RESEARCH ARTICLE

Genetic structure and ecological niche space of lentil's closest wild relative, *Lens orientalis* (Boiss.) Schmalh.

A. Guerra-García[1,2] (iD), O. Trněný[3] (iD), J. Brus[4] (iD), J. P. Renzi[5] (iD), S. Kumar[6], M. Bariotakis[7,8] (iD), C. J. Coyne[9] (iD), A. Chitikineni[10], K. E. Bett[1] (iD), R. Varshney[10,11] (iD), S. Pirintsos[7] (iD), J. Berger[12] (iD), E. J. B. von Wettberg[13] (iD) & P. Smýkal[14] (iD)

1 Department of Plant Sciences, University of Saskatchewan, Saskatoon, Canada
2 Departamento de Biotecnología y Bioquímica, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Irapuato, Guanajuato, Mexico
3 Agriculture Research Ltd, Troubsko, Czech Republic
4 Department of Geoinformatics, Palacký University, Olomouc, Czech Republic
5 Instituto Nacional de Tecnología Agropecuaria, Buenos Aires, Argentina
6 International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat, Morocco
7 Department of Biology, University of Crete, Heraklion, Greece
8 Botanical Garden, Rethymnon, Greece
9 Western Regional Plant Introduction Station, USDA-ARS, Pullman, WA, USA
10 International Crop Research Institute for the semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, India
11 Murdoch University, Murdoch, WA, Australia
12 CSIRO Plant Industry, Wembley, WA, Australia
13 Department of Plant and Soil Sciences, Gund Institute for the Environment, University of Vermont, Burlington, VT, USA
14 Department of Botany, Palacký University, Olomouc, Czech Republic

ABSTRACT

- Crops arose from wild ancestors and to understand their domestication it is essential to compare the cultivated species with their crop wild relatives. These represent an important source of further crop improvement, in particular in relation to climate change. Although there are about 58,000 *Lens* accessions held in genebanks, only 1% are wild.
- We examined the geographic distribution and genetic diversity of the lentil's immediate progenitor *L. orientalis*. We used Genotyping by Sequencing (GBS) to identify and characterize differentiation among accessions held at germplasm collections. We then determined whether genetically distinct clusters of accessions had been collected from climatically distinct locations.
- Of the 195 genotyped accessions, 124 were genuine *L. orientalis* with four identified genetic groups. Although an environmental distance matrix was significantly correlated with geographic distance in a Mantel test, the four identified genetic clusters were not found to occupy significantly different environmental space. Maxent modelling gave a distinct predicted distribution pattern centred in the Fertile Crescent, with intermediate probabilities of occurrence in parts of Turkey, Greece, Cyprus, Morocco, and the south of the Iberian Peninsula with NW Africa. Future projections did not show any dramatic alterations in the distribution according to the climate change scenarios tested.
- We have found considerable diversity in *L. orientalis*, some of which track climatic variability. The results of the study showed the genetic diversity of wild lentil and indicate the importance of ongoing collections and *in situ* conservation for our future capacity to harness the genetic variation of the lentil progenitor.

INTRODUCTION

Crops arose from wild ancestors and to understand their domestication it is essential to compare the cultivated species with their crop wild relatives (CWRs). CWRs are a critical pool of genetic diversity for crop breeding, particularly for the improvement of important traits such as disease resistance, abiotic stress tolerance, nutritional quality traits, and adaptation to changing climates (Smýkal *et al.* 2015; Dempewolf *et al.*

2017; Coyne *et al.* 2020). A key component of the value of CWR for the resilience of agricultural systems is that CWR populations have evolved for longer periods of time compared with crops, responding to changes in climate, to pressures from diseases and herbivores, and more recently, to anthropogenic habitat change (*e.g.* Warschefsky *et al.* 2014).

The Fertile Crescent, in modern southern Turkey, Syria, Iran, Lebanon, and Israel, is the site of the domestication of several crops, such as cereals and the three grain legumes:

lentils, peas and chickpeas. Archaeological evidence suggests nearly simultaneous domestication of these crops around 10,000 years ago, although there has long been debate about the speed of domestication and the possibility of multiple domestication events (*e.g.*, Larson *et al.* 2014; Abbo & Gopher 2020; Allaby *et al.* 2022). This region has also undergone significant habitat and climatic change over the past 10,000 years, with widespread loss of forest cover, intensive grazing, and multiple shifts in climate (Lawrence *et al.* 2021).

Lentil belongs to the group of grain legumes with the oldest domestication, as it appears early in archaeological records in the Fertile Crescent (Sonnante *et al.* 2009). The archaeological records provide evidence of lentil domestication in Syria and south-eastern Turkey in approximately 8500 BC. Thereafter, lentil cultivation spread throughout the Mediterranean region to Asia, and Europe during the Bronze Age. Between 5000 and 4000 BC, lentil cultivation moved eastwards to Georgia and finally reached India and Pakistan around 2000 BC (Zohary & Hopf 1973; Cubero 1981; Sonnante *et al.* 2009). Archaeological evidence shows that Indians have enjoyed lentil-based dishes since at least the early Harrapan Period, which started around 2800 BC with the establishment of the first urban centers (Pokharia & Srivastava 2013).

Geographical distribution of genetic diversity and its relationship to the environment have been studied in several crop progenitors, such as common bean (Rodriguez *et al.* 2016), barley (Thormann *et al.* 2016), soybean (Leamy *et al.* 2016; Sedivy *et al.* 2017), chickpea (van Oss *et al.* 2015) and pea (Smýkal *et al.* 2017; Hellwig *et al.* 2020, 2021, 2022). Such studies are essential for understanding their spatial distribution, identifying areas and highly diverse populations, and prioritizing their conservation *in* and *ex-situ*. The characterization of CWR collections can lead to the discovery of variants that can be then used for breeding purposes (Warschefsky *et al.* 2014; Coyne *et al.* 2020). Previous lentil diversity studies have focused primarily on the cultivated *Lens culinaris* Medik. (*e.g.*, Guerra-Garcia *et al.* 2022) and smaller *Lens* wild species sets sourced from a single genebank without precise georeferenced data (Dikshit *et al.* 2015; Wong *et al.* 2015; Khazaei *et al.* 2016; Koul *et al.* 2017). These studies on *Lens* diversity used various molecular approaches, from anonymous diversity assessment where the genomic position of the markers is unknown (Ferguson *et al.* 1998a,b; Alo *et al.* 2011), to genome-wide analysis (Wong *et al.* 2015; Dissanayake *et al.* 2020; Liber *et al.* 2021; Guerra-Garcia *et al.* 2022) and pan-transcriptome analysis (Gutierrez-Gonzalez *et al.* 2022). More extensive information on the association of wild *Lens orientalis* with varying climates will help expand the usefulness of this germplasm is breeding. As is true in many other crops (*e.g.*, Hajjar & Hodgkin 2007), wild *Lens* germplasm has already proven highly useful in breeding for disease resistance in lentils (e.g, Tullu *et al.* 2006; Singh *et al.* 2018; Coyne *et al.* 2020; Mohar *et al.* 2020; Gela *et al.* 2021; Asghar *et al.* 2021). However, the broad distribution of wild *Lens* suggests that there may also be potential for climatic adaptation (*e.g.*, Coyne *et al.* 2020; Guerra-Garcia *et al.* 2022). Better information on the association of climate at the site of origin with genetic variation can help guide future introgression efforts. For example, if breeders are selecting for drought tolerance, they may want to not only select wild parents that may be from drier locations, but also wild material from populations that have higher levels of diversity or are more genetically distinct from other parents used in other crosses.

The genus *Lens* Miller (2n = 2 × =14) is a member of the Fabaceae family and its taxonomy has gone through several modifications. Seven taxa are recognized, including *L. culinaris* (the domesticated species) and its closest wild relative *L. orientalis* (Boiss.). The six wild *Lens* taxa are widely distributed in the Mediterranean region, and the occurrence of all species overlaps in the southwestern part of Turkey and the Aegean islands (Guerra-Garcia *et al.* 2022). *L. orientalis* has an eastern distribution from Turkey and the Middle East up to Uzbekistan (Guerra-Garcia *et al.* 2022). The genus *Lens* has undergone several revisions in the past few decades (*e.g.*, Wong *et al.* 2015; Ogutcen *et al.* 2018; Dissanayake *et al.* 2020; Guerra-Garcia *et al.* 2022), with the most recent rearrangement by Wong *et al.* (2015) considering *L. orientalis* as its own species because of its genetic separation from cultivated *L. culinaris*. The Euro+Med PlantBase (2006) and ILDIS World Database of Legumes (Roskov *et al.* 2010) still retain *L. culinaris* subp. *orientalis*. All *Lens* species are autogamous, with observed selfing rates above 95% (*e.g.*, Wong *et al.* 2015).

Here we examine the geographic distribution and genetic diversity of lentil progenitor *L. orientalis*. We used Genotyping by Sequencing (GBS) to characterize differentiation among accessions of *L. orientalis* with available geographical data held at germplasm collections to assess population structure. We investigate whether there is evidence of systematic selection for ecotypes between environments, or if population structure is better explained by proximity, in which case relatedness and distance are correlated.

