scientific reports

Check for updates

OPEN

Unlocking genetic diversity for low-input systems in a changing climate through participatory characterization and GWAS of lentil landraces

Elisa Lorenzetti, Mercy Wairimu Macharia, Svenja Mager, Matteo Dell'Acqua, Stefano Carlesi✉ & Paolo Bàrberi

Lentils are a vital staple crop in a world seeking sustainable and secure food, but their cultivation face a threat due to yield instability, mainly arising from a lack of genetic diversity in breeding programmes. In this study, we assembled and characterized the genetic and phenotypic diversities of a collection of 106 lentil genotypes, to evaluate their breeding and cropping potential. Lentil landraces from Italy and beyond, either abandoned or still cultivated, were collected from genebanks, seed savers, universities and farmers. We characterized their phenotypic diversity with an augmented block design, using a control plot enabling a spatial analysis. We phenotyped the collection during two cropping seasons for its agronomic performance, involving local practitioners in a participatory variety evaluation. Meanwhile, we genotyped the landrace collection with a DNA sequencing approach, obtaining 91,136 high quality single nucleotide polymorphisms (SNPs). We used SNPs to describe the phylogenetic relation among landraces, unveiling their uniqueness, and combined SNP data with measured traits to conduct a genome-wide association study (GWAS) that led to the identification of 32 unique marker-trait associations highlighting lentil genomic loci related with adaptation and performance. The results of this study offer new tools to unlock agrobiodiversity for lentil breeding in the Mediterranean, towards the identification of genetic factors responsible for traits of agronomic interest and providing possible sources of parental material.

In the last decade, grain legumes have gained ever increasing attention by the international research community due to their crucial role in crop rotations, human nutrition and livestock feeding, and for being a source of sustainable proteins[1,2,3].

In recent years, grain legumes have experienced significant growth in Italy, with chickpea production increasing by 380% and lentil production by 320% from 2006 to 2021, according to ISTAT. They are also being promoted through quality labels and innovative culinary offerings. However, despite of their recent success, grain legumes still represent a minor crop sector in Italy (1.7% of total Utilized Agricultural Area) and in Europe (1.5% of total arable land)[4], where they are often grown in marginal lands and largely neglected by genetic improvement. As a result, European and worldwide farmers report difficulties in cultivation and yield instability, which hinder the expansion of these crops[5].

Within pulses, lentil (*Lens culinaris*) is a species with great untapped potential. It is a diploid (2n = 14), self-pollinating, annual crop at the foundation of the Italian and Mediterranean food culture. Its domestication is dated back to 11,000 BC[6–8]and its long-lasting cultivation history in the region resulted in the evolution of countless landraces adapted to local agroecosystems. In Italy, many of these landraces have been continuously cultivated until recent times, possibly due to their high adaptation to the microclimate of specific location[9]. Until the second half of the 20th century, Italy was amongst the top lentil producers in the Mediterranean basin[6]; however, since then lentil cultivation was progressively abandoned: to date, only 2% of national consumption is locally produced[10]. Currently, landraces' cultivation is phasing out, resulting in genetic erosion of the local lentil gene pool. Landraces have hardly undergone breeding processes and have therefore a high genetic variability and adaptability to local pedoclimatic conditions, making them a precious resource for breeding.

Institute of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, Italy. ✉email: stefano.carlesi@santannapisa.it

Despite progresses of breeding for improved cultivars, the genetic gain in grain legumes is still much lower than that potentially attainable: while cereals productivity increased by 300% in the past 50 years, that of legumes only increased by 60%, and this seems contributed by the limited genetic diversity options exploited in legume breeding programmes[11]. In addition, local landraces, traditionally cultivated in marginal lands, constitute a precious material for breeding for low input conditions: often times, the genetic material available in the seed market is not adapted to low-input agroecosystems, being cultivar selection and seed production carried out in high input systems[9].

A promising approach to exploit diversity for the selection of locally adapted material is the participatory varietal selection approach, that directly involve farmers in the evaluation of breeding materials[12] (Fig. 1). Joint cultivar selection by researchers and farmers has been proven successful due to the manifold skills and perspectives applied in the process and to the more efficient use of resources. The approach proven even more fruitful when utilized on marginal land or in low input systems, where decentralization of the selection environment is crucial to achieve proper adaptation[13].

The aim of this paper is to evaluate the phenotypic and genotypic diversity of a collection of lentil landraces reflecting Italian lentil diversity and to analyse their relationships with collection origin and farmers' preferences. By doing so, we aimed also at connecting lentil phenotypes with allelic diversity, disclosing potential for breeding. We chose to capture the diversity of Italian landraces, leveraging their extensive local adaptation, thus serving as an exemplary case study for illustrating the exploitation of traditional germplasm. However, to enrich the study, we also incorporated select representatives from other Mediterranean regions. We collected lentil landraces, either abandoned or still cultivated, from genebanks and custodian farmers, and characterized them during two cropping seasons (2020 and 2021) from a phenotypic and genetic perspective. In addition, farmers and practitioners were asked to evaluate the landraces at the end of the second-year cropping season. The combined phenotypic, genetic and farmers' evaluations data served for a genome-wide association study in order to identify genomic loci and candidate genes for adaptation and performance. The results of this study offer new tools to unlock agrobiodiversity for breeding in Mediterranean environments, identifying genetic factors responsible for traits of agronomic interest and providing possible sources of parental material for future breeding programmes.

Results

Phenotypic diversity

In 2020 the first dimension of the PCA accounted for 42% of the total variability, with the contribution of seven parameters, mainly linked to plant vitality: homogeneity of the canopy, percentage of soil cover, biomass, vigour, height at harvest, yield and lodging. The second dimension, explaining 22% of the variance, represented instead flower precocity and two parameters of pathogen sensitivity (Fig. 2).

In 2021 the first dimension explained 34% of the variance, representing the main production parameters: plant height at harvest, biomass, yield, homogeneity of the canopy and vigour. The second dimension, explaining 21% of the variance, was primarily associated with the contribution of resistance to pest and precocity parameters (Fig. 2).

Notably, in both years, parameters related to pathogen resistance and precocity were correlated and showed no correlation with yield and biomass production.

In 2020 the landraces with highest plant height came from Sicily mainland and Sardinia, ranging from 27 to 28.6 cm. In 2021, a few landraces from Apulia and Latium also exhibited notable plant heights, in addition to those observed the previous year, resulting in an increased height range of 35–39 cm. Overall, the average plant height increased from 21 to 27 cm when compared to 2020. It should be stressed that the control lentil (cv. Robin) was among the tallest in both years (Supplementary Information T1).

Concerning flower precocity, in 2020 the first flowering landraces came from Sicilian small islands, from Sicily mainland, from Sardinia, Apulia, Ethiopia and Latium. In 2021 the Ethiopian landraces resulted as the earliest flowering, together with two landraces from Sicily mainland, closely followed by those from the Sicilian small islands. In both years the control lentil did not show any tendency towards flower precocity.



Fig. 1. Participatory evaluation of the collection of lentil landraces in summer 2021.(courtesy of E. Lorenzetti).

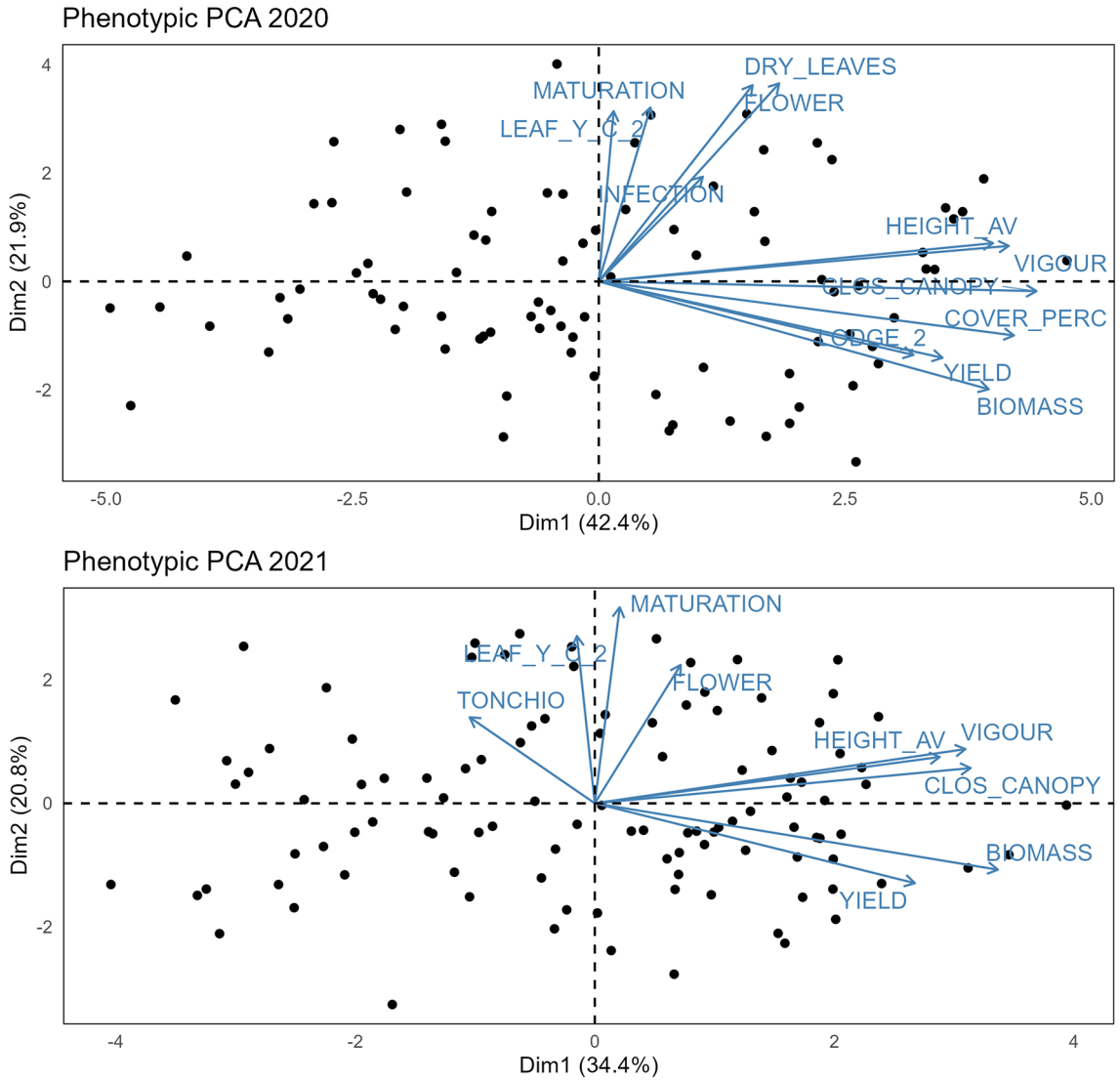


Fig. 2. Principal Component Analysis of phenotypic diversity in the lentil collection. Dots represent lentil landraces, vectors represent the explanatory magnitude and direction of phenotypic drivers of lentil differentiation.

