taxa currently known to occur in Tunisia.
(Table 2). We use ecogeographic information to discuss the potential adaptations of the taxa to abiotic stresses of interest for crop breeding. We finally assess the conservation status of Tunisian populations of the taxa, both in genebanks and botanic garden living plant collections (ex situ), and in protected areas (in situ).

MATERIALS AND METHODS

To comprehend the native ranges of Tunisian carrot wild relatives, estimate their potential for contribution to crop improvement, and assess their conservation status, the following steps were followed: (i) compiling and processing of occurrence information, (ii) distribution modeling, (iii) ecogeographic characterization, and (iv) conservation gap analysis.

Occurrence Information

We gathered occurrence data for all records of *Daucus*, *Pseudorouya*, and *Rouya* from the Global Biodiversity Information Facility (GBIF, 2018a, 2018b), the Global Crop Wild Relative Occurrence Database (Global Crop Diversity Trust, 2018b), and from the authors’ own botanical explorations. We compiled genebank conservation occurrence data from the Genesys plant genetic resources portal (Global Crop Diversity Trust, 2018a), the USDA National Plant Germplasm System (Germplasm Resource Information Network [GRIN] Global) (USDA-ARS NPGS, 2018a), the National GeneBank of Tunisia (NGBT) (NGBT, 2018) and the FAO World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (WIEWS) (FAO, 2018). Duplicates in the databases were removed with preference for original and most recently updated data providers (e.g., USDA NPGS dataset instead of equivalent USDA records in Genesys or in the Global Crop Wild Relative Occurrence Database).

Taxonomic names were standardized against USDA GRIN Global Taxonomy (USDA-ARS NPGS, 2018b). Cultivated taxa, records listed in sample status fields as other than wild, weedy, or null (e.g., landrace, improved, breeding material, cultivated, etc.), and fossil specimens (GBIF dataset) were removed. In preparation for the conservation analysis, we classified each record according to whether it would be used only as an input into distribution modeling (labeled H, as most of these records are from herbaria), or whether it would also be considered a “site where collected” location of an existing plant genebank or botanic garden conservation accession (labeled G, as most records are from genebanks). For GBIF, this classification was performed by filtering the “basis of record” field, assigning “living specimen” records as G, with the other categories (observation, literature, preserved specimen, human observation, machine observation, material sample, and unknown) assigned as H. All records in GRIN Global (note all occurrences were listed as “active” status), Genesys, and WIEWS were assigned G, and records from the Global Crop Wild Relative Database had already been categorized accordingly.

To review the occurrence data in preparation for distribution modeling, we uploaded records with coordinates for each taxon to an interactive mapping platform (Google, 2019). We then corrected or removed occurrences occurring in bodies of water or in clearly incorrect locations. Records were further constrained to the distribution modeling study area (western Tunisia).
Mediterranean, defined as Mediterranean Sea-facing Morocco, Algeria, Tunisia, and Libya in the south, and Portugal, Spain, France, Italy, Bosnia and Herzegovina, Montenegro, Albania, Greece, and Cyprus to the north). Due to the large number of occurrences for one taxon (D. carota subsp. carota), these occurrences were further proportionally thinned by country to less than a total of 2000, to facilitate distribution modeling, using the nstrata function in the R package SamplingUtil R package, as per Khoury et al. (2019a). Refined occurrence data were extracted from the interactive mapping platform for use in the distribution modeling and conservation analyses. The final occurrence dataset is available in Supplemental Dataset S1.

**Distribution Modeling**

We used the maximum entropy (MaxEnt) algorithm (Phillips et al., 2017) accessed through the R statistical package dismo (Hijmans et al., 2017) to model the potential distributions of the taxa across the western Mediterranean, following processes outlined in Khoury et al. (2019a). We modeled the taxa at the western Mediterranean scale so that sufficient occurrence data could mobilized to produce high-quality distribution models, especially important for the rarer Tunisian species with few occurrences in the country.

We assembled 26 ecogeographic predictors (Supplemental Table S1). These included 19 bioclimatic variables, plus solar radiation, water vapor pressure, and wind speed, derived from WorldClim 2 (Fick and Hijmans, 2017). Altitude was compiled from the CGIAR-Consortium for Spatial Information (CSI) dataset based on the NASA Shuttle Radar Topography Mission (STRM) data (Jarvis et al., 2008b). Variables for slope and aspect were also incorporated, calculated from the altitude dataset using the terrain function in R package raster (Hijmans, 2017). All ecogeographic predictors were processed at a spatial resolution of 2.5 arcminutes (~5 km² at the equator).