MATERIAL AND METHODS

Plant material

A total of 195 lentil accessions registered as *L. orientalis* from the *ex-situ* collections were used: 145 from ICARDA, 22 from USDA NPGS, 23 from Millennium Seed Bank, Kew, UK, and five from Israel Gene Bank (Table S1 and Appendix S2). Plants were grown in a greenhouse and were inspected morphologically to verify their taxonomic status, excluding accessions that presented large seeds and indehiscent pods, using the same conditions as Smýkal *et al.* (2017). In order to corroborate the identity of the samples, 51 accessions genotyped by Wong *et al.* (2015) from different *Lens* species were also included.

Genotyping by sequencing

Following previous work characterizing diversity in allogamous wild *Lens* (*e.g.*, Alo *et al.* 2011; Wong *et al.* 2015; Dissanayake *et al.* 2020; Liber *et al.* 2021), one individual from each of the 195 lentil accessions was genotyped. We used one individual because *Lens* is autogamous and all of the genebanks from which we sourced material have maintained their collections by selfing since their initial field collection. For this, high molecular weight genomic DNA was isolated from plant leaves collected from 10-day-old seedlings using the high throughput mini-DNA extraction method (DNeasy 96 Plant Kit; Qiagen, Valencia, CA, USA). The quality and quantity of the DNA were assessed using a spectrophotometer (Shimadzu UV160A, Japan). Accessions were genotyped using the Genotyping by

Sequencing (GBS) approach as described by Elshire *et al.* (2011). About 10 ng genomic DNA from each sample was restriction-digested using *ApeKI* (recognition site: G/CWCG) endonuclease. The digested product was ligated with uniquely barcoded adaptors using T4 DNA ligase enzyme and was further incubated at 22 °C for 1 h and heated at 65 °C for 30 min to inactivate the T4 ligase. Digested ligated products from each sample were mixed in equal proportion to construct the GBS libraries, which were then amplified and purified to remove the excess adapters. Libraries were sequenced on the HiSeq 2500 platform (Illumina, San Diego, CA, USA) to generate genome-wide reads.

Single nucleotid polymorhism (SNP) calling and variant filtering

Fastq files were demultiplexed with GBSx 1.3 (Herten *et al.* 2015) and reads were aligned with Nextgenmap 0.5.4 (Sedlazeck *et al.* 2013) to the *Lens culinaris* genome (Ramsay *et al.* 2021). The resulting sequences aligned (sam files) were then converted to binary files using samtools 1.12 (Kaisers *et al.* 2015). SNPs were discovered for each sample using the HaplotypeCaller tool and genotypes were merged with GenotypeGVCFs. Both tools are from the Genome Analysis Toolkit (GATK 4.2.0.0). VCFtools 0.1.16 (Danecek *et al.* 2011) was used to perform the variant filtering according to the following parameters: minimum mean depth 4×; maximum missingness per sample 0.20; minimum allele frequency (MAF) 0.05. Only biallelic sites were kept, and accessions with missing data >0.50 were removed.

Population structure and genetic diversity

We explored the population structure of the accessions identified as *L. orientalis* using Admixture version 1.3 (Alexander *et al.* 2009). A Principal Components Analysis (PCA) was performed with SNPrelate (Zheng *et al.* 2012), and a maximum-likelihood phylogenetic tree was constructed with FastTree (Price *et al.* 2009) including the accessions from the different *Lens* taxa that were genotyped due to apparent germplasm misclassification as *L. orientalis* (see Results). The PCA and the phylogenetic analysis were repeated only with the *L. orientalis* samples. For each genetic cluster of *L. orientalis* heterozygosity (H_O and H_E) and inbreeding coefficient (F_{IS}) per site (SNP) were estimated and then averaged with the Hierfstat package (Goudet 2005), performing a bootstrap (1,000) to obtain confidence intervals for the inbreeding coefficient.

Environmental data analysis

We used WorldClim (Fick & Hijmans 2017) version 2.1 as a database of climatic variables and 19 variables impacting biological distributions (BIO variables). For analyses based on elevation, digital elevation data of SRTM included in WordClim database was used. A raster processing in ArcGIS Pro was used to extract point data information from available datasets. We developed a script to download desired raster information from point data within a Google Earth Engine (GEE) environment (Gorelick *et al.* 2017). Some of the acquired information was downscaled by averaging as needed to match the scale of the other datasets (Fick & Hijmans 2017).

Relationships among *L. orientalis* genetic clusters and environmental variables

We assessed correlation among the 19 bioclimatic variables (BIO1–BIO19) in the WorldClim dataset. To avoid over-parameterization among the bioclimatic variables, we used Pearson correlation coefficients (assessed in Infostate) to measure pairwise correlations among the variables, and one of the two paired variables correlated above 0.8 was eliminated. Based on this analysis, ten bioclimatic variables were selected for further analysis: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), the maximum temperature of the warmest month (BIO5), temperature annual range (BIO7), annual precipitation (BIO12), precipitation of the wettest month (BIO13), precipitation seasonality (BIO15), and precipitation of the driest quarter (BIO17; Hijmans *et al.* 2005). The matrix of the reduced set of 10 Worldclim variables was analysed by PCA using Infostat software.

To determine whether the differences between genetic clusters could be explained by the environmental variables, a linear model was performed for each environmental variable using the genetic cluster as a fixed factor and accessions as a random factor. The model was performed with R version 3.6.3 using Infostat software interface to R. Clusters were compared by post-hoc mean Fisher's Least Significant Difference test ($P < 0.05$).

Relationships between genetic, geographic, and environmental distance matrices

The relationships between genetic, and both geographic and environmental distance matrices were assessed for the accessions identified as *L. orientalis*. For this, three matrices were prepared and examined using the Mantel test (Smouse *et al.* 1986) with vegan R package (Dixon 2003). The physical distance between accessions was estimated using geographic distance (GGD) for latitude (x)/longitude (y) values: $GGD = (xi - xj)^2 + (yi - yj)^2$ (Peakall & Smouse 2006). The Nei's genetic distance between the *L. orientalis* accessions was calculated with Tassel (Bradbury *et al.* 2007). Euclidean distances of the first four principal components obtained from the Worldclim variables (see previous section) were used for a partial Mantel test. The significance of the normalized Mantel coefficient was calculated using a two-tailed Monte Carlo permutation test with 10,000 permutations.

Ecological niche analysis

To develop Ecological Niche Models (ENMs) we used geographic locations of a broader set of 710 accessions (Appendix S2). We began with sites from a recent set of surveys we performed (Berger *et al.* unpublished, further described below, n = 192), supplemented with locations from Global Biodiversity Information Facility (GBIF, n = 234), and including locations from genebanks including both the accessions we verified as well as accessions for which seed is not available ICARDA (n = 169), USDA (n = 22), Israeli database (n = 58), Millenium Seed Bank, Royal Botanic Garden Kew, UK (n = 25), Australian genebank (n = 10) (Appendix S2). Points from our own analysis were sites where wild *Lens* was observed

in southeastern Turkey, as part of a wild *Cicer* collection that was performed in 2013–2015 (Berger pers. obs) with results of collections reported in other publications (*e.g.* von Wettberg *et al.* 2018; Toker *et al.* 2021). As environmental predictors, all 19 bioclimatic variables (Fick & Hijmans 2017) were used. We selected Maxent (version 3.4.1, Phillips *et al.* 2017) as our modelling approach. Maxent is a maximum-entropy based machine learning method, which has been shown to perform better than other methods when sample sizes are small (Elith *et al.* 2006; Hernandez *et al.* 2006) and avoids over-parameterization by means of regularization (Phillips *et al.* 2006). It estimates the potential niche instead of the realized distribution of the modelled entity (Phillips *et al.* 2006), which is useful for the targets of the present study.

The potential distribution was estimated according to the grouping suggested by Admixture which showed the lowest error rate (see Results). We also projected the models in past climatic conditions of the Last Glacial Maximum (around 22,000 YBP) using all 19 bioclimatic variables. Although these projections assume niche conservatism, they can provide valuable historical information about the groups. We also constructed models for future conditions projected for 2050 and 2070 based on two separate climate change scenarios: Representative Concentration Pathway (RCP) 4.5 and RCP 6.0. The first one is an optimistic scenario that projects a medium-high effort to reduce greenhouse gas emissions, and the second scenario projects a lower effort to curb the emission and a higher temperature increase (van Vuuren *et al.* 2011). For both the future and past projections the climatic variables were retrieved from Worldclim website (http://www.worldclim.com; Hijmans *et al.* 2005). We assessed whether high and low altitudes are characterized by different suitability values for the 710-point model. For that purpose, we downloaded an elevation model with the same spatial resolution as our predictions (2.5 arc minutes) and classified elevation as low (<=700 m) and high (>700 m). The suitability values were then compared between the two classes. Furthermore, we employed Shannon's diversity index as a measure of the genetic diversity and its geographical patterns, following the methodology used in Smýkal *et al.* (2017). Briefly, Shannon's diversity index was calculated using the genetic groups that were found during the STRUCTURE analysis, and the results of the haplotypes' distribution modelling. Therefore, instead of 'proportions of species', we calculated the index based on the Maxent modelling results (*i.e.*, relative suitability or relative probability of occurrence) of the genetic groups. The index was calculated using a custom R script on a per-pixel basis, using the same geographical grid of pixels that was used for the distribution modelling.