Concerning grain yield, with the control cultivar's yield set as the baseline, we observed percentage differences in the kriged yield of the studied landraces (Fig. 3). It emerged that, in 2020, only a limited number of landraces (12.4%) produced more than the control, while in 2021 their number increased (22.4%). The trend was confirmed with actual yield means, expressed as g m^{-2} (Fig. 4).

In 2020, on the basis of kriged values (Figs. 3), 11 landraces performed better than the control, while in 2021 they were 24; only 7 landraces though are repeated over the two years: such landraces comes from Latium, Apulia, Molise, Abruzzo, Basilicata, Sicily and Marche regions.

We clustered the phenotypic PCA on the base of the landraces' origin of collection, but clusters rarely showed consistency of distribution: in 2020 only landraces from Greece and Sardinia aggregated in clear groups, while in 2021 it was the turn of landraces from Ethiopia (that could not be included in 2020 analysis due to lack of plant development) (Supplementary Information, Figure F1).

Genetic diversity

The sequencing of the lentil landrace collection yielded approximately 998 million raw reads (Supplementary Information T2). Mean mapping rate was 77.7%. We mapped the paired end reads on the *Lens culinaris* reference genome, yielding a total of 428,872 SNP markers distributed across all chromosomes and mapped on at least 75% of the samples with a heterozygosity lower than 0.8. We then retained a total of 91,136 high-quality SNPs markers with minor allele frequency greater than 5%. This set of markers was used for all subsequent analysis. Detailed alignment statistics, including the alignment rate for each accession, are provided in Supplementary Information T3. A Neighbour Joining (NJ) tree, based on these high-quality SNPs, shows that Italian landraces are highly diverse and cluster in different monophyletic groups, partially in accordance with local geographic

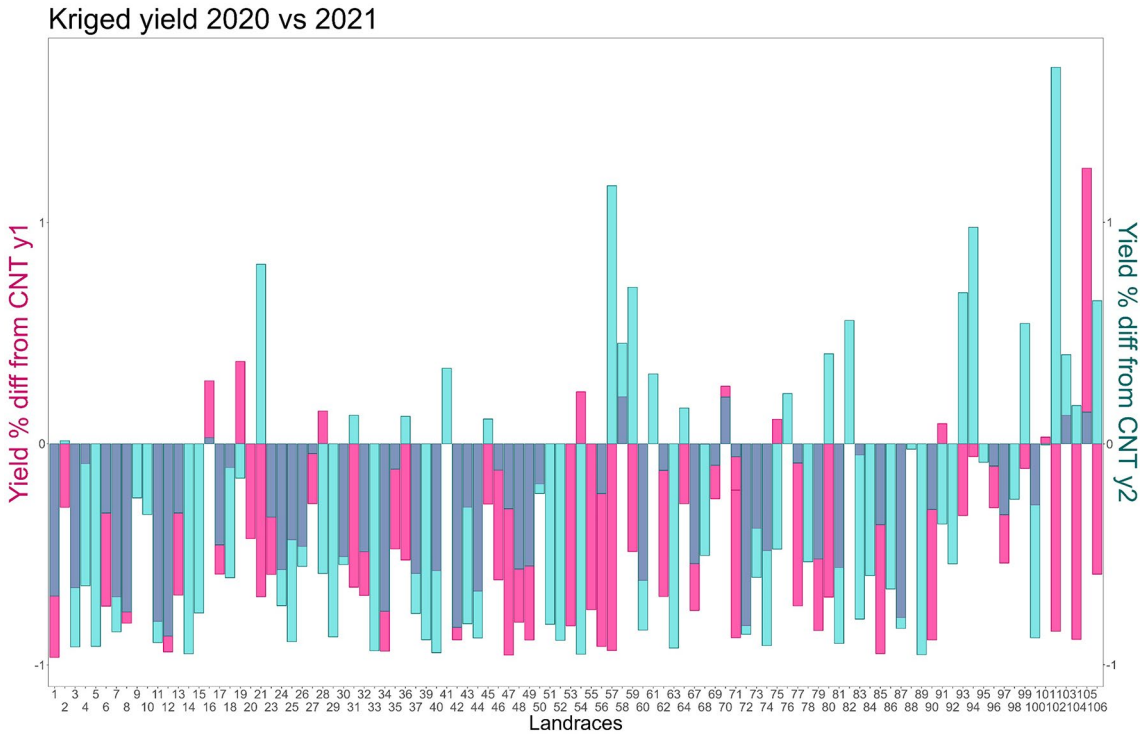


Fig. 3. Landrace grain yield expressed as percent difference from yield of the control cultivar. The percent differences recorded in 2020 and 2021 are represented with pink and blue bars, respectively: the bars may overlap. 1 and − 1 indicate + 100% and − 100% grain yield compared with the control, respectively.

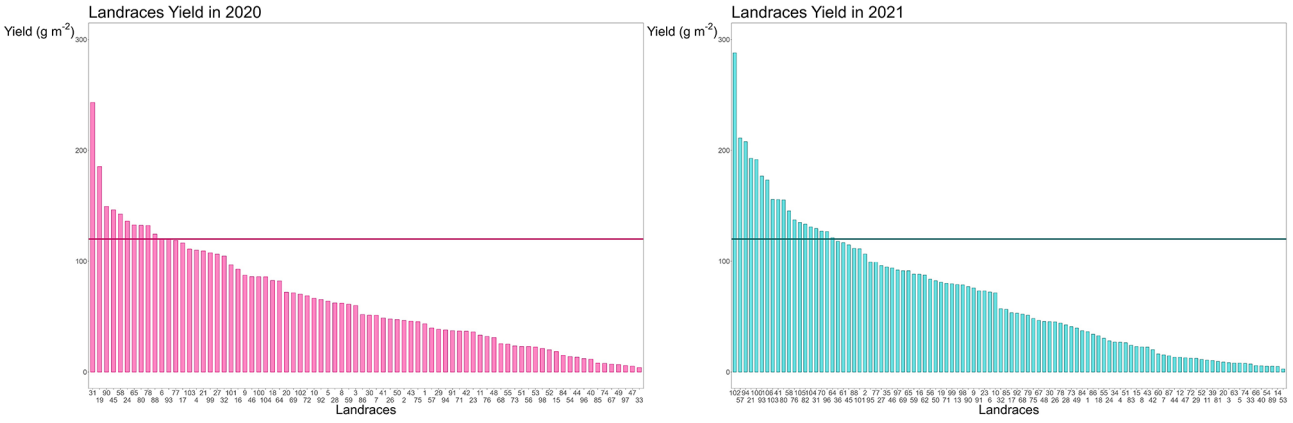


Fig. 4. Landrace yield (g m^{-2}) in 2020 (on the left) and 2021 (on the right). Average yield of the control cultivar is represented with the horizonal line (120 g m^{-2} in 2020 and 107 g m^{-2} in 2021), while average yield of landraces was 64 g m^{-2} in 2020 and 73 g m^{-2} in 2021.

distribution (Fig. 5). We then analysed the population structure of lentil landraces using a Discriminant Analysis of Principal Components (DAPC), which identified six genetic clusters. A PCA on SNP data showed these clusters (Supplementary Information, Figure F2), with PC1 and PC2 explaining 28.09% and 12.13% of the variance. Overall, the genetic clusters showed a moderate correspondence with collection origin, differently to what happened with phenotypic data (see Sect. 2.1).

Concerning the non-Italian landraces, lentils from Ethiopia formed a very coherent group (DAPC cluster 1) in 6 cases out of 8, while the remaining 2 landraces were positioned in very distant regions of the phylogenetic tree and placed in DAPC clusters 3 and 4. Lentils from Tunisia instead were integrated in a cluster of Italian landraces (from southern Italy and islands, cluster 2) without showing the attitude to create monophyletic groups. Finally, lentils from Greece were positioned partially together with Tunisian landraces in cluster 2 and partially in cluster 3, the most heterogeneous cluster (Supplementary Information T1).