Ecogeographic variables were selected per taxon using the R package VSURF (Genuer et al., 2018). All variables that made no measurable impact on model performance were removed and the remaining variables were ranked in order of importance. Following the most important predictor, variables that were correlated with a Pearson correlation coefficient of >0.7 were removed. This process was performed for the top five predictor variables, with the remaining variables used in the modeling process.

The number of comparative background points (pseudo-absences) were defined per taxon in proportion to the total area of the spatial background, which was calculated based on pertinent ecoregion boundaries (i.e., the ecoregions defined in Olson et al., 2001; available from Khoury et al., 2019b) wherein occurrence data fell within the native countries of the taxa as listed in USDA GRIN Global Taxonomy (USDA-ARS NPGS, 2018b), with a maximum of 5000 pseudo-absences per taxon. Pseudo-absence points that fell within the same cell as a presence point were removed.

For each species, we calculated the modeled distribution as the median of 10 MaxEnt model replicates (K = 10), using linear, quadratic, hinge and product features, with a regularization parameter $\beta = 1.0$. Following previous gap analysis studies (Khoury et al., 2019a; Ramírez-Villegas et al., 2010), we evaluated the MaxEnt model output using the area under the receiver operating characteristic curve (AUC), the standard deviation of the AUC across replicates (STAUC), the proportion of the potential distribution model with a standard deviation (of the $K = 10$ replicates) above 0.15 (ASD15), and the calibrated area under the receiver operating characteristic curve (cAUC). A robust model as per the previous studies required AUC $\geq 0.7$, STAUC $< 0.15$, ASD15 $\leq 10\%$, and cAUC $\geq 0.4$. Those models that did not pass these criteria were individually evaluated for quality of fit based on the authors’ field experiences. Distribution models were thresholded using the maximum sum of sensitivity and specificity (Liu et al., 2005, 2013) and clipped to the extent of the pertinent ecoregion boundaries (Olson et al., 2001).

**Ecogeographic Characterization**

Ecogeographic predictor information, at a resolution of 30 arcseconds (~1 km² at the equator) for 23 pertinent variables from the WorldClim 2 and CGIAR-CSI datasets were extracted for all occurrence data with coordinates, for all taxa (Supplemental Table S1). These data were used to characterize taxa with regard to their ecogeographic niches for each variable. We also assessed the representation of these niches in ex situ conservation by comparing the distributions of G points for each taxon within the full spread of its occurrences.

**Conservation Gap Analysis**

We assessed the degree of representation of Tunisian populations of each taxon in conservation systems, both ex situ and in situ, following methods outlined in Khoury et al. (2019a). To do so, we evaluated the modeled distributions and occurrence data found within the boundaries of the country. Ex situ, three scores were calculated. The sampling representativeness score (SRS) provides a general indication of the completeness of genebank and botanic garden conservation collections from Tunisia for each taxon, comparing the total count of germplasm accessions (G) collected in Tunisia and available in such repositories against the total count of reference (H) records sampled in Tunisia, with an ideal ratio of 1:1. Unique among the conservation metrics, this score makes use of all compiled reference and germplasm records in the country, regardless of whether they possess geographical coordinates. In this and all other metrics, SRS was bounded between 0 and 100, with 0 representing an extremely poor state of conservation, and 100 comprehensive (complete) conservation. If no G or H records existed, taxa were automatically considered to be of high priority for further conservation action due to the lack of information, and assigned a value of 0.

The geographical representativeness score ex situ (GRSex) is a geographic measurement of the proportion of the range of the taxon in the country that can be considered to be conserved ex situ. Buffers (“CA50”) of 0.5° (~50-km radius) were created around each G collection point in Tunisia to estimate geographic areas already collected within the distribution models. Comprehensive conservation under this metric was considered to have been accomplished when the buffered areas covered the entire distribution model in the country.