RESULTS

Population structure and genetic diversity

Of the 195 accessions genotyped, eight were not further examined due to missing data. Over 3 million genetic variants were identified and 3,271 SNPs were kept after filtering. The mean site depth was 16.77× and the mean missingness per individual was 0.12. In genebank-derived samples incorrect taxonomic assignment is common, particularly in groups with a history of frequent taxonomic revisions. Although the retrieved accessions were originally taxonomically assigned as *L. orientalis*, we

used previously identified accessions of the seven *Lens* species (six wild and 1 cultivated species) from Wong *et al.* (2015) to verify the identity of the accessions. The results of the PCA (Fig. 1a) and phylogenetic (Fig. 1b) analysis showed that the dataset with the 187 genotypes included all *Lens* species. A total of 124 accessions were identified as genuine *L. orientalis* and 63 corresponded to other taxa: 31 to cultivated lentil (*L. culinaris*), 6 to *L. ervoides*, 1 to *L. lamottei*, 4 to *L. nigricans*, 8 to *L. odemensis*, and 13 to *L. tomentosus* (Table S1 and Appendix S2).

The filtering was repeated including only the confirmed *L. orientalis* accessions using the same filtering parameters. The resulting dataset contained 124 accessions and 6,969 SNPs with a mean site depth of 16.75× and a mean missing value of 0.07 per sample. For the 124 *L. orientalis* accessions Admixture suggested four genetic clusters (Figures S1 and S2a). Two groups were clearly observed with the first principal component of the PCA (Fig. 2b). *K* values from two to five were explored (Figures S2 and S3) and the two groups observed with the PCA reflected the $k = 2$ clustering pattern. $K = 4$, the most likely grouping value, showed further details of the population structure and did not contradict the $k = 2$ grouping assignment. Interestingly, placing the genetic clusters (from $K = 2$ to $K = 5$) on a map revealed no clear geographic pattern (Fig. 2c, Figure S3). No differences in the genetic diversity in terms of heterozygosity were found (Figure S4a). As expected in autogamous species, the four clusters presented high values of inbreeding coefficient ($F_{IS} = 0.85$) and it was slightly higher in cluster 2 ($F_{IS} = 0.88$). The highest genetic differentiation was observed between clusters 2 and 3 (Figure S4b).

Relationships between genetic, geographical and environmental distance matrices

The environmental distance matrix was significantly correlated with the geographic distance matrix (Figure S5a), suggesting that environmental (BIOs) conditions diverge with increasing geographic distance. However, geographic distance was not significantly correlated with genetic distance (Figure S5b), consistent with the lack of geographic clustering observed in the accessions (Fig. 2c). The genetic and environmental distance matrices showed a weak but significant positive correlation ($R = 0.10$, $P < 0.01$; Figure S5c). From the biplot showing the four genetic clusters of the *L. orientalis* together with environmental factors, it can be seen that the clusters are not separated significantly by environmental factors (Fig. 3). Cluster 2 overlaps with clusters 1 and 4, while cluster 3 is more separated from them. Samples of cluster 3 originated from environments with lower temperature seasonality and range, and higher precipitation in the wettest month (Fig. 4). No significant differences were found between clusters for the rest of the bioclimatic variables ($P > 0.05$).

Niche analysis of wild lentil distribution

Maxent modelling of *Lens orientalis* gave a distinct potential distribution pattern (Fig. 5a), situated in the Fertile Crescent, with intermediate probabilities of occurrence in parts of Turkey, Greece, Cyprus, Morocco, and south of the Iberian Peninsula with NW Africa close to the Strait of Gibraltar, while higher probabilities of occurrence resulted in parts of the region of Cyrenaica in present-day Libya. Projections of
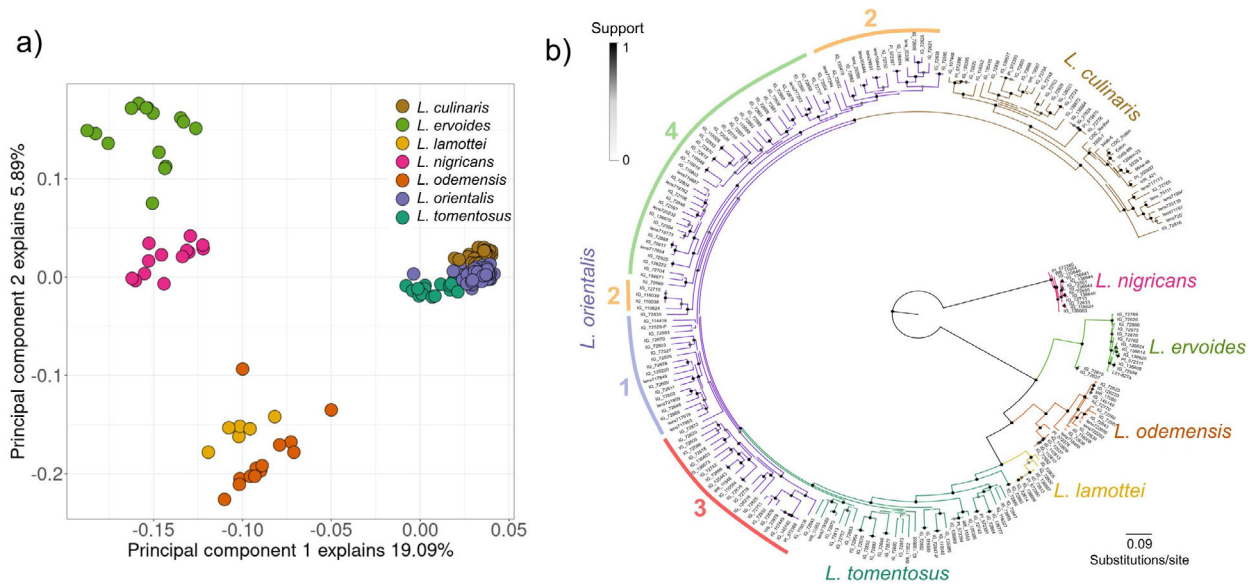
Fig. 1. a: PCA b: Maximum-likelihood tree showing the phylogenetic relationships of the 238 genotyped lentil accessions of *Lens* species included in this study (n = 187) and the accessions used by Wong *et al*. (2015) (n = 51). Dots in the nodes indicate the support. The genetic clusters identified in the accessions of *Lens orientalis* are shown outside the phylogenetic tree.
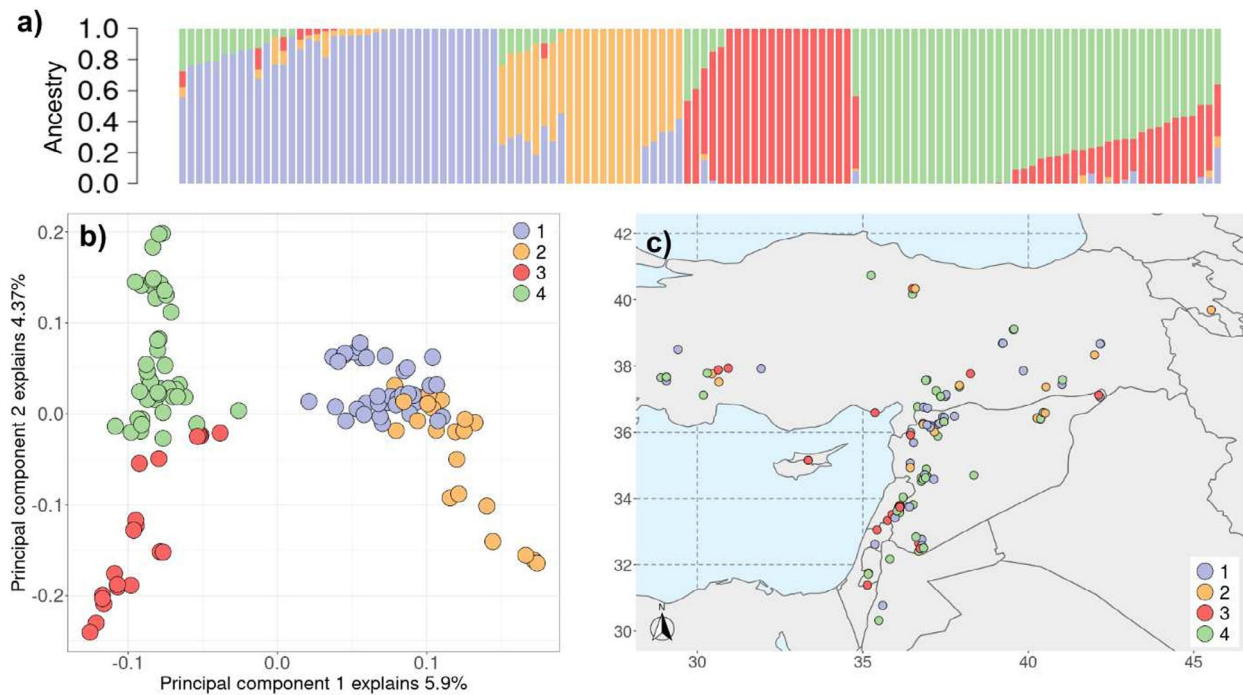

Fig. 2. a: Ancestry proportion of the four genetic clusters identified within the *Lens orientalis* accessions using Admixture. b: PCA of the *L. orientalis* samples. c: Their geographic distribution coloured according to the genetic groups.