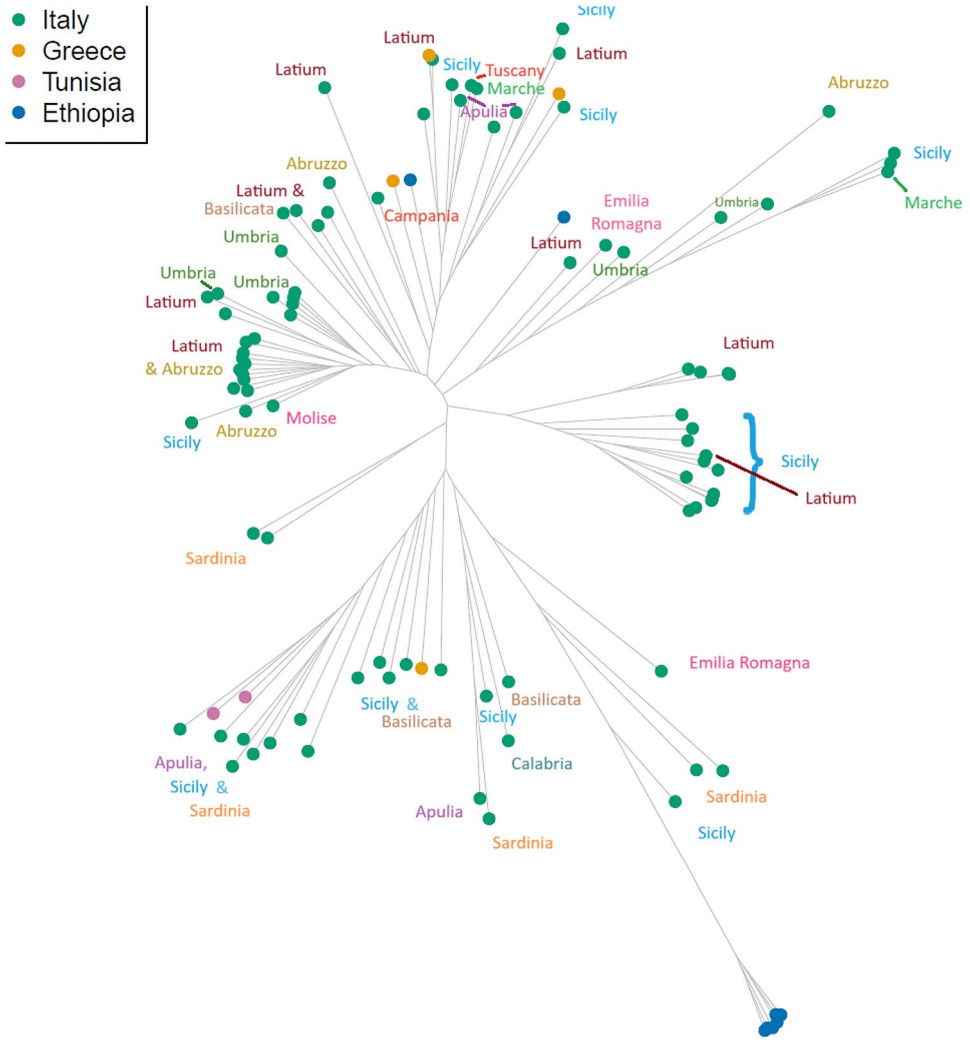


Fig. 5. Phylogenetic tree constructed from SNP-based genetic similarity matrices of ten seed DNA samples. Different dot colours represent the country of origin, each landrace is labelled with the region of collection, indicated in a distinct text colour.

Concerning Italian landraces, a few homogeneous clusters emerged: landraces from Sicilian islands (Lipari, Ustica, Linosa, Lampedusa and Pantelleria) fell in the same cluster as most of the landraces from Latium's island, Ventotene (cluster 6). Another group emerged (cluster 4) composed mainly by landraces following the distribution of the Apennine ridge: Rascino upland (Latium), Abruzzo (Santo Stefano di Sessanio in particular), Molise and Umbria. Lentils from mainland Sicily were instead spread all over the tree, even with wide distance between landraces of the same neighbourhood (e.g., Modica or Villalba). Two more clusters of landraces mainly from southern Italy and the islands emerged (clusters 2 and 3), while the remaining (cluster 5) did not show geographical coherence.

The landraces showed a different degree of heterozygosity according to the genetic clusters: cluster 6 from the small islands showed the lowest heterozygosity (0.06), closely followed by cluster 1 from Ethiopia (0.08); the highest degree was found in clusters 2 (0.19), 3 (0.16) and 5 (0.16), while cluster 4 from the Apennine ridge stayed in between (0.12). We could not discriminate whether the heterozygosity lies within the different individuals of the same landraces or within the landrace genome itself, but the distribution of heterozygosity within the genetic clusters suggests varying levels of historical gene flow and adaptation among lentil landraces from different origins.

The groups identified in the DAPC were used to generate a new clustering of the phenotypic PCA, replacing the original factor of collection origin (Fig. 6).

In 2020, clear phenotypic differences among DAPC clusters emerged, primarily driven by the second dimension, reflecting precocity and pathogen sensitivity traits. The landrace group from the Appennins and central Italy exhibited late flowering and maturation, and health traits, while the group from Sicilian and Latium islands displayed early flowering and maturation, as well as early senescence and pathogen sensitivity traits (Fig. 6).

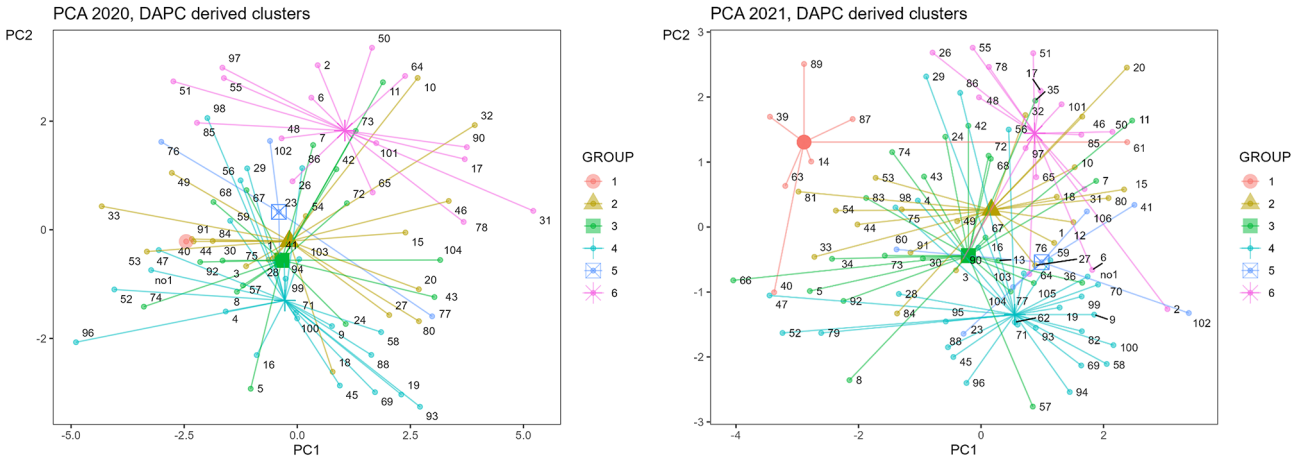


Fig. 6. Spider plot of the Principal Component Analysis of phenotypic diversity (Fig. 2): each of the genetic clusters determined through the DAPC is represented by a different colour and shape (see legend on the right: clusters' number refer to Sect. 2.2). The shapes are positioned at the median point of each cluster. On the left is displayed the PCA of 2020, on the right that of 2021.

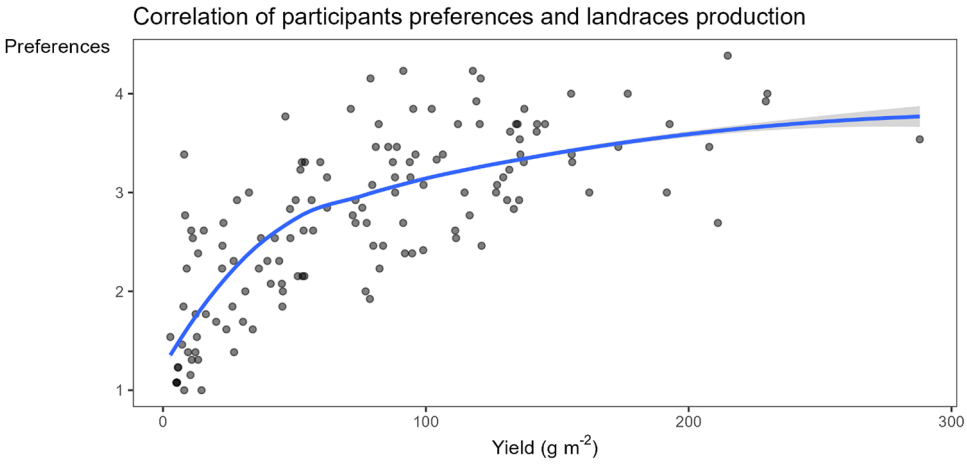


Fig. 7. Relationship between mean preference score assigned to each landrace by participants in the participatory evaluation (y axis; from 1 -not satisfactory- to 5 -satisfactory-) and the actual grain yield of landraces (x axis) in 2021 (g m^{-2}).

In 2021, this trend persisted, with more pronounced differences, possibly due to harsher climate conditions. Small islands' landraces distinctively separated from those of mainland Sicily, and the Apennines group showed increased compactness, indicating increased diversity compared to other groups. Finally, some shifts along the first principal components suggest advantages in plant vitality parameters for the Apennines groups during the 2021 cropping season (Fig. 6).

Participatory evaluation

Participants' preferences resulted correlated to the actual grain yield of lentil: the observation was confirmed through the Spearman correlation test, that estimated a rho value of 0.73 with a $p < 0.0001$. The distribution fitted a logarithmic model ($p < 0.0001$), with participants' preferences expressed as a function of yield (Eq. (1)) (Fig. 7).

$$preference \sim log\,(yield) + e \quad (1)$$

The preferred landraces come from Basilicata, Marche, Apulia, Sardinia, Molise and Tuscany. The control lentil (i.e. cv. Robin) also received good appreciation: for each of the preferred landraces, one control was voted (therefore the 18 preferred landraces consist in nine controls and nine landraces) (Supplementary Information, Table T1).

The participants had the opportunity to express the criteria of their evaluation, revealing that the most preferred traits were soil coverage and canopy homogeneity (indicated by 12 participants out of 14 participants),

followed by plant height (10 participants), pod number and size and percentage of fruit set (7 participants), ability to compete with weeds and erect habitus (6 participants). Additional criteria, mentioned by a minority of the group, were reduced lodging and pod shattering, seed ripening precocity and homogeneity, seed shape and size (Supplementary Information, Table T2).