The ecological representativeness score ex situ (ERSex) is an ecological measurement of the proportion of the range of the taxon in the country that can be considered to be conserved in ex situ repositories. The ERSex compares the ecoregional diversity in Tunisia encompassed in ex situ conservation
repositories to the diversity throughout the distribution models in the country, considering comprehensive conservation to have been accomplished only when every ecoregion potentially inhabited by a taxon was included at least once within the set of CA50 buffered areas. The layer used for estimating the ERSex contained 867 distinct terrestrial ecoregions worldwide (Olson et al., 2001). A final conservation score for ex situ (FCSex) was derived by calculating the average of the three individual ex situ conservation metrics.

For the analysis of the state of in situ conservation in Tunisia, two metrics were calculated based on the extent of representation of the range of each taxon within officially recognized protected areas in the country. We used the World Database of Protected Areas (WDPA) (IUCN, 2019), selecting terrestrial and coastal reserves marked as designated, inscribed, or established. The geographical representativeness score in situ (GRSin) is a geographic measurement of the proportion of the range of a taxon in the country that can be considered to be conserved in protected areas. The calculation compares the area (in km$^2$) of the distribution model located within protected areas in Tunisia vs. the total area of the distribution model in the country, considering comprehensive conservation to have been accomplished only when the entire distribution occurs within protected areas.

The ecological representativeness score in situ (ERSin) is an ecological measurement of the proportion of the range of a taxon in Tunisia that can be considered to be conserved in protected areas. The ERSin compares the ecological variation encompassed within the range located inside protected areas in Tunisia to the ecological variation encompassed within the total area of the distribution model in the country, considering comprehensive conservation to have been accomplished only when every ecoregion potentially inhabited by a taxon is included within the distribution of the species located within a protected area. A final conservation score for in situ (FCSin) was derived by calculating the average of the two individual in situ conservation metrics.

A final combined conservation score (FCS-mean) was calculated for each taxon by averaging its ex situ (FCSex) and in situ (FCSin) totals. As additional useful metrics, we computed a minimum combined score (FCS-min) by selecting whichever score (between the FCSex and FCSin) was smaller, reflecting the stance that both conservation strategies are important and both need to be increased to an ideal level. We also calculated a maximum combined score (FCS-max) by selecting whichever score was greater, reflecting the stance that the total comprehensiveness of conservation is what is ultimately important, regardless of the conservation strategy used. Taxa were categorized based on their final mean combined conservation score, with high priority for further conservation action for species where FCS < 25, medium priority where 25 ≤ FCS < 50, low priority where 50 ≤ FCS < 75, and sufficiently conserved for taxa whose FCS ≥ 75.

RESULTS
Distributions of Assessed Crop Wild Relatives of Carrot
A total of 2477 occurrence records were compiled and processed for the 13 taxa for distribution modeling across the western Mediterranean, including 2113 H records and 364 G records (Table 2, Supplemental Table S3). The number of records for the western Mediterranean per taxon ranged from two (D. reboudii) to 1349 (D. carota subsp. carota).

All of the models for the 10 taxa with sufficient data to run the MaxEnt algorithm passed our evaluation criteria, except for the model for D. setifolius, which displayed one metric slightly outside the cAUC threshold (Supplemental Table S2). Based on current knowledge of the taxa, we consider all these models robust. Due to lack of data, it was not possible to produce distribution models for three taxa: D. gracilis, with only five occurrence points; D. reboudii, with only two occurrence points; and D. virgatus, with only two occurrence points. Although these taxa were not modeled, we discuss their distributions and conservation status below, based on the reference occurrences in Tunisia. Interactive distribution models and associated evaluation criteria for all modeled taxa are available in Supplemental Dataset S2.

Potential distributions of the assessed carrot crop wild relatives occur throughout the great majority of the coastal areas of the western Mediterranean (Fig. 1). Although nine of the taxa are widespread in the western Mediterranean, D. carota subsp. capillifolius and D. syriicus are limited to Tunisia and adjacent Libya, and D. virgatus and D. reboudii are limited to Algeria and adjacent Tunisia. Taxon richness was highest in Morocco, Tunisia, and southern Spain, with up to nine taxa potentially overlapping in the same approximately 5-km$^2$ areas. Cohabitation (at 1 m or less) of D. carota subsp. carota with D. carota subsp. capillifolius in the south and the center of Tunisia, and with D. muricatus in the north, was often observed in the field.