Ecological Niche Modelling (ENM) in past conditions of the Last Glacial Maximum (LGM) gave a similar potential distribution pattern in the Mediterranean area with higher probabilities of occurrence in the Fertile Crescent and remarkably high in south Turkey, in parts of the region of Cyrenaica and in NW Africa close to the Strait of Gibraltar (Fig. 5b). Future projections of the ENM did not show any dramatic alterations in the distribution according to the selected climate change scenarios, although a slight shrinking of the high concentration areas in Israel and Lebanon was observed in the different scenarios (Fig. 6). The potential niche of the whole species seems to be largely stable in the projected future conditions. The suitability values of the model showed that altitudes above 700 m were characterized by a wider range of suitability values, as well as a higher mean value (Figure S6). Furthermore, this trend seemed to be stable across different temporal scales since the same
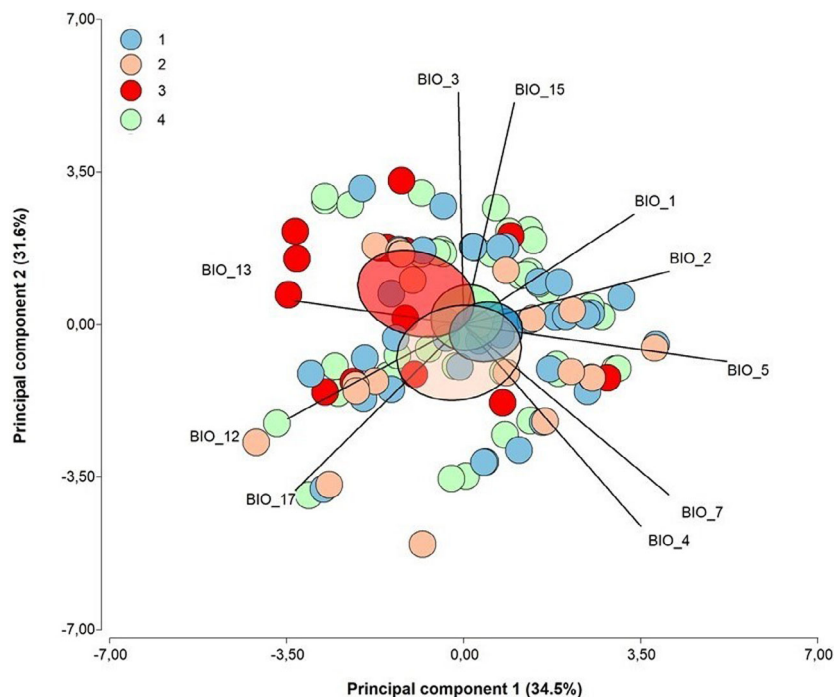
Fig. 3. Principal Component Analysis of selected bioclimatic variables of sites of origin of *Lens orientalis* accessions. Colours indicate the genetic clusters identified.
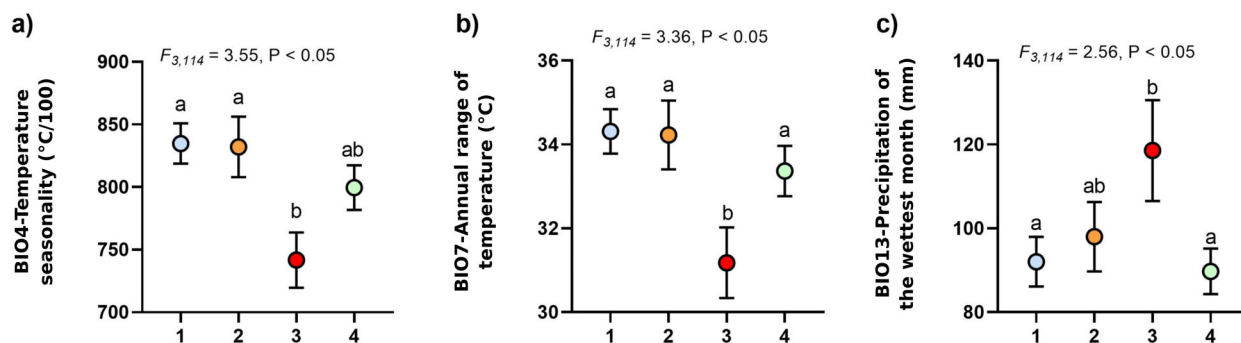

Fig. 4. Relationship of the four genetic clusters of *Lens orientalis* with a: Temperature seasonality (BIO4), b: Temperature annual range (BIO7) and c: Precipitation of the wettest month (BIO13) from the sites of origin. For each environmental variable, clusters marked with the same letters do not differ by LSD test ($P \leq 0.05$).

pattern is observed for the LGM (Fig. 5, Figure S7), and the predictions for 2050 and 2070 (Fig. 6). The current potential distribution of *L. orientalis* based on the relatively small dataset selected for genotyping (Table S1 and Appendix S2) seems to follow a similar pattern to results from the complete dataset (Figure S7), indicating that the small dataset is a representative subset. Genetic groups 1, 3 and 4 present a similar potential distribution (Fig. 4, Figure S7), following the pattern of the whole species. Genetic group 2 presents higher probabilities of occurrence westwards in areas such as Turkey, Greece, North Africa, and Spain (Figure S7). Concerning the projection for the LGM, only genetic group 1 followed a similar pattern to the present distribution, while the other three groups presented a broader probability of occurrence westwards, indicating a higher tolerance to the cold conditions during the LGM in the

Mediterranean area (Figure S7). A small-scale regional shift was observed in genetic group 2 since the LGM scenario presented higher probabilities of occurrence eastwards (Figure S6), but no change was observed for the future projections. Shannon's diversity index showed an increase of variation in the current distribution compared to the LGM (Fig. 7), particularly in Greece, Turkey, North Africa and Spain. This shift in the center of the genetic diversity of *L. orientalis* towards those areas is accentuated in the predictions of future distribution (Figure S8).

DISCUSSION

A century after Vavilov's seminal work (Vavilov 1926) we still study the origin of crops and their relationship to progenitors.
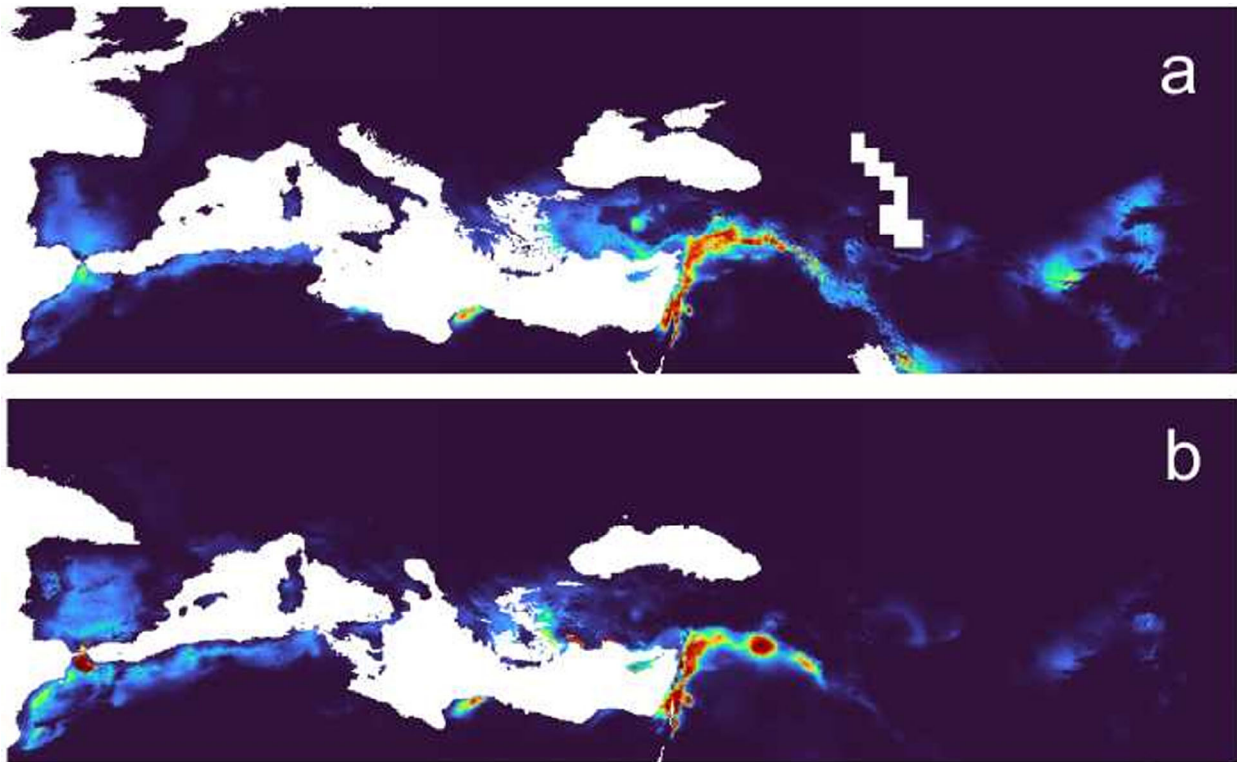
Fig. 5. Predicted potential distribution of *Lens orientalis* based on all the available records (710). a: Map for present conditions and b: Map shows climatic projections for the Last Glacial Maximum. The coloration is a heat map, with warmer (redder) colours indicating higher climatic suitability, and cooler (bluer) colours indicating lower climatic suitability.