GWAS

In each year separately we run an association analysis looking for genomic loci related to the expression of measured traits. For the year 2020 we identified 17 marker-trait associations (MTAs) surpassing the stringent threshold across all chromosomes for five traits: flower precocity, pods precocity, plant senescence, plant height and leaf yellowing (Table 1). The highest number of MTAs (four) was found for flower precocity (on chromosomes 1, 2, 5 and 6) (Fig. 8), for pod precocity (on chromosomes 1, 2, 5 and 6) and for plant height (on chromosomes 1, 2 and 5). Three MTAs were identified for leaf yellowing on chromosomes 2, 6 and 7 and two MTAs for plant senescence on chromosomes 4 and 7 (Supplementary Information Manhattanplots 2020 and Supplementary Information QQplots 2020).

For the year 2021, we identified 15 MTAs surpassing the stringent threshold: flower precocity was associated with the most markers (six), again on chromosomes 1, 2, 5 and 6 (Fig. 8), in similar position than the previous year (Table 1). We also identified five MTAs for plant senescence on chromosomes 1, 2, 3 and 7, three MTAs for pod precocity on chromosome 6 and one MTA for yield on chromosome 5 (Supplementary Information Manhattanplots 2021 and Supplementary Information QQplots 2021). For each significant marker, ± 150 bp flanking sequences were extracted to provide context for further research and breeding applications. The flanking sequences are detailed in Supplementary Information T4.

Considering flower precocity, most of the MTAs correspond between the two years, being located maximum 100 million bases apart on each chromosome (except from one association on chromosome 2 in 2020 that is not confirmed in 2021). Only one association for pod precocity correspond in the two seasons (on chromosome.

6, around 200k bases difference), while plant senescence associations do not correspond in 2020 and 2021, being only one association repeated on chromosome 7, but in distant genetic regions. The recurring MTAs harbour putative genes, presented in Supplementary_Information_T5.

Discussions

The main phenotypic variation between the two cropping seasons lies in the weather pattern. In 2021, with reduced precipitation and higher temperatures compared to 2020, landraces exhibited a superior adaptation capacity, suggesting their genetic readiness for environmental stresses. In contrast, the control's yield average (cv. Robin) dropped from 120 g m-2 in 2020 to 107 g m-2 in 2021. This is notably interesting, consisting most of landraces in farmer varieties or heterogeneous material, from non-professional breeding processes[6,8]. This suggests that classic breeding programmes (i.e. cv Robin), operating under a high-input selection environment, may led to genotypes that are less adaptable to environmental variations, potentially posing a disadvantage amid forecasted climatic instability.

This finding is in line with the fact that conventionally bred cultivars are frequently unsuitable for low input conditions, characterized by variable soil type, nutrient and water availability[14,15,16,17]. Since poor and unstable yields pose the main barrier against farmers' adoption of grain legume cultivation[4,18–20], the landraces highlighted merit consideration for breeding programmes tailored to Mediterranean sandy soils and low input systems. The lack of commercial cultivars adapted to specific growing conditions is indeed a major factor contributing to poor and unstable yields, being lentil often cropped in resource-poor and marginal lands[19,21].

We clustered the phenotypic PCA on the base of the landraces' origin of collection, but clusters rarely showed consistency of distribution: in 2020 only landraces from Greece and Sardinia aggregated in clear groups.

Concerning the interpretation of phenotypic variation in view of collection origin, the observed clustering of Ethiopian landraces aligns with prior research[22], highlighting how landraces from Ethiopia showed the most coherent trait-origin association.

The observation was confirmed by genetic variation. Lentils from Ethiopia indeed formed a coherent group, unlike the other foreign landraces (Greek and Tunisian): historical and geographical barriers may offer potential explanations. The Tunisian and Greek coasts are only 150 km from Italy, fostering historical commercial exchanges and migration in the Mediterranean. In contrast, Ethiopia, a hotspot of agrobiodiversity, has had limited interactions with Mediterranean germplasm[23,24,25]. Although two Ethiopian landraces exhibit phylogenetic proximity to Italian ones, the exact reasons remain speculative: possible connections may be rooted in the historical colonization of Ethiopia, with Italians settling in the Horn of Africa around the early 20th century.

Regarding genetic variation of Italian landraces, DAPC findings mostly traced Italian geography yet suggesting that landrace collection origin may not always correspond to their area of evolution.

Concerning Sicilian islands landraces, their limited genetic diversity can be attributed to the discontinuous population flow and recent crop reintroduction[26], creating a strong founder effect and leading to genetic uniformity[8,27]. Nevertheless, each lentil landrace is locally distinguished, often marketed as distinct varieties. The genotyping results could spur a reconsideration, or enlargement, of local denominations, as is the case of the Sicilian archipelago.

Despite a geographical separation of almost 300 km, the landrace from Ventotene island shares a genetic affinity with those of the Sicilian islands. Racheli[28] suggested a migration of these populations from the nearby mainland (Naples area) in the 18th century. Contextually, Laghetti et al.[7]referred to older reports indicating that legume cultivation was introduced to the small Sicilian islands from the Naples area, likely starting from Pantelleria and Ustica[6].

Trait	Year	SNP	Chrom	Position	*P*.value	maf	FDR Adj *P*	effect	Year	SNP	Chrom	Position	*P*.value	maf	FDR Adj *P*	effect
SENESCENCE	2020	438897:377:-	4	3,89E + 08	4.29e-20	0,2381	4.09e-15	−0,8339	2021	1010:332:+	1	4,267,412	3.33e-14	0,0825	2.57e-09	−0,4963
SENESCENCE	2020	662697:258:+	7	14,642,262	4.64e-07	0,1190	0,022	−0,7648	2021	738522:308:-	7	3,55E + 08	5.38e-14	0,1602	2.57e-09	0,5479
SENESCENCE									2021	296340:109:-	3	1,84E + 08	1.59e-13	0,2330	5.08e-09	0,5030
SENESCENCE									2021	307601:42:+	3	2,34E + 08	4.05e-10	0,0777	9.65e-06	0,6364
SENESCENCE									2021	196766:369:-	2	3,36E + 08	6.19e-07	0,4903	0,012	−0,1772
FLOWERS	2020	503277:291:+	5	1,93E + 08	1.55e-15	0,0765	1.47e-10	−2,3007	2021	202147:98:-	2	3,61E + 08	1.94e-09	0,4466	0,0002	−0,4478
FLOWERS	2020	657977:102:-	6	4,15E + 08	2.43e-10	0,4353	1.16e-05	−1,0008	2021	527037:277:-	5	2,94E + 08	3.11e-08	0,2282	0,001	−0,4587
FLOWERS	2020	92122:323:-	1	4,05E + 08	2.76e-07	0,2588	0,009	−0,9246	2021	121302:84:+	1	5,36E + 08	4.28e-08	0,4272	0,001	0,3813
FLOWERS	2020	194586:321:-	2	3,27E + 08	4.04e-07	0,2059	0,010	0,8172	2021	644244:68:-	6	3,51E + 08	1.09e-07	0,2670	0,003	0,3690
FLOWERS									2021	110596:57:+	1	4,88E + 08	3.82e-07	0,2718	0,006	−0,2499
FLOWERS									2021	131758:40:-	2	45,333,964	8.41e-07	0,3544	0,011	0,2324
PODS	2020	499720:122:+	5	1,78E + 08	1.53e-12	0,3471	1.46e-07	−0,8852	2021	580268:132:+	6	66,200,043	7.59e-09	0,1165	0,001	0,6682
PODS	2020	28012:340:-	1	1,24E + 08	2.51e-11	0,4882	1.19e-06	−0,4976	2021	567049:10:-	6	4,268,673	1.01e-07	0,2087	0,005	0,5326
PODS	2020	580203:103:+	6	65,982,871	6.53e-08	0,1176	0,002	−0,7202	2021	655257:373:-	6	4,03E + 08	5.10e-07	0,2184	0,016	−0,3880
PODS	2020	205035:334:+	2	3,74E + 08	8.82e-07	0,4706	0,021	0,2951								
YIELD									2021	512821:373:+	5	2,33E + 08	1.84e-14	0,0882	1.76e-09	87,8180
HEIGHT	2020	77693:73:+	1	3,4E + 08	1.89e-12	0,1118	1.81e-07	5,5363								
HEIGHT	2020	562756:56:-	5	4,6E + 08	9.33e-09	0,4765	0,0004	−2,1903								
HEIGHT	2020	158077:80:+	2	1,66E + 08	1.99e-07	0,0882	0,006	−3,0742								
HEIGHT	2020	210295:318:+	2	3,97E + 08	3.31e-07	0,4647	0,008	1,8594								
LEAF_YELLOW	2020	632549:360:+	6	2,98E + 08	1.62e-10	0,2471	1.54e-05	−0,9271								
LEAF_YELLOW	2020	211786:405:-	2	4,04E + 08	3.50e-08	0,4882	0,002	0,5743								
LEAF_YELLOW	2020	769234:77:+	7	4,91E + 08	2.48e-07	0,1294	0,008	0,7244								

Table 1. MTAs (marker trait associations) in GWAS (Genome-Wide Association Studies) based on phenotypic data of the 111 lentils' accessions characterized over 2020 and 2021 seasons.