Ecogeographic Characterization
Throughout the entire ranges of the taxa in the western Mediterranean, substantial variation was found across taxa (e.g., with median occurrence in the highest maximum temperatures in the warmest month of the year for D. syriicus, D. carota subsp. capillifolius, and D. aureus, and the lowest temperatures in the coldest month measured by median of occurrences for D. carota subsp. carota and D. crinitus). Variation was also seen between populations within taxa, particularly for D. carota subsp. carota (Supplemental Fig. S1). Temperature, precipitation, and other ecogeographic variation was also evident in Tunisian-only occurrences, both across taxa and between populations within taxa (Supplemental Fig. S2).

A principal component analysis (PCA) revealed that 71.3% of variation across all occurrences and all taxa was explained by the first two components. Both components (with 43 and 28.3% of variance, respectively) were generally positively correlated with temperature and precipitation variables with additional correlations of the first component to water vapor pressure and solar radiation,
and the second component to wind speed (Supplemental Fig. S3A–S3B). Performing a cluster analysis based on the PCA results allowed the segregation of occurrence data into three ecogeographic groups. The first and the third clusters were characterized generally by high variability related to precipitation and low temperature variables, whereas the second cluster was characterized by high temperature and low precipitation (Supplemental Fig. S3C–S3E). Analyses of populations of taxa per cluster (Supplemental Fig. S3F–S3G) showed that Cluster 1 is mainly represented by *D. carota* subsp. *carota*, whereas Clusters 2 and 3 contain the other taxa. More than 85% of *D. syrticus* and *D. carota* subsp. *capillifolius* occurrences fell in Cluster 2, and >84% of *D. carota* subsp. *gummifer*, *D. muricatus*, and *D. rouyi* occurrences fell in Cluster 3.

**Conservation Status**

For Tunisia, 219 occurrence records were used in the conservation assessment, including 51 H records and 168 G records. With regard to ex situ conservation, taxa ranged from no representation at all from Tunisia (*D. crinitus*, *D. gracilis*, *D. reboudii*, *D. setifolius*, and *D. vinigatus*) to a very high representation in genebanks and botanic gardens (*D. carota* subsp. *capillifolius*, with an FCSex of 98.36, and *D. muricatus*, with 94.17), as an average across sampling, geographic, and ecological representation (Fig. 2, Table 2, Supplemental Table S3). Due to the relatively small size of Tunisia and thus few ecoregions outlined by Olson et al. (2001) occurring in the country, sufficiency ex situ with regard to ecoregions sampled was higher than the geographic metric in all cases except for *D. carota* subsp. *carota*. The mean FCSex across all taxa was 41.25.

In total, with regard solely to the state of ex situ conservation, five taxa (*D. crinitus*, *D. gracilis*, *D. reboudii*, *D. setifolius*, and *D. vinigatus*) could be considered high priority for further collecting, three (*D. carota* subsp. *gummifer*, *D. pumilus*, and *D. aureus*) medium priority, one (*D. rouyi*) low priority, and four (*D. carota* subsp. *carota*, *D. syrticus*, *D. muricatus*, and *D. carota* subsp. *capillifolius*) sufficiently conserved. We note that *D. rouyi* and all the medium-priority species are represented by four or less germplasm (G) accessions and thus may warrant higher conservation prioritization than that given by our gap analysis results.

Further collecting priorities for crop wild relatives of carrot in Tunisia are concentrated in the northeast, north, and northwest coasts, with uncultivated populations of up to seven taxa potentially occurring in the same ~5-km$^2$ areas (Fig. 3A).

The analysis of representation ex situ of ecogeographic variation in Tunisian populations of the target taxa indicated that, in general, those taxa with relatively large numbers of samples (e.g., *D. carota* subsp. *capillifolius*, *D. carota* subsp. *carota*, *D. syrticus*, and *D. muricatus*) are well represented in genebanks and botanic gardens across their ecogeographic variation ranges (Supplemental Fig. S2). Those taxa with no (*D. crinitus*, *D. gracilis*, *D. reboudii*, *D. setifolius*, and *D. vinigatus*) or very few (*D. carota* subsp. *gummifer*, *D. rouyi*, *D. pumilus*, and *D. aureus*) germplasm samples are of course poorly represented ecogeographically.