With the advancement of molecular and ecological tools, we can ask questions on distribution range and comparison between species within geographical range. Furthermore, a comparison of CWRs can improve our capacity to harness adaptive traits from CWR (Hübner & Kantar 2021). Several studies focused on the diversity of domesticated lentil have been conducted (Khazaei *et al*. 2016; Liber *et al*. 2021; Guerra-Garcia *et al*. 2022), but this is not the case for the wild lentil species, not even for its closest relative *L. orientalis*. Here we performed the most extensive molecular analysis of global wild lentil collection to assess their genetic diversity and geographic distribution patterns. We have found considerable diversity in *Lens orientalis*, some of which track climatic variability. The distribution is suggestive of both the complex patterns of climatic variability in the Fertile Crescent, as well as patterns of possible human-mediated movement of wild materials (*e.g.*, hitchhiking seeds on livestock).

Genetic diversity of wild lentils

An important proportion of the lentil accessions included in the study were misclassified as *L. orientalis*. This shows the difficulty of identifying *Lens* species and the need to analyse the wild accessions that are preserved in genebanks in order to have more accurate information about CWR. It also shows the necessity of checking for misclassification in gene banks. We suspect that misclassification is quite common and that many researchers doing studies like ours simply excluded mislabelled accessions from further analysis and did not report them.

Genebank amplification could even create hybrids, if wild lentils of different taxa are amplified in neighbouring plots and low-frequency outcrossing occurs. Although we think this problem is more common than appreciated across many autogamous crops, we suspect many studies simply drop these samples from analyses without reporting them. Ignoring misclassified specimens fails to correct genebank passport annotations, and makes our collections less useful for breeding and research. Transparency about this challenge helps make the case for why genebanks require expanded resources for collection maintenance.

Four genetic clusters were identified in the wild lentil *L. orientalis*, which showed similar levels of genetic variation, and no geographic pattern in their current distribution was observed. Despite this, relatively significant differentiation was estimated for genetic cluster 3, particularly compared to cluster 2 (Figure S4b). All genetic groups showed mixed ancestry, particularly cluster 2 (Fig. 2a). This can be due to gene flow among the groups and/or incomplete sorting lineage, which is expected in populations that diverged recently. As a small-seeded species occurring in disturbed habitats that were likely recolonized since the Last Glacial Maximum, both incomplete lineage assortment and gene flow are reasonable explanations for our observed lack of geographic patterning.

A set of 83 *L. orientalis* accessions was analysed by Alo *et al*. (2011), sequencing 22 conserved genes. The accessions from *L. orientalis* were clustered into three genetic groups and one of them was geographically isolated and distributed in Central Asia, mainly in Turkmenistan, Uzbekistan, Kyrgyzstan, and
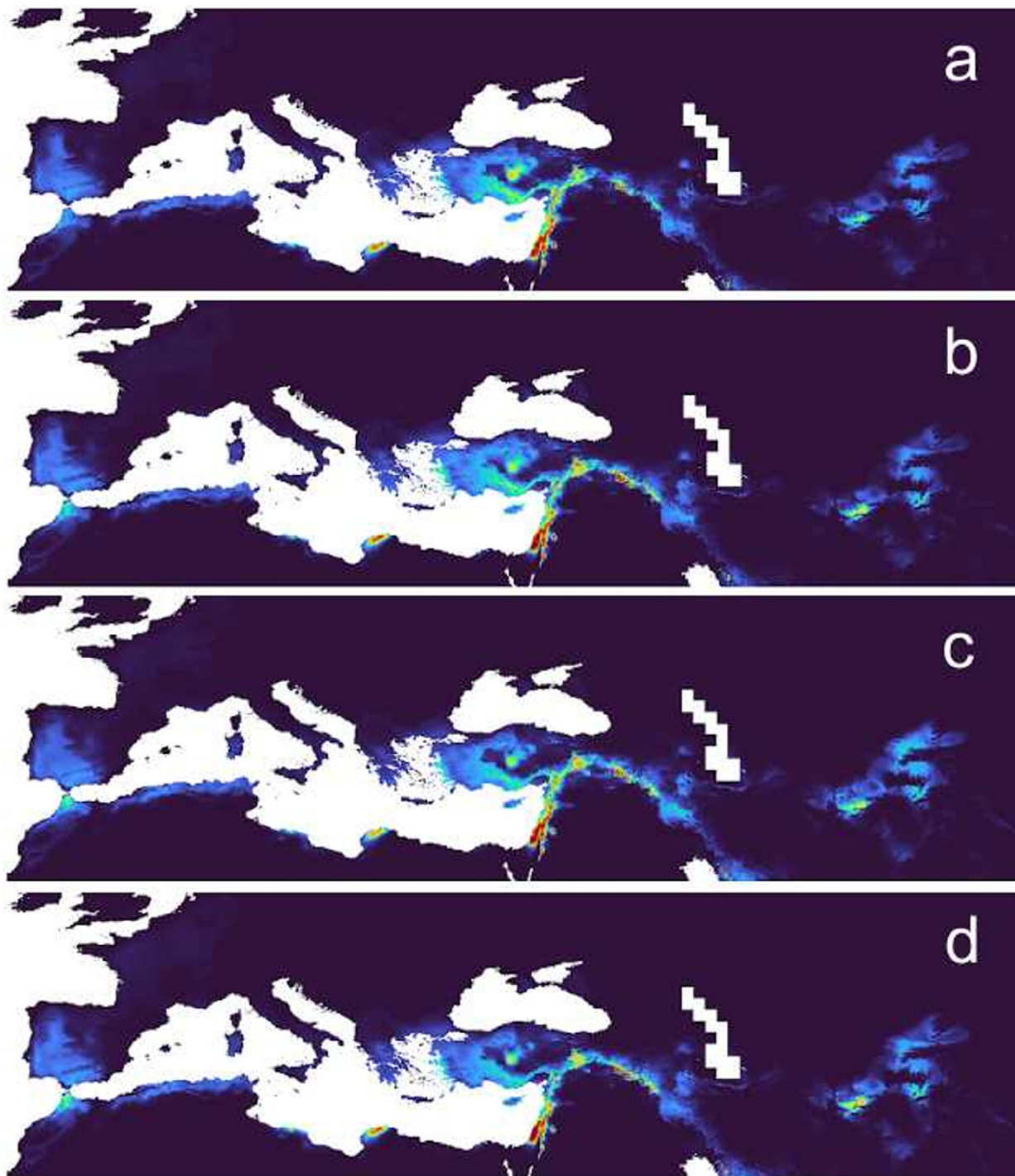
Fig. 6. Potential distribution of *Lens orientalis* for future conditions projected for 2050 based on a: RCP 4.5 and b: RCP 6.0. Projection for 2070 based on c: RCP 4.5 and d: RCP 6.0. The coloration is a heat map, with warmer (redder) colours indicating higher climatic suitability, and cooler (bluer) colours indicating lower climatic suitability.

Tajikistan. No accessions identified as *L. orientalis* in this study presented a distribution in Central Asia (Fig. 1) but the estimated potential distribution showed that there are intermediate probabilities of occurrence in that area (Fig. 5) that increase when only the 124 accessions from this study are used to construct the potential distribution (Figure S7). A gap in our sampling could explain why we did not find any *L. orientalis* in

Central Asia. Another explanation could be a misidentification of the accessions used by Alo *et al.* (2011). The low genetic distance of this *L. orientalis* cluster to *L. culinaris* accessions, even lower than to the other *L. orientalis* clusters (Alo *et al.* 2011), might support the hypothesis of misclassification. More recently, Dissanayake *et al.* (2020) found a weak correlation between geographic origin and genetic relationships in *Lens* species, similar
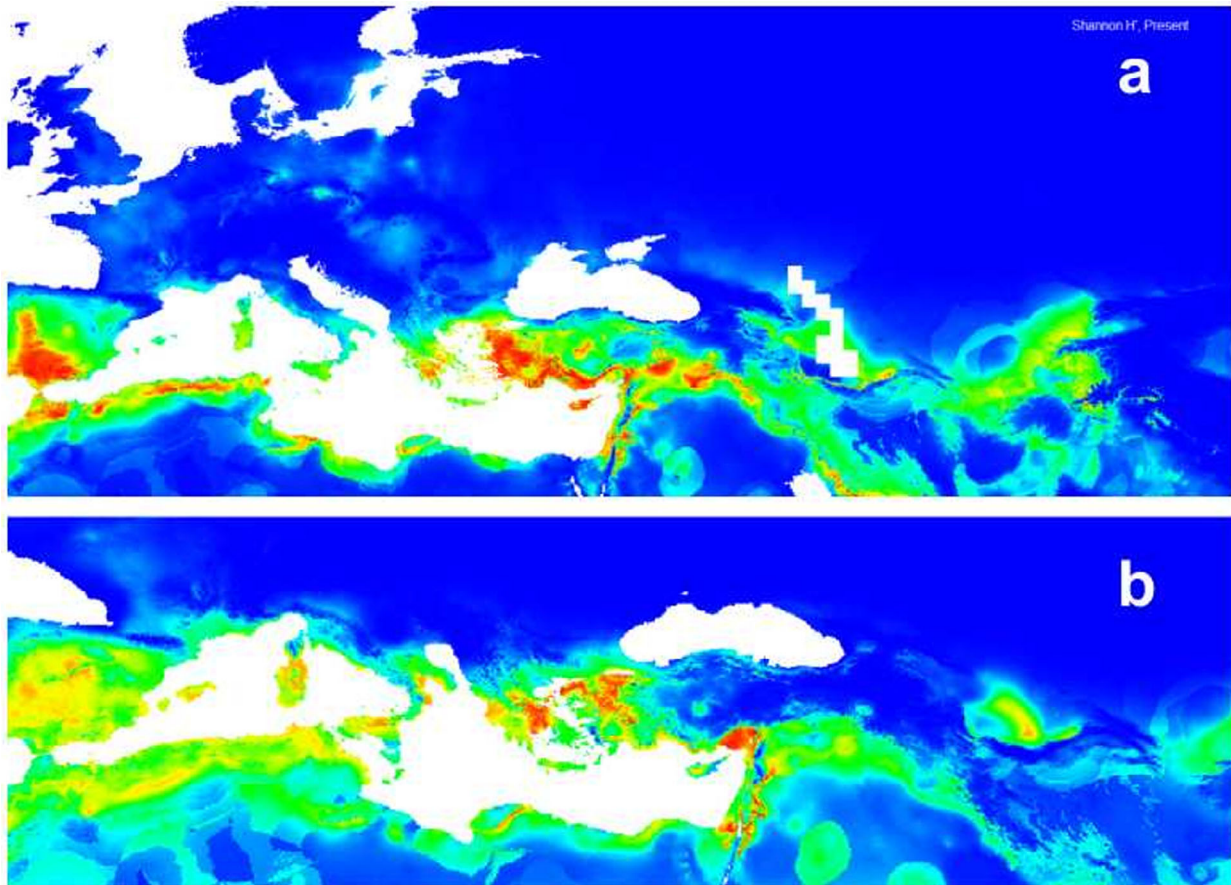
Fig. 7. Geographical distribution of Shannon's diversity index for a: Present conditions and b: LGM calculated from the predictions of the Maxent models. The coloration is a heat map, with warmer (redder) colours indicating higher climatic suitability, and cooler (bluer) colours indicating lower climatic suitability.