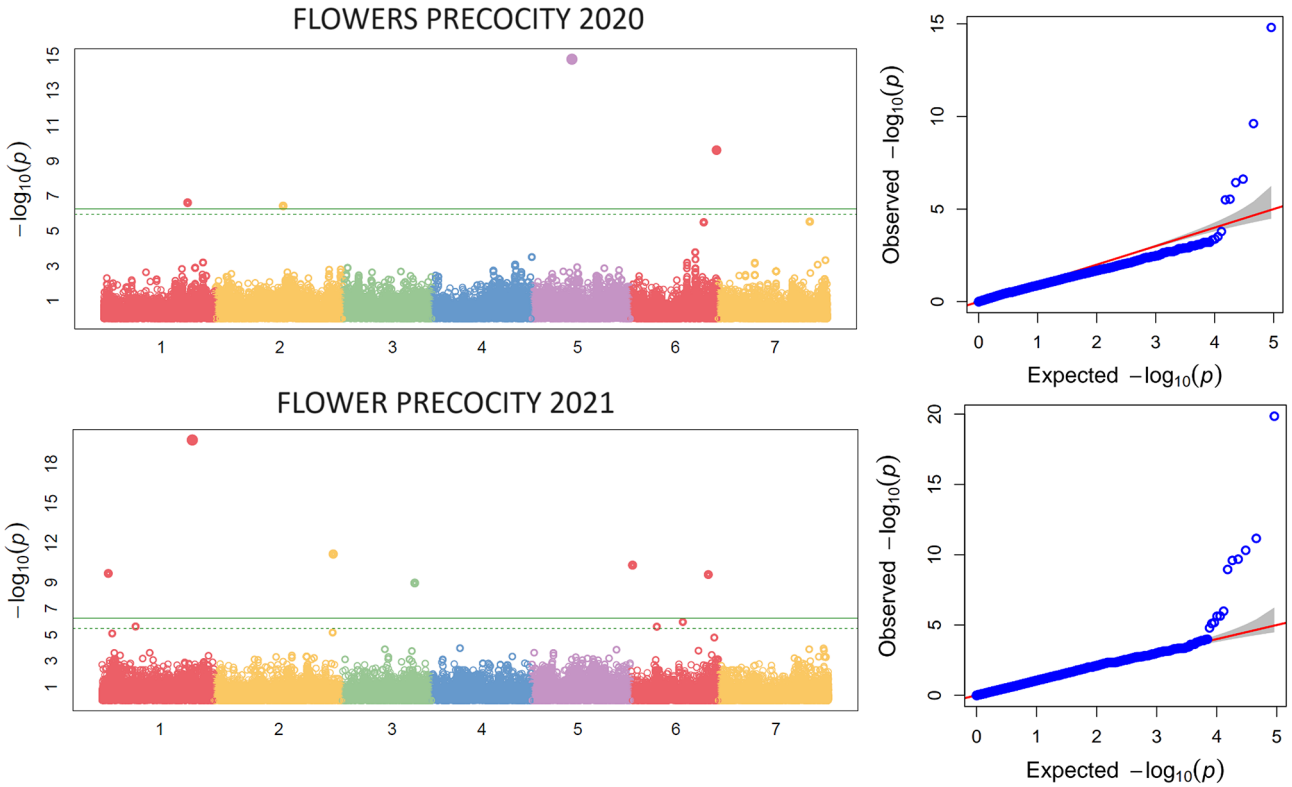


Fig. 8. GWAS results for flower precocity in 2020 (top) and 2021 (bottom). On the left, Manhattan plots depicting individual SNPs across all chromosomes (x-axis) and their corresponding -log10 P values (y-axis). Horizontal lines indicate a stringent Bonferroni threshold for a nominal p-value of 0.05. SNPs exceeding this threshold signify genomic loci strongly associated with variations in the collection's phenotypic traits. On the right, The QQ plot visualizes the comparison between observed and expected p-values. Each point on the plot represents an SNP in the study. On the x-axis, expected -log10 transformed p-values are plotted under the assumption of no association, while the y-axis displays the observed -log10 transformed p-values obtained from the GWAS analysis. Deviations from the diagonal line (y = x) indicate departures from the null hypothesis of no association.

In the marginal Apennine ridge areas, lentil cultivation has been prevalent since the early 20th century[6]. The region's pedological barriers and lack of developed infrastructure have led to highly fragmented cultivation, with minimal farm-to-farm exchange[29]. This likely explains the distinct monophyletic grouping within the main cluster. Apulia and Sicily regions also have a lentil cultivation tradition, but landraces from these regions don't exhibit clear monophyletic grouping and are widely distributed in the tree. Commercial reasons likely contributed, as lentils from these regions, historically well-known[6], probably spread along commercial routes giving rise to local populations in different regions.

Finally, the clustering of the phenotypic PCA based on DAPC groups provides a more uniform and coherent categorization of landraces, showing that phenotypic traits align with phylogenetic distance rather than collection origin, as previously observed[30,31]. This can be attributed to the possibility that the harvest location may not correspond to the area of landrace evolution and adaptation. Additionally, Italian regions are delineated by political borders, often not aligning with pedoclimatic differences.

Marker-trait associations (MTAs) of lentil agronomic performances are rarely available in literature, therefore we do not exclude that our observations could be among the first references for further studies. Indeed, the few existing studies often pinpoint different associations than those we observed, highlighting the importance not to rely on existing data before more differentiated analysis are carried out. For instance, a recent study by Rajendran et al[32]. discovered MTAs with flowering time on chromosome 3, potentially linked to the Hr locus regulating the flowering response to photoperiod in legumes[33]. Neupane and collaborators[34]instead suggested associations with chromosome 5, 2, and 6, similarly to our study: indeed, we noticed on both years significant markers located in the same general regions than Neupane et al. (around 0,4 Mbp on ch. 2, around 4 Mbp on ch. 6). Flowering time is a crucial trait to control, especially in the face of increasing climate instability. New trends in frost and heat wave timing indeed are damaging flowers, reducing production[36,37]. Additionally, given the recent interest in intercropping, flowering time becomes a significant trait for breeders, playing a key role in regulating species complementarity[38]. Our study found that MTAs for flower precocity also aligned with other precocity indicators: pod precocity in 2020 (same flower positions on chromosomes 2 and 5) and senescence in 2021 (same positions of flowers and pods on chromosome 2). However, the consistency of pod precocity and senescence across the two years was not confirmed.

We observed another interesting association on chromosomes 2, 6 and 7, linking foliar pathogen sensitivity: significant markers for pathogen infection indeed emerged in year 1. This trait is broadly related to stress conditions rather than just pathogen sensitivity: therefore, the identified MTA may be related to lentil stress response through leaf depletion. Among the general pathogen sensitivity, special attention was given to Bruchus spp. resistance, because few information is available and the chemical control of this pest presents several environment concerns[39]. Although we did not identify significant markers for Bruchus sensitivity in our analyses, the importance of this trait warrants continued investigation.

Plant height also exhibited strong associations with specific markers in our study. Plant height is a trait of interest for breeding, but it is often mistakenly considered the sole cause of lodging. The correlation of lodging and plant height indeed is not straightforward, as lodging is also influenced by traits related to culm and roots[40,41]. Proving this, short plants have been proven susceptible to lodging in our study and previous ones[41]. Suitability to mechanization, a major challenge in lentil cultivation, may benefit from the success of molecular breeding in selecting for plant height, as it was the case for other crops[43,44,45]. Further research is necessary tough to identify lodging-specific markers and to effectively select for plant height in lentils.

Finally, we identified one MTA with grain yield in 2021 season. Yield is a complex trait: its interaction with the environment plays a crucial role and we expect the occurrence of different regulating genes according to the specific environments. In addition, yield is typically controlled by manifold small-effect loci and thus is challenging to be dissected by experiments with relatively low power as this one. In face of these challenges, even a single association can be valuable, given the limited information on the genetic determinants of complex traits in lentils. Notably, only one GWAS lentil study to date considered yield as a study trait[46].

Given the significant marker-trait associations (MTAs) identified in our preliminary study, we recommend that future studies aiming to validate these findings consider increasing plot size, expanding the variety of environments, and incorporating a broader range of genotypes to further improve the accuracy of agronomic performance evaluations.

Moving to the appreciation of the studied landraces, despite lacking specific expertise, participants' preferences correlated with actual lentil grain yield. This experience demonstrates the feasibility of indirectly estimating lentil grain yield based on plant morphological traits, thus highlighting the potential for applying participatory selection in lentil breeding. Participants also mentioned specific criteria for peculiar situations, such as the importance of a prostrate growth habitus for wheat-lentil intercropping. This suggests that unconventional breeding programs need to consider parameters beyond grain yield, aligning with prior research[14,47. 48,49].

Participatory breeding, with its decentralized selection approach, has proven efficient, particularly for early generation material, in providing targeted seed resources for diverse environments[13,15,50]. This is particularly crucial for lentil breeding, that faces a shortage of genetic material and locally adapted varieties: e.g., one of the most common lentil cultivated in Tuscany, cv. Robin, was selected in Canada and multiplied internationally before reaching the Italian market. This study marks an initial step in participatory lentil selection tailored to Mediterranean sandy soils, offering genetic material and insights applicable beyond. Additionally, the explored lentil collection could serve as a pool for selecting heterogeneous material, providing an alternative to conventional breeding or a parallel pathway. Heterogeneous material demonstrates adaptability to local conditions, addressing the shortage of adapted cultivars and offering a viable solution for climate change challenges[51,52].

While acknowledging the boundaries posed by the limited number of accessions and seasons, our study maintains its focus on providing valuable insights into the genetic landscape of Mediterranean lentils. Our year-specific GWAS has identified robust signals, often validated across both years, emphasizing the significance of our findings for specific environmental conditions. Due to the limited genotyping coverage and incomplete annotation of the lentil genome, we chose not to pursue the identification of -potentially biased- candidate genes. Instead, we focused on providing breeders with actionable results to enhance screening and improve lentil traits. By highlighting the pioneering nature of our work, we aim to inspire future studies to build on our methodology and results to explore lentil agrobiodiversity across additional regions, fostering a deeper understanding of the genetic factors influencing their agronomic potential.

Materials and methods

Site, climate and soil

After amplification of the genetic material in 2019, a field experiment was conducted in 2020 and 2021 at San Piero a Grado (43.66 lat. N, 10.35 long. E), ca. 9 km SW of Pisa, Central Italy. The topsoil (0–0.3 m) at the experimental site was predominantly sandy, with a composition of 795 g kg^{-1} of sand, 129 g kg^{-1} of silt and 76 g kg^{-1} of clay. The soil had a pH (water) of 7.5, 2% Organic Matter, 1.4 g kg^{-1} total Nitrogen, C/N ratio of 8.9, 29.1 mg kg^{-1} of P_2O_5 (Olsen method) and 91.5 mg kg^{-1} K (BaCl2).