With regard to in situ conservation in officially recognized protected areas, taxa ranged from no representation at all in Tunisia (*D. gracilis*, *D. reboudii*, and *D. vinigatus*; note that since models were not available for these taxa, we assessed whether their occurrence records were located in protected areas and found that none were) to a

---

**Fig. 1.** Taxonomic richness map for carrot wild relative potential distribution models, (A) across the western Mediterranean, with focus on (B) Tunisia. Darker colors indicate greater numbers of taxa potentially overlapping in the same areas.
moderate level of representation (D. syrticus, with an FCSIn of 56.15; D. crinitus with 52.17; D. rouyi with 51.7; D. setifolius with 51.59; D. muricatus with 51.24; and D. carota subsp. capillifolius with 50.74), as an average across geographic and ecological representation (Fig. 2, Table 2, Supplemental Table S3). As with the ex situ analysis, sufficiency with regard to ecoregions covered in protected areas was higher than geographic coverage, in this case for all taxa and with six of the taxa being fully represented with regard to diversity of ecoregions in protected areas. On average across all taxa, in situ conservation was determined to be slightly less well accomplished than ex situ conservation, with a mean FCSIn across all taxa of 34.33.

In total, with regard solely to the state of in situ conservation, three taxa (D. gracilis, D. reboudii, and D. virgatus) could be considered high priority for further in situ conservation, four (D. carota subsp. gummifer, D. pumilus, D. carota subsp. carota, and D. aureus) medium priority, and six (D. syrticus, D. crinitus, D. rouyi, D. setifolius, D. muricatus, and D. carota subsp. capillifolius) low priority, with none considered sufficiently conserved. Paralleling ex situ conservation collecting priorities, the most efficient establishment of additional protection for carrot crop wild relatives in protected areas in Tunisia with regard to geographic gaps would be concentrated in the northeast, north, and northwest coastal areas (Fig. 3B).

With regard to combined conservation status (both ex situ and in situ), taxa ranged from no conservation at all in Tunisia (D. gracilis, D. reboudii, and D. virgatus) to a high level of conservation (D. carota subsp. capillifolius, with an FCS-mean of 74.55; D. muricatus with 72.7; and D. syrticus with 71.35) (Fig. 2, Table 2, Supplemental Table S3). The average FCS-mean across all taxa was 37.79. In combination, three taxa (D. gracilis, D. reboudii, and D. virgatus)
could be considered high priority for further conservation, five (D. setifolius, D. crinitus, D. carota subsp. gummifer, D. pumilus, and D. aureus) medium priority, and five (D. rouyi, D. carota subsp. carota, D. syrticus, D. muriatus, and D. carota subsp. capillifolius) low priority, with none considered sufficiently conserved.

**DISCUSSION**

With 61.5% of the crop wild relatives of carrot in Tunisia assessed as high or medium priority for further collecting for ex situ conservation, 53.9% as high or medium priority for further protection in situ, and 61.5% as high or medium priority with regard to the strategies in combination, it is clear that further conservation action is needed to safeguard the wild genetic resources of this important crop occurring in the country. Included in this list of priorities are taxa with zero germplasm accessions accessible to the global community in genebanks and botanic gardens (D. crinitus, D. gracilis, D. reboudii, D. setifolius, and D. virgatus), as well as another four with less than five accessions conserved ex situ. The rarity of some taxa (e.g., D. crinitus) and the high level of endemism (e.g., D. virgatus and D. reboudii) (Pottier Alapetite, 1979) provide further insights into the low level of representation of these species in genebanks and botanic gardens. Other taxa (D. pumilus, D. rouyi, and D. carota subsp. gummifer) were observed to grow in coastal areas that are quickly being modified by road construction and tourism activities.

Some taxa appear to be confined to quite specific habitats, including D. carota subsp. gummifer, D. rouyi, and D. pumilus (lowland coastal areas) and D. syrticus (dry sandy soils near the beach below 300 m in altitude). In addition to habitat preferences, many Daucus species have distinct phenological differences, not modeled here, but critical to efficient field collecting. This was demonstrated during recent collections in Spain in 2016 by coauthors Martínez Flores, Simon, and Spooner. As Daucus germplasm must be collected with mature fruits, three trips were required to optimize the collections. In Tunisia, D. pumilus was best collected in June; D. aureus, D. carota (all subspecies), D. crinitus, D. muriatus D. rouyi, and D. syrticus in were best collected in mid-August and early September; and D. setifolius was best collected in early October. Additional fieldwork is needed to understand the best times to collect D. gracilis, D. reboudii, and D. virgatus.