to our results. However, most of the accessions included by Dissanayake *et al.* (2020) were domesticated lentils and no population structure, genetic diversity and differentiation were assessed for the wild lentils. Lentil species have large (~4 Gb) and complex genomes (Ramsay *et al.* 2021). The genotyping method used for this study (GBS) covers a small proportion of the lentil genome but enough to infer the population differentiation and genetic diversity of the wild lentil accessions used in this study. Nevertheless, the discovery of genes involved in important traits or correlated with the environment requires a genotyping approach that allows a higher coverage across the genome, as has been shown in the model legume *Medicago truncatula* across the Mediterranean region (Burgarella *et al.* 2016; Renzi *et al.* 2020).

We have performed the largest population genetics analysis of the closest wild relative of the domesticated lentil, finding similar levels of genetic diversity in all genetic clusters of *L. orientalis*. Extensive exploration of the genetic variation found in the other wild lentil species remains pending. The examination of CWR is key to understanding the domestication of crops and CWRs are a critical pool of genetic diversity for crop improvement, particularly for traits including disease resistance, abiotic stress tolerance, and adaptation to changing climates (Smýkal *et al.* 2015; Dempewolf *et al.* 2017; Coyne *et al.* 2020).

We note that a possible shortcoming of our study, as well as others in lentil, is that ICARDA and most national genebanks made their wild lentil collections by taking many seeds from a population, putting them in a bag, and taking them back to the genebank. The seeds from this bag would be labelled as one accession, with the initial understanding that it is possibly heterogeneous. Later amplification might use many of these seeds to increase the accession and make it available for distribution. However, consistent methods have not been used between genebanks in maintaining population size, meaning that diversity is inherently lost during accession curation, and that different genebanks with the same original collection may have had different levels of loss of the original diversity due to different subsampling or effective population sizes. Our use of one seed to represent an accession, although consistent with past wild lentil diversity studies (*e.g.*, Alo *et al.* 2011; Wong *et al.* 2015; Dissanayake *et al.* 2020; Liber *et al.* 2021), could overlook within accession diversity. Recent collections of other wild cool season legumes, like that described in von Wettberg *et al.* (2018) intentionally overcome this limitation. To the best of our knowledge, similar work has not yet been permitted in wild *Lens*.

Harnessing diversity based on niche models and its implications for breeding and germplasm conservation

An important way to explore spatial patterns of crop wild relatives is niche modelling. Niche modelling can show predicted distributions based on presence-only occurrences. This approach is important for identifying gaps in collections, as

well as material that may have useful adaptation traits such as drought tolerance. In some crops species distribution alone has been used to identify potentially useful germplasm utilizing a Focused Identification of Germplasm Strategy (FIGS) approach (Khazaei *et al.* 2013). The potential of wheat progenitors was recently reviewed in relation to the climate change scenario, especially considering allelic variation for flowering time, cold and drought tolerance, and possibly also photosynthesis rate (Leigh *et al.* 2022).

Wild lentil plants grow in shallow stony habitats where they form tiny populations with a small number of individuals per site together with other annual legumes. They are generally poor competitors and highly palatable to grazing animals (Ferguson & Erskine 2001). They are found predominantly in primary, ungrazed habitats where they are not subject to competition by aggressive colonizer plants like grasses. The presumed distribution of *L. orientalis* is from Greece in the west to Tajikistan and southern Kirghizia in the east, from the Crimean Peninsula in the north to Jordan in the south (Ladizinsky 1979; Cubero 1981; Ladizinsky & Abbo 2015). The current analysis was based on available germplasm stored accessions, retrieved after a thorough scrutiny of worldwide collections. Although the distribution of *L. orientalis* covers a great area, where lentil has also been cultivated for a long time, our niche modelling results suggested high probabilities of occurrence restricted to the Fertile Crescent, in the region of Turkmenistan, in parts of the region of Cyrenaica and the Maghreb (west North Africa) and the Strait of Gibraltar (Fig. 6a).

Concerning the projections of the ecological niche model (ENM) into the last glacial maxima (LGM), the whole species and genetic groups 2, 3 and 4 presented a broader potential distribution along the Mediterranean Basin (Figure S7), where the climatic conditions were colder and drier than at present. Genetic cluster 2 had a broader distribution both in the present and during the LGM. The distribution of cluster 1 was the most restricted. The distribution of all genetic clusters seemed to have shrunk since the LGM (Figure S7). The lack of geographic patterns in the genetic clusters might be at least partially explained by the changes in the distribution of the genetic groups. The diversity patterns of the wild lentil genetic clusters are likely due to historic processes, during which the distribution of the species and its populations has changed. These processes are likely to be multifactorial, influenced by habitat change, natural and anthropogenic disturbance, disease and herbivory and are therefore not captured by broad-scale climatic changes. Furthermore, the self-pollination system might facilitate that the populations remain genetically differentiated. The broader potential distribution of wild lentil during the LGM might suggest a cold tolerance character of *L. orientalis*. Therefore, we cannot exclude a westwards broader distribution of the species and its specific genetic groups during the glaciations in the Pleistocene, followed by an eastwards retroflexion during the warmer Holocene. It is worth mentioning that the potential genetic centers of the species (the areas of high potential co-occurrence of genetic groups) in the Holocene and in future projections remain around the Mediterranean Basin (Fig. 6, Figure S8). The cold tolerance found in *L. orientalis* could be used for *L. culinaris* breeding programs, noting accessions from the genetic cluster 1 did not present this trait. Quantitative Trait Loci (QTL) have been identified in

cultivated lentils (Kahraman *et al.* 2004), which can also be utilized in breeding cold-tolerant cultivars, but further research is required.

The lack of accessions of *L. orientalis* from the different regions where high probabilities of occurrence were estimated showed that substantial genetic diversity is currently not available in *ex-situ* conserved collections and that it is incompletely sampled. In the framework of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), there is an urgent need for the development of both great *ex situ* and *in situ* conservation of wild lentil. For example, Alo *et al.* (2011) identified a genetic cluster distributed in Central Asia and we detected intermediate probabilities of occurrence in that area. However, no *L. orientalis* accession that we genotyped was distributed in Central Asia. Until the 1970s, there was no wild *Lens* conserved *ex situ* (Ladizinsky & Abbo 2015). The gap not only limits its use but also renders precious genetic diversity inaccessible and vulnerable to extinction (Harlan 1976). Although there are 58,405 *Lens* accessions held in various genebanks (FAO 2010), only 1% are wild accessions (Coyne *et al.* 2020). The gap in the gene bank collections might even be aggravated due to species misclassification. These observations show the importance of ongoing collections and *in situ* conservation for our future capacity to harness the genetic variation of wild lentil.

CONCLUDING THOUGHT

Our results show considerable single nucleotide diversity within *L. orientalis*, the closest wild relative of cultivated lentil. Analysis of the spatial pattern of this diversity shows the need for expanded collections in regions of high predicted suitability, where collections we were able to access are sparser. Although we did not observe marked relationships between genetic diversity and climatic factors, the genetic diversity in *L. orientalis* is essential to breeding cultivated lentils for resilience against shifting climatic conditions. The high rates of misidentification or mislabeling of wild *Lens* accessions held in germplasm show the difficulties of identifying wild lentils. Altogether this shows a need for expanded resources to both maintain and expand existing collections of wild *Lens*.

ACKNOWLEDGEMENT

AUTHOR CONTRIBUTIONS

EvW and PS conceived the study and led the discussion and writing of the manuscript. AGG and OT performed the genetic analysis and with SP contributed to the paper writing. JBrus, JBerger and JPR retrieved the environmental data and analysed it. SK performed biological material analysis. MB and SP

performed the ecological niche modelling analysis. CJC provided USDA biological material and conducted the DNA isolation. AC and RV performed the GSB sequencing. CTC and KEB GBS conducted the mapping against the lentil reference genome.