The experimental site features a Mediterranean climate, characterized by mild and relatively rainy winters, and hot dry summers. Total mean annual precipitations are 910 mm, ranging from 25 mm in the driest month (July) to 145 mm in the wettest (November), with an irregular pattern throughout the year, autumn being the rainiest season.

In 2020 lentil was sown on the 5th of March and was harvested on the 9th of July. During that time, we registered a total precipitation of 276 mm; the average maximum temperature was 22.6 °C (peaking at 32 °C and 9.5 °C, with 11 days > 30 °C and 1 day < 10 °C), while the average minimum temperature was 9 °C (peaking at 18 °C and − 6.5 °C, with no days > 20 °C and 8 days < 0 °C) (Fig. 9).

In 2021 lentil was sown on the 25th of February and harvested on the 6th of July. During the growing season we registered a total precipitation of 142 mm; the average maximum temperature was 21.8 °C (peaking at 33.8 °C and 13.2 °C, with 18 days > 30 °C and no days < 10 °C), while the average minimum temperature was 8 °C (peaking at 18.5 °C and − 4.2 °C, with no days > 20 °C and 11 days < 0 °C) (Fig. 9).

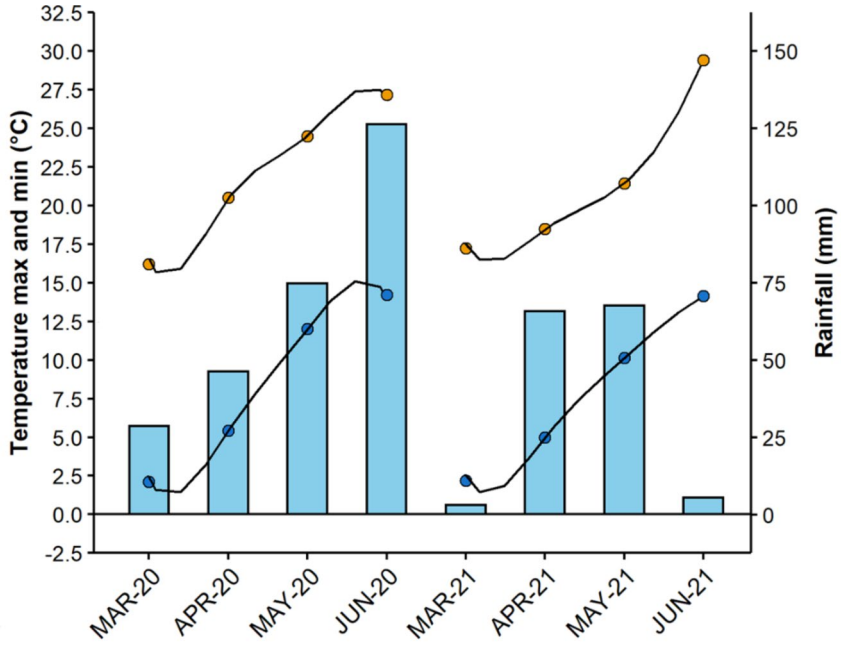


Fig. 9. Weather pattern registered at San Piero a Grado during the two lentil growing seasons (March to June 2020 and 2021). The left y axis shows the maximum (top lines) and minimum (bottom lines) monthly average temperatures, while the right y axis shows the monthly total rainfall (blue bars).

The experimental field was organically managed and consisted in a fallow that hosted lettuce as a cultural precedent the year before the trial. No fertilizers were provided to the crop, weeding was managed manually.

Plant materials and experimental design

All methods were conducted in strict adherence to applicable national and international guidelines and regulations, with the approval of the Scuola Superiore Sant'Anna. The plant material used in this study was exclusively cultivated within Scuola Superiore Sant'Anna facilities, and the utilization of seeds followed previously obtained permissions in accordance with relevant regulations.

In 2019, we assembled a collection of 106 lentil cultivars from various sources and we multiplied it in the same environment of the experimental fields. The aim of the collection was to capture the historical lentil diversity across the Italian peninsula, therefore the main choice criterion for the inclusion in the collection was the geographical origin: all Italian lentil landraces available at each provider were collected, with the addition of 15 landraces from Greece, Tunisia and Ethiopia. The landraces were provided from the CNR Bari (Italy) and the IPK Gatersleben (Germany) gene banks, that share a common collection (39 landraces in total, with 14 in common); from the University of Perugia - ITA 363 (17 landraces); from ARSIAL -the regional agency for development and innovation of the Latium region- in collaboration with Rete Semi Rurali (14 landraces); from Catania University (13 landraces), Naples University and the CNR Pontecagnano (4 landraces). Seed exchange events and some farmers also provided part of the material (4 landraces). When available, we collected information about the collection location for each of the landraces (Supplementary Information T1).

The experimental design for the phenotypic characterization consisted in an augmented block design with spatial analysis[53]. The commercial cultivar Robin, a Canadian cultivar[54], was used as a control plots, repeated 32 times; plots were 1×1 m, sown at a density of 300 seeds m^{-2} (Supplementary Information, Figure F3).

Phenotypic diversity

The target phenotypic traits were defined in 2019 during a focus group where 45 lentils farmers, members of the Lenticchia di Altamura PGI, shared their perspectives on trait preferences and limitations for lentil production. The findings were further discussed with lentil breeders from Caussade Semence (France) and Saatgut-Forschung (Germany) who provided additional information on traits' interest from a breeding perspective.

In 2020, eight phenotypic traits were recorded on whole plot at different growth stages using a Likert scale[55]: overall vigour (defined as general health and growth rate), leaf yellowing, flower precocity, canopy closure, pods precocity, lodging, plant senescence, sensitivity to pathogen infections. Sensitivity to pathogens was assessed based on the occurrence of stem, foliar and pod symptoms of the most common lentil diseases: Pythium spp., Fusarium, Anthracnose, Stemphylium and Ascochyta blight, leaf roll virus. In 2021 this characterization was repeated by adding an additional trait: sensitivity to *Bruchus* spp. (Coleoptera, Chrysomelidae), upon the occurrence of a seasonal pest attack. *Bruchus* attack was characterized in post-harvest stage according to the percentage of seed damage. The scale ranged from 0 (indicating absence) to 5 (indicating maximum severity or 100% of presence): e.g. a mean value of 3 for lodging means that over the two years 60% of the plot was affected

by lodging. In both years, quantitative measures were also recorded for three agronomic traits: plant height (cm), total biomass (g m^{-2}) and grain yield (g m^{-2}). Plant height was recorded from three randomly selected plants, while biomass and yield from the whole plot.

A spatial analysis was conducted with the kriging method: x-y coordinates were assigned to each plot to create a spatial grid. Within the grid, kriging values were estimated on the base of check plots performance, used to quantify the spatial variation with a semivariogram[56]. Deviation from the kriging was then computed to estimate the performance of each landrace under study. The kriging analysis was performed in R for Windows, version 4.0.3 (R Core Team R, The R Project for Statistical Computing. https://www.r-project.org/), with the gstat package[57].

A multivariate analysis of the kriged traits was then performed: a principal component analysis (PCA) was computed in order to evaluate the unbiased relationship between the phenological traits and the lentils landraces. PCA was conducted with R package FactorMiner[58], while data visualization with the R package ggplot2, version 3.3.5[59]. The analysis was conducted separately in the two experimental years: the choice is based on the different environmental conditions (climate) that interacted with the phenotypic observations.

Genetic diversity

Ten lentil seeds per landrace were germinated in towels. The seedlings were combined into a single pooled sample for each landrace, frozen in liquid nitrogen and stored at −80°C until grounding. Genomic DNA was extracted at the Sant'Anna School of Advanced Studies (Pisa, Italy) laboratories using the GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. DNA quality was checked using a NanoDrop ND-1000 spectrophotometer. Samples were normalized and sent to IGATech (Udine, Italy), where they were pooled into multiplexing batches. DNA was purified using 1.5 volumes of AMPure XP beads (Agencourt). The library was then prepared for sequencing, with fragments selected for a size range of 400 bp to 550 bp. Sequencing was performed using double-digestion restriction site-associated DNA markers (ddRAD) with the enzymes SphI and MboI on an Illumina NovaSeq 6000 platform. The sequencing was conducted in paired-end mode over 150 cycles. Raw reads were aligned on the reference lentil genome of cv. CDC Redberry[61]using the BWA aligner (Burrows-Weeler-Aligner v.0.7.12)[62] and single nucleotide polymorphisms (SNPs) were called using the STACKS software (Stacks v2.0)[63] in population mode. SNPs failing in more than 20% samples and with a heterozygosity higher than 50% were removed from the dataset.

SNPs were loaded in TASSEL 5.2.29[64] as a .hmp file and further filtered for quality, with the following parameters: markers with MAF > 0.05, 95% sites present data and maximum observed heterozygosity of 80%.

Genetic diversity analyses were conducted on the refined set of high-quality SNPs (91,135) using R for Windows. To unveil potential genetic groupings within the core collection, a discriminant analysis of principal components (DAPC) was carried out using R/ adegenet[65]. DAPC describes genetic clusters and reveals the best possible groupings within the core collection, providing insights into the underlying population structure. The optimal number of clusters for the DAPC was identified using the find.cluster function that, after transforming data through a PCA, computes successive K-means with an increasing number of clusters. For each model, a statistical measure of goodness of fit (BIC) is computed, which allows to choose the optimal number of clusters (Supplementary Information, Figure F4). The clusters were used to perform a spider plot of the phenotypic PCA with the R package ggplot2, in order to evaluate correspondences between phenotypic and genetic traits.

To illustrate the genetic relationships among the genotypes, a neighbor-joining tree was constructed using R/ adegenet[65], offering a visual representation of genetic distances among the genotypes, with the country of origin indicated for each genotype. Complementing this, a principal component analysis (PCA) was conducted on the SNP dataset using the same R package. The PCA depicted the DAPC clusters and included information about the country of origin for each genotype allowing for a visual comparison of genetic clusters with geographical origins, thereby enhancing the understanding of the genetic diversity and structure within the core collection. To assess genetic variability within each genetic cluster (DAPC clusters), observed heterozygosity was calculated using R/adegenet.