Two taxa (D. syrticus and D. carota subsp. capillifolius) considered endemic to Tunisia and adjacent Libya were identified as exhibiting potential adaptation to high temperatures and low precipitation. The contribution of such traits from wild relatives has not yet advanced to their use in currently grown carrot cultivars, but given anticipated abiotic and biotic challenges in future production agriculture, the identification of germplasm sources for these important traits is valuable. Major carrot production regions in Asia, southern Europe, and North America already experience elevated temperatures reducing carrot...
yields and quality. Severe water shortages in some of these regions may be mitigated by access to more abundant irrigation, but this often increases salinity. These abiotic stressors often enhance the damage from diseases and pests, to further threaten crop productivity. The occurrence of *D. carota* subsp. *gummifer* (subsp. *gingidium*) (as well as several other *Daucus* species) growing in close proximity to the salt spray of the Atlantic Ocean and Mediterranean Sea suggests these carrot wild relatives as candidate germplasm sources of salt stress tolerance. Although our results classified *D. syrticus* and *D. carota* subsp. *capillifolius* as low priority for further collecting, further field exploration in extreme high-salinity soils may lead to the discovery of new populations with particularly valuable adaptation to this abiotic stressor. This is further justified as these taxa are closely related (in Gene Pool 1) to the crop. Natural hybridization between *D. carota* subsp. *capillifolius* and the cultivated carrot (*D. carota* subsp. *sativus*) was frequently observed in Tunisia (Mezghani et al., 2014, 2018).

As the geographic gaps in both ex situ and in situ conservation largely align with taxonomic richness geography in Tunisia, hotspots, especially in the northern coastal areas, represent particularly high-value regions for further efficient collection of the taxa for ex situ conservation, as well as for further protection in situ. Additional collecting across the country will be needed to form germplasm collections that are comprehensive at the population level, and additional protected areas will be needed to sufficiently conserve the taxa in situ.

One challenge inherent to the modeling analysis is that distributions of taxa are likely also driven by factors beyond the present 26 ecogeographic predictors used here. These include biotic (e.g., mycorrhizae, pollinators, and dispersal agents), other abiotic (e.g., soil types), and stochastic factors. Furthermore, the habitat suitability models are unable to account for extirpation of populations due to habitat degradation or destruction due to anthropogenic factors. In sum, the models in many cases may be considered optimistic relative to actual distributions. Our results, therefore, should be considered as planning tools to guide explorations but which should be verified with additional field work. Further taxonomic and genotyping analyses are also needed to fill research gaps for this difficult genus of global importance.

With regard to the conservation analyses, openly available databases on genebank and botanic garden holdings are not fully representative of all such institutions and coordinate and/or other locality information is also presently lacking for a large number of records that are available. If either or both of these constraints were to be remedied, it is possible that the ex situ conservation status of some taxa could be revised in a positive direction. Moreover, although the lands listed in the World Database on Protected Areas afford collateral protection to carrot wild relatives because of overall land conservation practices, robust long-term protection of these taxa in these areas will require the formation of active taxon- and population-specific management plans.

**Supplemental Material**

Supplemental material is included with this manuscript. This includes the full occurrence dataset, including ecogeographic data values (Supplemental Dataset S1); taxon-level modeling and gap analysis results (Supplemental Dataset S2); and ecogeographic characterizations for the western Mediterranean (Supplemental Fig. S1) and for Tunisia (Supplemental Fig. S2), including a principal component and cluster analysis (Supplemental Fig. S3), a list of ecogeographic variables used in the potential distribution modeling and ecogeographic characterization (Supplemental Table S1), distribution modeling metrics per taxon (Supplemental Table S2), and extended conservation gap analysis results per taxon (Supplemental Table S3). All code implemented in our analysis is available at https://github.com/dcarve1/cwrSDM. The ecogeographic predictors, ecoregions, and protected areas datasets are openly available for use through Khoury et al. (2019b).

**Conflict of Interest**

The authors declare that there is no conflict of interest.

**Acknowledgments**

The authors thank the botanists, taxonomists, plant collectors, geospatial scientists, and genetic resource professionals who compiled and made available through open access repositories the occurrence and ecogeographic information used in this analysis. The USDA is an equal opportunity employer and provider.

**References**


Global Biodiversity Information Facility (GBIF). 2018b. GBIF occurrence download. doi:10.15468/dl.n82beg