CONFLICT OF INTEREST

Author O. Trněný is employed by Agriculture Research Ltd., Czechia. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The genetic data can be retrieved from https://osf.io/5acus/?view_only=c6c7341719744b7bb8eefe7a51041bba.

SUPPORTING INFORMATION

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

Appendix S1. Supporting information.

Table S1. *Lens* accessions genotyped for this study and that were kept after filtering. The genetic classification is based on the results observed in this work. Eight accessions for which high levels of missing sequence data were present are excluded.

Table S2. Information including data source, location, and bioclimatic variables for the 710 locations used for environmental niche modeling.

Table S3. Information including source, location, and bioclimatic variables for the 195 accessions that were sequenced. Bioclimatic information was not gathered for accessions that were not *Lens orientalis*.

Table S4. Information including source, location, and bioclimatic variables for the 51 accessions from Wong *et al.* 2015 that were used to determine if accessions labelled as *Lens orientates* from genebanks were in face *Lens orientalis*. Bioclimatic information was not gathered for these accessions.

Figure S1. Error probability estimated by Admixture according to different genetic clusters.

Figure S2. Ancestry proportions according to different values of *k*. The order of the accessions is kept across the different ancestry plots.

Figure S3. Geographic distributions of the *Lens orientalis* accessions, based on clustering them with STRUCTURE into either two genetic clusters ($k = 2$) or five ($k = 5$) clusters. Each point is colored by the genetic group to which they have been assigned.

Figure S4. (a) Components of genetic diversity (F_{is}, the inbreeding coefficient; H_o, observed heterozygosity; and H_e, expected heterozygosity) and (b) differentiation (F_{st}) with its confidence interval among the four genetic cluster identified in the *L. orientalis* accessions.

Figure S5. Relationship between genetic, environmental and geographical distances in the 124 *L. orientalis* accessions. R corresponds to the regression coefficient.

Figure S6. Predicted suitability values for *Lens orientalis* during the LGM (a), the present (b), predicted for 2050 (c) and for 2070 (d) that correspond to the elevation classes below and above 700 m. *Lens orientalis* locations are based on our own surveys as well as observations in GBIF and in several genebanks.

Figure S7. Predicted potential distribution of *L. orientialis* using the 124 genotyped accessions of *L. orientalis* and the four genetic groups found. The left column shows present conditions and climatic projections of the Last Glacial Maximum (LGM) are in the right column.

Figure S8. Geographical distribution of Shannon's diversity index for the genetic groups as calculated from the predictions of the Maxent models. Maps show predictions for future conditions.

REFERENCES

Abbo S., Gopher A. (2020) Plant domestication in the Neolithic Near East: the humans-plants liaison. *Quaternary Science Reviews*, **242**, 106412.

Alexander D.H., Novembre J., Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Allaby R.G., Stevens C.J., Kistler L., Fuller D.Q. (2022) Emerging evidence of plant domestication as a landscape-level process. *Trends in Ecology & Evolution*, **37**, 268–279.

Alo F., Furman B.J., Akhunov E., Dvorak J., Gepts P. (2011) Leveraging genomic resources of model species for the assessment of diversity and phylogeny in wild and domesticated lentil. *The Journal of Heredity*, **102**, 315–329.

Asghar M.J., Hameed A., Khan M.R., Shahid M., Atif R.M. (2021) Lentil wild genetic resource: a potential source of genetic improvement for biotic and abiotic stress tolerance, *Wild germplasm for genetic improvement in crop plants*. Academic Press, London, UK.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Burgarella C., Chantret N., Gay L., Prosperi J.-M., Bonhomme M., Tiffin P., Young N.D., Ronfort J. (2016) Adaptation to climate through flowering phenology: a case study in *Medicago truncatula*. *Molecular Ecology*, **25**, 3397–3415.

Coyne C.J., Kumar S., Wettberg E.J.B., Marques E., Berger J.D., Redden R.J., Ellis T.H.N., Brus J., Zablatzká L., Smýkal P. (2020) Potential and limits of exploitation of crop wild relatives for pea, lentil, and chickpea improvement. *Legume Science*, **2**, e36.

Cubero J.I. (1981) Origin, taxonomy and domestication [lentils]. In: Webb C., Hawtin G. (Eds), *Lentils*. CABI, London, UK, pp 15–38.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dempewolf H., Baute G., Anderson J., Kilian B., Smith C., Guarino L. (2017) Past and future use of wild relatives in crop breeding. *Crop Science*, **57**, 1070–1082.

Dikshit H.K., Singh A., Singh D., Aski M.S., Prakash P., Jain N., Meena S., Kumar S., Sarker A. (2015) Genetic diversity in lens species revealed by EST and genomic simple sequence repeat analysis. *PLoS One*, **10**, e0138101.

Dissanayake R., Braich S., Cogan N.O.I., Smith K., Kaur S. (2020) Characterization of genetic and allelic diversity amongst cultivated and wild lentil accessions for germplasm enhancement. *Frontiers in Genetics*, **11**, 546.

Dixon P. (2003) VEGAN, a package of R functions for community ecology. *Journal of vegetation science:*

official organ of the International Association for Vegetation Science, **14**, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Elith J., Graham C.H., Anderson R.P., Dudík M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M.C.M., Townsend Peterson A., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberón J., Williams S., Wisz M.S., Zimmermann N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Euro+Med (2006) Euro+Med PlantBase – the information resource for Euro-Mediterranean plant diversity. https://europlusmed.org/

FAO (2010)FAOSTAT. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat [19th September 2022].

Ferguson M., Erskine W. (2001) Lentils (*Lens* L.). In: Maxted N., Bennett S.J. (Eds), *Plant genetic resources of legumes in the Mediterranean*. Springer, Dordrecht, pp 125–133.

Ferguson M.E., Ford-Lloyd B.V., Robertson L.D., Maxted N., Newbury H.J. (1998a) Mapping the geographical distribution of genetic variation in the genus lens for the enhanced conservation of plant genetic diversity. *Molecular Ecology*, **7**, 1743–1755.

Ferguson M.E., Robertson L.D., Ford-Lloyd B.V., Newbury H.J., Maxted N. (1998b) Contrasting genetic variation amongst lentil landraces from different geographical origins. *Euphytica*, **102**, 265–273.

Fick S.E., Hijmans R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.

Gela T.S., Adobor S., Khazaei H., Vandenberg A. (2021) An advanced lentil backcross population developed from a cross between *Lens culinaris* × *L. ervoides* for future disease resistance and genomic studies. *Plant Genetic Resources: Characterization and Utilization*, **19**, 167–173.

Gorelick R., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. (2017) Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, **202**, 18–27.

Goudet J. (2005) Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Guerra-Garcia A., Haile T., Ogutcen E., Bett K.E., von Wettberg E.J. (2022) An evolutionary look into the history of lentil reveals unexpected diversity. *Evolutionary Applications*, **15**, 1313–1325.

Gutierrez-Gonzalez J.J., García P., Polanco C., González A.I., Vaquero F., Vences F.J., Pérez de la Vega M., Sáenz de Miera L.E. (2022) Multi-species transcriptome assemblies of cultivated and wild lentils (Lens sp.) provide a first glimpse at the lentil pangenome. *Agronomy*, **12**, 1619.

Hajjar R., Hodgkin T. (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica*, **156**, 1–13.

Harlan J.R. (1976) Genetic resources in wild relatives of crops. *Crop Science*, **16**, 329–333.

Hellwig T., Abbo S., Ophir R. (2022) Drivers of genetic differentiation and recent evolutionary history of an Eurasian wild pea. *Journal of Biogeography*, **49**, 794–808.

Hellwig T., Abbo S., Sherman A., Coyne C.J., Saranga Y., Lev-Yadun S., Main D., Zheng P., Ophir R. (2020) Limited divergent adaptation despite a substantial environmental cline in wild pea. *Molecular Ecology*, **29**, 4322–4336.

Hellwig T., Abbo S., Sherman A., Ophir R. (2021) Prospects for the natural distribution of crop wild-relatives with limited adaptability: the case of the wild pea *Pisum fulvum*. *Plant Science*, **310**, 110957.

Hernandez P.A., Graham C.H., Master L.L., Albert D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.

Herten K., Hestand M.S., Vermeesch J.R., Van Houdt J.K.J. (2015) GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics*, **16**, 73.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Limatology*, **25**, 1965–1978. https://doi.org/10.1002/joc.1276

Hübner S., Kantar M.B. (2021) Tapping diversity from the wild: from sampling to implementation. *Frontiers in Plant Science*, **12**, 626565.

Kahraman A., Kusmenoglu I., Aydin N., Aydogan A., Erskine W., Muehlbauer F.J. (2004) QTL mapping of winter hardiness genes in lentil. *Crop Science*, **44**, 13–22.

Kaisers W., Schaal H., Schwender H. (2015) Rbamtools: an R interface to samtools enabling fast accumulative tabulation of splicing events over multiple RNA-seq samples. *Bioinformatics*, **31**, 1663–1664.

Khazaei H., Caron C.T., Fedoruk M., Diapari M., Vandenberg A., Coyne C.J., McGee R., Bett K.E. (2016) Genetic diversity of cultivated lentil (*Lens culinaris* Medik.) and its relation to the World's agro-ecological zones. *Frontiers in Plant Science*, **7**, 1093.

Khazaei H., Street K., Bari A., Mackay M., Stoddard F.L. (2013) The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One*, **8**, e63107.