Participatory evaluation

A participatory evaluation of lentil traits was conducted in 2021 close to physiological maturity. Fourteen participants, including farmers (not strictly lentil farmers), technicians and researchers from the region, were asked to evaluate each landrace of the trial according to their preferences from 1 to 5 (Likert scale), and then to express their evaluation criteria. This approach ensured that all participants evaluated every landrace, thereby avoiding any potential imbalance due to the unreplicated design. Participants had no prior information on the landraces. In order to take into account the field effect in defining the most preferred landraces, participants' preferences were transformed into kriging values on the base of check plots performance, then deviation from the kriging was computed.

Participants' preferences were analysed in relation to landrace yield: a correlation analysis was computed through the Spearman correlation test, with the cor.test function of R, and tested through a general linear model with yield as a function of participants' votes.

Participants evaluation criteria were analysed with a qualitative approach: each criterion was categorized and each category was ranked on a frequency basis.

Informed consent to image and information publication was obtained from all subjects participating to the evaluation.

Genome-wide Association studies

The genome wide association study (GWAS) was conducted utilising the FarmCPU model[66], incorporating both fixed and random circulating probability unification methods. This was executed through the in R/GAPIT package[67], with the first principal component serving as a covariate. Marker-trait associations (MTAs) were identified based on SNPs exceeding a false discovery rate threshold of 0.05, as determined by Storey's method[68].

FarmCPU is recognized for its superior ability to control false positives while maintaining high sensitivity for true positives compared to traditional models like GLM and MLM[66]. Unlike MLMs such as MLMM, which may compromise true positive detection while adjusting for population structure and kinship, FarmCPU employs an iterative approach with Fixed Effect Model (FEM) and Random Effect Model (REM). This iterative method separates testing markers from associated markers, using them to define kinship and unify P values across iterations. In addition, FarmCPU not only enhances statistical power compared to existing approaches but also offers efficient computing times, enabling the analysis of large datasets. Comparative analyses consistently demonstrate FarmCPU's superiority over MLM and MLMM in crop-based GWAS[69].

The phenotypes considered in the GWAS were leaf yellowing (at two stages), plant vigour, flowers precocity, maturation precocity, soil cover, lodging (at two stages), plant senescence, pathogen sensitivity, grain yield, biomass production, plant height. Phenotypes were treated separately for the two years in which the experiment was run to account for environmental variation. Phenotypic data was corrected using a generalized linear model and a population stratification correction based on multi-dimensional scaling (5 dimensions)[70]. MTAs were defined by a stringent statistical threshold, devised on a false discovery rate of 5% and based on multiple testing correction of associations using the R package qvalue[71]. For this threshold, we minimize type I errors (false positives) over type II errors (false negatives). To identify gene functions potentially linked with significant MTAs, annotations were searched in the *Lens culinaris*CDC Redberry Genome Assemblyv1.2[61]. Gene models, associated protein-coding genes and gene functions were searched within ±200 Kb of the MTAs loci for each significant MTA confirmed across both years.

Data availability

The genomic dataset generated and analysed during the current study is available in the SRA repository BioProject ID: PRJNA1125613. Additionally, all relevant dataset and information were deposited on Figshare at this permanent address https://doi.org/10.6084/m9.figshare.26197109.v1.

Received: 5 February 2024; Accepted: 16 December 2024
Published online: 30 December 2024

References

1. Multari, S., Stewart, D. & Russell, W. R. Potential of fava bean as future protein supply to partially replace meat intake in the human diet. *Compr. Rev. Food Sci. Food Saf.* **14** (5), 511–522. https://doi.org/10.1111/1541-4337.12146 (2015).
2. Ferreira H, Pinto E, Vasconcelos MW (2021) Legumes as a cornerstone of the transition toward more sustainable agri-food systems and diets in Europe. Front Sustain Food Syst 5:694121. https://doi.org/10.3389/fsufs.2021.694121s3Martinelli F, Vollheyde AL, Cebrián-Piqueras MA, von Haaren C, Lorenzetti E, Barberi P, … & Yousefi S (2022) LEGU-MED: Developing biodiversity-based agriculture with legume cropping systems in the Mediterranean basin. Agronomy 12 (1):132. https://doi.org/10.3390/agronomy12010132.
4. Watson CA, Reckling M, Preissel S, Bachinger J, Bergkvist G, Kuhlman T, … Zander P(2017) Grain Legume Production and Use in European Agricultural Systems. In Advances in Agronomy, 144, 235–303. https://doi.org/10.1016/bs.agron.2017.03.003.
5. Malik, D. P., Devi, M. & Reddy, A. A. Global status of lentil production with special reference to India. *Indian J. Agric. Sci.* **92** (4), 474–479 (2022).
6. Piergiovanni, A. R. The evolution of lentil (Lens culinaris Medik.) Cultivation in Italy and its effects on the survival of autochthonous populations. *Genet. Resour. Crop Evol.* **47**, 305–314. https://doi.org/10.1023/A:1008789614680 (2000).
7. Laghetti, G., Perrino, P., Olita, G. & Hammer, K. Multicrop collecting expeditions in aeolian archipelago (Italy). *Plant. Genet. Resour. Newsl.* **128**, 26–34 (2001).
8. Sonnante, G. & Pignone, D. The major Italian landraces of lentil (Lens culinaris Medik.): their molecular diversity and possible origin. *Genet. Resour. Crop Evol.* **54**, 1023–1031. https://doi.org/10.1007/s10722-006-9153-x (2007).
9. Laghetti, G., Piergiovanni, A. R., Sonnante, G., Lioi, L. & Pignone, D. The Italian lentil genetic resources: a worthy basic tool for breeders. *Eur. J. Plant. Sci. Biotechnol.* **2**, 48–59 (2008).
10. Crepaldi, G. Il 98% delle lenticchie e il 95% dei fagioli secchi è importato. Il fatto alimentare. (2018). https://ilfattoalimentare.it/legumi-secchi-importazione.html Accessed June 2023.
11. Foyer, C. et al. Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants*. **2**, 16112. https://doi.org/10.1038/nplants.2016.112 (2016).
12. Ceccarelli, S. & Grando, S. Decentralized-participatory plant breeding: an example of demand-driven research. *Euphytica* **155**, 349–360. https://doi.org/10.1007/s10681-006-9336-8 (2007).
13. Dawson, J. C., Murphy, K. M. & Jones, S. S. Decentralized selection and participatory approaches in plant breeding for low-input systems. *Euphytica* **160** (2), 143–154. https://doi.org/10.1007/s10681-007-9533-0 (2008).
14. Atlin, G. N. & Frey, K. J. Breeding crop varieties for low-input agriculture. *Am. J. Altern. Agr.* **4** (2), 53–58. https://doi.org/10.1017/S0889189300002721 (1989).
15. Ceccarelli, S. Adaptation to low/high input cultivation. *Euphytica* **92**, 203–214. https://doi.org/10.1007/BF00022846 (1996).
16. Bänziger, M. & Cooper, M. Breeding for low input conditions and consequences for participatory plant breeding: examples from tropical maize and wheat. *Euphytica* **122**, 503–519. https://doi.org/10.1023/A:1017510928038 (2001).
17. Reid, T. A. et al. Should spring wheat breeding for organically managed systems be conducted on organically managed land? *Euphytica* **169**, 239–252. https://doi.org/10.1007/s10681-009-9949-9 (2009).
18. Malhotra, R. S. & Erskine, W. Breeding lentil for High Yield, Stability, and adaptation. *Adv. Pulses Res. Bangladesh* **235**. (1989).
19. Magrini MB, Anton M, Cholez C, Corre-Hellou G, Duc G, Jeuffroy MH, … & Walrand S (2016) Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. Ecol Econ 126:152–162. https://doi.org/10.1016/j.ecolecon.2016.03.024.
20. Lorenzetti, E., Carlesi, S. & Bàrberi, P. Mixtures of Commercial Lentil cultivars Show Inconsistent results on agronomic parameters but positive effects on Yield Stability. *Agronomy* **12** (9), 2107. https://doi.org/10.3390/agronomy12092107 (2022).
21. Akibode, C. S. & Maredia, M. K. Global and regional trends in production, trade and consumption of food legume crops. (2012).