Koul P.M., Sharma V., Rana M., Chahota R.K., Kumar S., Sharma T.R. (2017) Analysis of genetic structure and interrelationships in lentil species using morphological and SSR markers. *3 Biotech*, **7**, 83.

Ladizinsky G. (1979) The origin of lentil and its wild genepool. *Euphytica*, **28**, 179–187.

Ladizinsky G., Abbo S. (2015) The *Lens* genus. In: Ladizinsky G., Abbo S. (Eds), *The search for wild relatives of cool season legumes*. Springer International, Cham, Switzerland, pp 1–28.

Larson G., Piperno D.R., Allaby R.G., Purugganan M.D., Andersson L., Arroyo-Kalin M., Barton L., Vigueira C.C., Denham T., Dobney K., Doust A.N., Gepts P., Gilbert M.T.P., Gremillion K.J., Lucas L., Lukens L., Marshall F.B., Olsen K.M., Pires J.C., Richerson P.J., de Casas R.R., Sanjur O.I., Thomas M.G., Fuller D.Q. (2014) Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 6139–6146.

Lawrence D., Palmisano A., de Gruchy M.W. (2021) Collapse and continuity: a multi-proxy reconstruction of settlement organization and population trajectories in the Northern Fertile Crescent during the 4.2kya rapid climate change event. *PLoS One*, **16**, e0244871. https://doi.org/10.1371/journal.pone.0244871

Leamy L.J., Lee C.-R., Song Q., Mujacic I., Luo Y., Chen C.Y., Li C., Kjemtrup S., Song B.-H. (2016) Environmental versus geographical effects on genomic variation in wild soybean (*Glycine soja*) across its native range in Northeast Asia. *Ecology and Evolution*, **6**, 6332–6344.

Leigh F.J., Wright T.I.C., Horsnell R.A., Dyer S., Bentley A.R. (2022) Progenitor species hold untapped diversity for potential climate-responsive traits for use in wheat breeding and crop improvement. *Heredity*, **128**, 291–303.

Liber M., Duarte I., Maia A.T., Oliveira H.R. (2021) The history of lentil (*Lens culinaris subsp. culinaris*) domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. *Frontiers in Plant Science*, **12**, 628439.

Mohar S., Kumar S., Basandrai A.K., Basandrai D., Malhotra N., Saxena D.R., Gupta D., Sarker A., Singh K. (2020) Evaluation and identification of wild lentil accessions for enhancing genetic gains of cultivated varieties. *PLoS One*, **15**, e0229554. https://doi.org/10.1371/journal.pone.0229554

Ogutcen E., Ramsay L., von Wettberg E.B., Bett K.E. (2018) Capturing variation in lens (Fabaceae): development and utility of an exome capture array for lentil. *Applications in Plant Sciences*, **6**, e01165. https://doi.org/10.1002/aps3.1165

van Oss R., Abbo S., Eshed R., Sherman A., Coyne C.J., Vandemark G.J., Zhang H.-B., Peleg Z. (2015) Genetic relationship in Cicer Sp. expose evidence for geneflow between the cultigen and its wild progenitor. *PLoS One*, **10**, e0139789.

Peakall R., Smouse P.E. (2006) Genalex 6: genetic analysis in excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Phillips S.J., Anderson R.P., Dudík M., Schapire R.E., Blair M.E. (2017) Opening the black box: an open-source release of Maxent. *Ecography*, **40**, 887–893.

Phillips S.J., Anderson R.P., Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pokharia A.K., Srivastava C. (2013) Current status of archaeobotanical studies in Harappan civilization: an archaeological perspective. *Heritage: Journal of Multidisciplinary Studies in Archaeology*, **1**, 118–137.

Price M.N., Dehal P.S., Arkin A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, **26**, 1641–1650.

Ramsay L., Koh C.S., Kagale S., Gao D., Kaur S., Haile T., Gela T.S., Chen L.-A., Cao Z., Konkin D.J., Toegelová H., Doležel J., Rosen B.D., Stonehouse R., Humann J.L., Main D., Coyne C.J., McGee R.J., Cook D.R., Banniza S., Edwards D., Bayer P.E., Batley J., Udupa S.M., Bett K.E. (2021) Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. *bioRxiv*. 2021.07.23.453237. https://doi.org/10.1101/2021.07.23.453237.

Renzi J.P., Duchoslav M., Brus J., Hradilová I., Pechanec V., Václavek T., Machalová J., Hron K., Verdier J., Smýkal P. (2020) Dormancy release in *Medicago truncatula* seeds is related to environmental variations. *Plants*, **9**, 503.

Rodriguez M., Rau D., Bitocchi E., Bellucci E., Biagetti E., Carboni A., Gepts P., Nanni L., Papa R., Attene

G. (2016) Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytologist*, **209**, 1781–1794.

Roskov Y.R., Bisby F.A., Zarucchi J.L., Schrire B.D., White R.J. (Eds) (2010) *ILDIS world database of legumes: draft checklist, version 10*. ILDIS, Reading, UK.

Sedivy E.J., Wu F., Hanzawa Y. (2017) Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytologist*, **214**, 539–553.

Sedlazeck F.J., Rescheneder P., von Haeseler A. (2013) NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, **29**, 2790–2791.

Singh M., Sharma S.K., Singh B., Malhotra N., Chandora R., Sarker A., Singh K., Gupta D. (2018) Widening the genetic base of cultivated gene pool following introgression from wild *lens* taxa. *Plant Breeding*, **137**, 470–485.

Smouse P.E., Long J.C., Sokal R.R. (1986) Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Smýkal P., Coyne C.J., Ambrose M.J., Maxted N., Schaefer H., Blair M.W., Berger J., Greene S.L., Nelson M.N., Besharat N., Vymyslický T., Toker C., Saxena R.K., Roorkiwal M., Pandey M.K., Hu J., Li Y.H., Wang L.X., Guo Y., Qiu L.J., Redden R.J., Varshney R.K. (2015) Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews in Plant Sciences*, **34**, 43–104.

Smýkal P., Hradilová I., Trněný O., Brus J., Rathore A., Bariotakis M., Das R.R., Bhattacharyya D., Richards

C., Coyne C.J., Pirintsos S. (2017) Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports*, **7**, 17384.

Sonnante G., Hammer K., Pignone D. (2009) From the cradle of agriculture a handful of lentils: history of domestication. *Rendiconti Lincei*, **20**, 21–37.

Thormann I., Reeves P., Reilley A., Engels J.M.M., Lohwasser U., Börner A., Pillen K., Richards C.M. (2016) Geography of genetic structure in barley wild relative *Hordeum vulgare* subsp. *spontaneum* in Jordan. *PLoS One*, **11**, e0160745.

Toker C., Berger J., Eker T., Sari D., Sari H., Gokturk R.S., Kahraman A., Aydin B., von Wettberg E.J. (2021) *Cicer turcicum*: a new Cicer species and its potential to improve chickpea. *Frontiers in Plant Science*, **12**, 662891.

Tullu A., Buchwaldt L., Lulsdorf M., Banniza S., Barlow B., Slinkard A.E., Sarker A., Tar'an B., Warkentin T., Vandenberg A. (2006) Sources of resistance to anthracnose (*Colletotrichum truncatum*) in wild lens species. *Genetic Resources and Crop Evolution*, **53**, 111–119.

Vavilov N. (1926) *Studies on the origin of cultivated plants*. Bulletin of Applied Botany and Plant Breeding, Leningrad, Russia.

van Vuuren D.P., Stehfest E., den Elzen M.G.J., Kram T., van Vliet J., Deetman S., Isaac M., Klein G.K., Hof A., Mendoza B.A., Oostenrijk R., van Ruijven B. (2011) RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C. *Climatic Change*, **109**, 95–116.

Warschefsky E., Penmetsa R.V., Cook D.R., von Wettberg E.J.B. (2014) Back to the wilds: tapping

evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany*, **101**, 1791–1800.

von Wettberg E.J.B., Chang P.L., Başdemir F., Carrasquila-Garcia N., Korbu L.B., Moenga S.M., Bedada G., Greenlon A., Moriuchi K.S., Singh V., Cordeiro M.A., Noujdina N.V., Dinegde K.N., Shah Sani S.G.A., Getahun T., Vance L., Bergmann E., Lindsay D., Mamo B.E., Warschefsky E.J., Dacosta-Calheiros E., Marques E., Yilmaz M.A., Cakmak A., Rose J., Migneault A., Krieg C.P., Saylak S., Temel H., Friesen M.L., Siler E., Akhmetov Z., Ozcelik H., Kholova J., Can C., Gaur P., Yildirim M., Sharma H., Vadez V., Tesfaye K., Woldemedhin A.F., Tar'an B., Aydogan A., Bukun B., Penmetsa R.V., Berger J., Kahraman A., Nuzhdin S.V., Cook D.R. (2018) Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications*, **9**, 649.

Wong M.M.L., Gujaria-Verma N., Ramsay L., Yuan H.Y., Caron C., Diapari M., Vandenberg A., Bett K.E. (2015) Classification and characterization of species within the genus lens using genotyping-by-sequencing (GBS). *PLoS One*, **10**, e0122025.

Zheng X., Levine D., Shen J., Gogarten S.M., Laurie C., Weir B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

Zohary D., Hopf M. (1973) Domestication of pulses in the Old World: legumes were companions of wheat and barley when agriculture began in the near east. *Science*, **182**, 887–894.