22. Erskine, W., Adham, Y. & Holly, L. Geographic distribution of variation in quantitative traits in a world lentil collection. *Euphytica* **43** (1), 97–103. https://doi.org/10.1007/BF00037901 (1989).
23. Caproni L, Lakew BF, Kassaw SA, Miculan M, Ahmed JS, Grazioli S, … Dell'Acqua M (2023)The genomic and bioclimatic characterization of Ethiopian barley (Hordeum vulgare L.) unveils challenges and opportunities to adapt to a changing climate. Global Change Biology. https://doi.org/10.1111/gcb.16560.
24. Fantahun, B. et al. Multivariate analysis in the dissection of phenotypic variation of Ethiopian cultivated barley (Hordeum vulgare ssp Vulgare L.) genotypes. *Cogent Food Agric.* **9** (1), 2157104. https://doi.org/10.1080/23311932.2022.2157104 (2022).
25. Vavilov, N. I., Vavilov, M. I. & Dorofeev, V. F. *Origin and Geography of Cultivated Plants* (Cambridge University Press, 1992).
26. Sansivente, B. M. *Report on Linosa Island for Ferdinand II of Bourbons* (Naples, 1849).
27. Laghetti, G., Volpe, N., Sonnante Ga, Pignone, D. & Sonnante Gi. & Salvaguardia, caratterizzazione e valorizzazione dell'antico agro-ecotipo pugliese ''Lenticchia di Altamura''. In VII National Congress Biodiver. (p. 148). Catania, Italy. (2005).
28. Racheli, G. Le isole Ponziane rose dei venti. Natura, storia, arte. Mursia (Ed.), Milan, Italy. (1987).
29. Zucchini, M. *Il cece, la lenticchia e la cicerchia. Biblioteca per l'insegnamento agrario professionale* (Ramo Editoriale degli Agricoltori, 1937).
30. Toklu, F., Biçer, B. & Karaköy, T. Agro-morphological characterization of the Turkish lentil landraces. *Afr. J. Biotechnol.* **8**, 17 (2009).
31. Lombardi, M. et al. Assessment of genetic variation within a global collection of lentil (Lens culinaris Medik.) Cultivars and landraces using SNP markers. *BMC Genet.* **15**, 150. https://doi.org/10.1186/s12863-014-0150-3 (2014).
32. Rajendran K, Coyne C, Zheng P, Saha G, Main D, Amin N, … & Kumar S (2021) Genetic diversity and GWAS of agronomic traits using an ICARDA lentil(Lens culinaris Medik.) Reference Plus collection. Plant Genet Resour Charact Util 19 (4):279–288. https://doi.org/10.1017/S147926212100006X
33. Weller JL, Liew LC, Hecht VF, Rajandran V, Laurie RE, Ridge S, … Lejeune-Hénaut I(2012) A conserved molecular basis for photoperiod adaptation in two temperate legumes.Proc Natl Acad Sci 109 (51):21158–21163. https://doi.org/10.1073/pnas.1207943110.
34. Neupane, S. et al. Focusing the GWAS Lens on days to flower using latent variable phenotypes derived from global multi-environment trials. bioRxiv. (2023). https://doi.org/10.1002/tpg2.20269
36. Zheng, B., Chenu, K. & Chapman, S. C. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivium) varieties? *Glob. Change Biol.* **18** (9), 2899–2914. https://doi.org/10.1111/j.1365-2486.2012.02724.x (2012).
37. Zheng, B., Chenu, K. & Chapman, S. C. Velocity of temperature and flowering time in wheat – assisting breeders to keep pace with climate change. *Glob. Change Biol.* **22** (2), 921–933. https://doi.org/10.1111/gcb.13118 (2016).
38. Davis, J. H. C. & Woolley, J. N. Genotypic requirement for intercropping. *Field Crops Res.* **34** (3–4), 407–430. https://doi.org/10.1016/0378-4290(93)90124-6 (1993).
39. Laserna-Ruiz, I., De-Los-Mozos-Pascual, M., Santana-Méridas, O., Sánchez-Vioque, R. & Rodríguez-Conde, M. F. Screening and selection of lentil (Lens Miller) germplasm resistant to seed bruchids (Bruchus spp). *Euphytica* **188** (2), 153–162. https://doi.org/10.1007/s10681-012-0752-7 (2012).
40. Berry, P., Spink, J., Gay, A. & Craigon, J. A comparison of root and stem lodging risks among winter wheat cultivars. *J. Agric. Sci.* **141** (2), 191–202. https://doi.org/10.1017/S002185960300354X (2003).
41. Ookawa T, Hobo T, Yano M, Murata K, Ando T, Miura H, … & Matsuoka M (2010) New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. Nat Commun 1 (1):1–11. https://doi.org/10.1038/ncomms1132.
42. Navabi, A., Iqbal, M., Strenzke, K. & Spaner, D. The relationship between lodging and plant height in a diverse wheat population. *Can. J. Plant. Sci.* **86** (3), 723–726. https://doi.org/10.4141/P05-144 (2006).
43. Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S, … Penner G (2003)Quantitative trait loci for lodging resistance, plant height and partial resistance to mycosphaerella blight in field pea (Pisum sativum L.). Theor Appl Genet 107 (8):1482–1491.https://doi.org/10.1007/s00122-003-1379-9.
44. Rao, Y., Li, Y. & Qian, Q. Recent progress on molecular breeding of rice in China. *Plant. Cell. Rep.* **33** (4), 551–564. https://doi.org/10.1007/s00299-013-1551-x (2014).
45. Tian X, Wen W, Xie L, Fu L, Xu D, Fu C, … Cao S (2017) Molecular mapping of reduced plant height gene Rht24 in bread wheat. Front Plant Sci 8:1379. https://doi.org/10.3389/fpls.2017.01379.
46. Kumar, J. et al. Identification of QTLs for agronomic traits using association mapping in lentil. *Euphytica* **214**, 75. https://doi.org/10.1007/s10681-018-2155-x (2018).
47. Crespo-Herrera, L. A. & Ortiz, R. Plant breeding for organic agriculture: something new? *Agric. Food Secur.* **4** (1), 1–7. https://doi.org/10.1186/s40066-015-0045-1 (2015).
48. Bourke PM, Evers JB, Bijma P, van Apeldoorn DF, Smulders MJ, Kuyper TW, … & Bonnema G (2021) Breeding beyond monoculture: putting the intercrop into crops.Front Plant Sci 12:734167. https://doi.org/10.3389/fpls.2021.734167.
49. Moore VM, Schlautman B, Fei SZ, Roberts LM, Wolfe M, Ryan MR, … & Lorenz AJ (2022) Plant Breeding for Intercropping in Temperate Field Crop Systems:A Review. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.843065.
50. Gesesse CA, Nigir B, de Sousa K, Gianfranceschi L, Gallo GR, Poland J, … & Dell'Acqua M (2023) Genomics-driven breeding for local adaptation of durum wheat is enhanced by farmers' traditional knowledge. Proc Natl Acad Sci 120 (14):e2205774119.https://doi.org/10.1073/pnas.2205774119.
51. Raggi, L., Negri, V. & Ceccarelli, S. Morphological diversity in a barley composite cross-derived population evolved under low-input conditions and its relationship with molecular diversity: indications for breeding. *J. Agric. Sci.* **154** (6), 943–959. https://doi.org/10.1017/S0021859615000921 (2016).
52. Ceccarelli, S. & Grando, S. Evolutionary plant breeding as a response to the complexity of climate change. *iScience* **23** (12), 101815. https://doi.org/10.1016/j.isci.2020.101815 (2020).
53. Müller, B. U., Schützenmeister, A. & Piepho, H. P. Arrangement of check plots in augmented block designs when spatial analysis is used. *Plant. Breed.* **129** (6), 581–589. https://doi.org/10.1111/j.1439-0523.2010.01803.x (2010).
54. Vandenberg A, Kiehn FA, Vera C, Gaudiel R, Buchwaldt L, Dueck S, … Slinkard E A (2002)CDC Robin lentil. Can J Plant Sci 82 (1):111–112. https://doi.org/10.4141/P01-003.
55. Allen, I. E. & Seaman, C. A. Likert scales and data analyses. *Qual. Prog.* **40** (7), 64–65 (2007).
56. Zas, R. Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genet. Genomes.* **2**, 177–185. https://doi.org/10.1007/s11295-006-0042-4 (2006).
57. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* **30**, 683–691 (2004).
58. Lê, S., Josse, J. & Husson, F. FactoMineR: A Package for Multivariate Analysis. *J. Stat. Softw.* **25** (1), 1–18. https://doi.org/10.18637/jss.v025.i01 (2008).
59. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. (2016). https://ggplot2.tidyverse.org Accessed on June 2023.
60. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 7. https://doi.org/10.1371/journal.pone.0037135 (2012).
61. Bett, K. E., Cook, D. R. & KnowPulse (eds) (n.d.). https://knowpulse.usask.ca/ Accessed June 2023.

62. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013). https://doi.org/10.48550/arXiv.1303.3997
63. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140. https://doi.org/10.1111/mec.12354 (2013).
64. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, Volume 23, Issue 19, October 2007, Pages 2633–2635, (2007). https://doi.org/10.1093/bioinformatics/btm308
65. Jombart, T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
66. Liu, X. et al. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genet.* **12**, e1005767. https://doi.org/10.1371/journal.pgen.1005767 (2016).
67. Lipka, A. E. et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399. https://doi.org/10.1093/bioinformatics/bts444 (2012).
68. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U S A.* **100**, 9440–9445. https://doi.org/10.1073/pnas.1530509100 (2003).
69. Kaler, A. S., Gillman, J. D., Beissinger, T. & Purcell, L. C. Comparing different statistical models and Multiple Testing Corrections for Association Mapping in Soybean and maize. *Front. Plant. Sci.* **25**, 10:1794. https://doi.org/10.3389/fpls.2019.01794 (2020).
70. Li, M., Reilly, M. P., Rader, D. J. & Wang, L. S. Correcting population stratification in genetic association studies using a phylogenetic approach. *Bioinformatics* **26** (6), 798–806. https://doi.org/10.1093/bioinformatics/btq025 (2010).
71. Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. R package version 2.30.0. (2022). http://github.com/jdstorey/qvalue Accessed on June 2024.

Acknowledgements

The authors would like to thank Giacomo Nardi, Cristiano Tozzini, Fabio Taccini and all the technicians of the Centre for Agri-Environmental Research "Enrico Avanzi" of the University of Pisa for their support in field trial organization and sampling. We thank Rosario Floriddia (Il Mulino a Pietra) for having provided cv. Robin and Rete Semi Rurali for the facilitation with many of the providers of the lentil genetic material, especially Bari CNR and ARSIAL. Our gratitude also goes to IPK Gatersleben, for University of Perugia, Catania University, Naples University (in the kind person of professor Frusciante) and the CNR Pontecagnano for the provision of genetic material.

Author contributions

Conceptualization, E.L. and P.B.; methodology, E.L., S.C., M.D. and P.B.; software, E.L., M.D., M.M., S.M. and S.C.; validation, E.L., S.C., M.D. and P.B.; formal analysis, E.L., M.D., M.M. and S.C.; investigation, E.L. and S.C.; resources, M.D. and P.B.; data curation, E.L., M.M., S.M. and S.C.; writing—original draft preparation, E.L.; writing—review and editing, E.L., S.C., M.M., S.M., M.D. and P.B.; visualization, E.L., M.M.; supervision, M.D., S.C. and P.B.; project administration, P.B.; funding acquisition, P.B. and M.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work has received funding from "Legumes in biodiversity-based farming systems in Mediterranean basin" project (LEGU-MED) funded by PRIMA Foundation and from Scuola Superiore Sant'Anna, PhD Programme in Agrobiodiversity, who funded E.L.'s study grant.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83516-y.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024